Organisation for Economic Co-operation and Development

DSTI/CDEP(2018)9

For Official Use

English - Or. English

27 April 2018

DIRECTORATE FOR SCIENCE, TECHNOLOGY AND INNOVATION
COMMITTEE ON DIGITAL ECONOMY POLICY

Cancels & replaces the same document of 27 April 2018

ARTIFICIAL INTELLIGENCE IN SOCIETY
PHASE 1

Paris, 16-18 May 2018

Note: This report corresponds to Phase 1 of the draft analytical report on Artificial Intelligence. Phase 2 will be finalised at the November meeting.

Action required: CDEP delegates are invited to discuss this report and to provide directions to the Secretariat.

This document is a contribution to the IOR 1.3.1.1.4 of the 2017-2018 CDEP Programme of Work and Budget

Karine Perset; Karine.Perset@oecd.org, +331 45 24 93 54
Nobuhisa Nishigata, Nobuhisa.Nishigata@oecd.org, +331 45 24 78 94
Anne Carblanc, Anne.Carblanc@oecd.org, +331 45 24 93 34

JT03430927

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

TABLE OF CONTENTS

NOTE BY THE SECRETARIAT

1. The present report constitutes **Phase 1** of the report, for review by the CDEP at its May 2018 meeting. **Phase 2** of the report is expected to be reviewed at the November 2018 meeting of the CDEP and will: *i)* further develop section 2 on trends in development and diffusion of AI, notably by developing a definition and taxonomy of AI that can be used to derive trends in AI-related research and innovation as proxied by patents and scientific publications data (leveraging the STI Micro-data Lab); *ii)* develop Section 3 on AI applications and case studies; and *iii)* further develop and finalise Section 4 on public policy issues associated with AI.

EXECUTIVE SUMMARY

2. The purpose of the report is to help build a shared understanding of Artificial Intelligence (AI) in the present and near-term. It maps some of the economic and social impacts of AI technologies and applications and their policy implications, presenting evidence and policy options. It is also developed to help coordination and consistency with discussions in other international fora, notably the G7 and the G20.

3. **Section 1 on the state of AI research** characterises AI as the broad category of approaches to help make machines "smart" and focuses on the subset of AI called "machine learning" (ML), which uses a statistical approach to teach machines by showing them examples. The goal of the section is to develop an AI research taxonomy that can help policymakers to understand AI trends and identify policy issues in what is now one of the most active research areas in computer science.

4. **Section 2 on trends in development and diffusion of AI** uses available data to show AI's growing prevalence in terms of investment, research, and innovation. In all sectors, financial investment in AI is growing fast with estimates ranging from USD 26 to 39 billion invested in AI in 2016: 70% internally, 10% through acquisitions and 20% through investments in start-ups (MGI, 2017). OECD analysis based on Crunchbase data is consistent with these figures and finds USD 8.5 billion was invested in AI start-ups in 2016. Investment in AI start-ups nearly doubled in 2017, to reach USD 15 billion. AI now represents over 10% of investments in start-ups and this share is increasing in all major economies.

5. Start-ups based in the United States represented about half of the financial investments in AI start-ups both in volume and in value in 2017. As of 2016, the dollar amount of investment in Chinese AI start-ups grew dramatically, to represent over 30% of worldwide investments in 2017. Average investments in Chinese start-ups were USD 150 million per investment transaction, about ten times the worldwide average. Investments in EU start-ups represented 10% of the dollar value of worldwide investments, with the UK representing over half of this. In 2017, Israel captured about 4% of the worldwide investments in AI start-ups and Canada 2%.

6. **Section 3 on AI applications and case studies** [to be developed in Phase 2] will review AI applications in some of the sectors that are seeing widespread or rapid uptake of AI technologies, namely: autonomous vehicles, marketing and advertising, health, scientific discovery and space, surveillance and digital security, public services, criminal justice, and virtual reality. The goal of the section will be to illustrate policy opportunities and challenges in major application areas, building on the OECD October 2017 Conference "*AI: Intelligent Machines, Smart Policies*"[1] and on work being undertaken throughout the OECD.

7. **Section 4 on public policy issues** [of which a preliminary version is included, to be further developed and finalised in Phase 2] details some of the salient public policy issues that AI raises. AI is expected to replace and/or augment components of human labour, requiring policies to facilitate professional transitions and to help workers develop the skills to benefit from, and to complement, AI. Another priority is to ensure that the development and deployment of AI respects human rights, fundamental values and

privacy. Other key issues are to ensure the transparency and accountability of AI-powered decisions that impact people and to prevent algorithmic biases and discrimination. AI also raises new liability, responsibility, security and safety questions as well as considerations related to the open and inclusive development and diffusion of AI.

8. **Section 5 on the policy landscape** first reviews the role of AI in the policy agendas of stakeholders at both national and international levels. Several governments have developed national initiatives to use AI for improving productivity and competitiveness. The section reviews government initiatives in Canada, China, Estonia, Finland, France, Germany, Korea, Japan, the United Kingdom and the United States. These AI strategies also highlight policy challenges associated with AI deployment, reflecting differences in national cultures, legal systems, country size and level of AI adoption.

9. AI is also a priority at the international level, such as at the G7, G20, European Union and U.N. levels. Following the G7 ICT Ministers' Meeting in Japan in April 2016, the G7 ICT and Industry Ministers Meeting in Turin, Italy in September 2017 shared a vision of "human centric" AI and agreed to lead international cooperation and multi-stakeholder dialogue on AI, supported by the OECD. At the G7 Innovation Ministers' Meeting in Canada in March 2018, G7 Innovation Ministers agreed to convene a multi-stakeholder conference on AI in the fall of 2018 to better understand and benefit from AI.

10. Stakeholder groups are actively engaged in discussions on how to steer AI development and deployment to serve all of society. The "Partnership on AI to Benefit People and Society" was originally created by companies leading AI development and now has broad membership. The "AI Initiative" of the Future Society at Harvard's Kennedy School of Government has launched an online platform for civic debate on AI. Several technical, governmental, business, academic and labour organisations have developed principles to guide AI development. They include the IEEE's "Ethically Aligned Design" principles (version 2) and the "Asilomar AI Principles" of the Future of Life Institute. The Japanese Ministry of Internal Affairs and Communications published "AI R&D Guidelines" as input into international discussions. Labour organisation UNI Global Union has developed AI principles.

11. Common themes emerge from these initiatives. There is a consensus that AI should be used for the benefit of people broadly, which translates into calls for appropriate safeguards to ensure that AI systems are designed and operated in a way that is transparent and explainable and that respects human values and rights, democracy, culture and diversity, non-discrimination, privacy and control, safety and security. Other key themes that emerge from existing initiatives are calls for awareness and empowerment, access to data, skills development, clear accountability and responsibility, whole of society dialogue, and measurement. Together these themes could serve as the foundation to develop OECD principles on AI in society. The proposed next step for the OECD is to develop a Council Recommendation on AI in Society, as discussed at the CDEP at its November 2017 meeting.

1. THE STATE OF AI RESEARCH

12. Artificial Intelligence (AI), and particularly the subset of artificial intelligence called “machine learning” (ML), is one of the most active research areas in computer science. A broader range of academic disciplines are leveraging AI techniques for a wide variety of applications.

13. There is no agreed-upon classification scheme for breaking AI into research streams that is comparable for example to the 20 major economics research categories in the Journal of Economic Literature’s (JEL) classification system. The goal of this section is to develop an AI research taxonomy that policymakers can use to understand AI trends and identify policy issues.

14. To start, it is important to clarify the core distinctions between AI and ML because the two are often confused. AI can be characterised the broad category of approaches to help make machines “smart” (Box 1.1). The field started in the 1950s in an area that is now known as symbolic AI research (Figure 1.1) where research was logic-based and required researchers to build detailed and human-understandable decision structures to help machines arrive at human-like decisions. The sheer complexity of developing logical systems to train computers means that applications are limited for this approach, and research on symbolic approaches has largely been supplanted by statistical approaches.

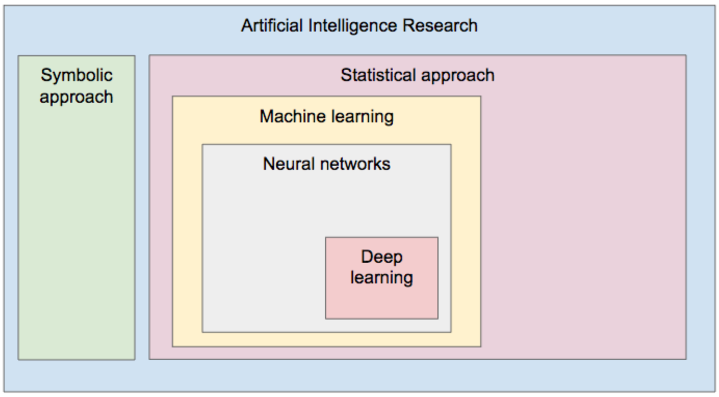
Box 1.1. What is AI?

There is no universally accepted definition of AI. A widely used definition is provided by Nils J. Nilsson (2010): “Artificial intelligence is that activity devoted to making machines intelligent, and intelligence is that quality that enables an entity to function appropriately and with foresight in its environment”. Machines understanding human speech, competing in strategic game systems, driving cars autonomously or interpreting complex data are currently considered to be AI applications. Intelligence in that sense intersects with autonomy and adaptability through AI’s ability to learn from a dynamic environment.

Source: DEO (2017)

15. These statistical approaches, which are collectively referred to as “machine learning,” teach machines to make decisions by showing them many examples of correct decisions. Machine learning itself contains many techniques that have been used by economists for decades. They range from linear and logistic regressions, decision trees, and principle component analysis to deep neural networks.

Figure 1.1. The relationship between artificial intelligence and machine learning



16. In economics, regression models use input data to make predictions in such a way that researchers can interpret the coefficients (weights) on the input variables, often for policy reasons. With machine learning, the primary goal is to make accurate predictions, but people may not be able to understand the models themselves. Additionally, machine learning problems tend to work with many more variables ("features" in machine-learning terminology) than is common in economics, typically in the thousands or higher, and much larger data sets ranging from tens of thousands to hundreds of millions of observations. At this scale, researchers rely on more sophisticated and less-understood techniques such as neural networks to make the predictions. Interestingly, one of the core research areas of machine learning is trying to reintroduce the type of explainability that economists are used to in these large-scale models (see Cluster 4 below).

17. Machine-learning techniques require minimal input from a human operator beyond the initial algorithmic design. The vast majority of today's AI-based systems use this statistical, machine learning-based approach to AI. The present report focuses on the statistical approach and uses "AI", "ML", "algorithms" interchangeably to mean machine learning systems.

18. In recent years, the combination of three trends--growing computational power, massive datasets ("big data"), and the maturity of a statistical modelling technique called "neural networks"--has powered the current expansion in AI development. The real technology behind the current wave of machine learning applications is a sophisticated statistical modelling technique called "neural networks". Finally, deep learning is a phrase that refers to particularly large neural networks; there is no defined threshold as to when a neural net becomes "deep".

19. This evolving dynamic in academic AI research, paired with continual advances in computational abilities, data availability, and neural network design means the statistical approach to AI will continue to overtake symbolic systems research and dominate AI research for the foreseeable future. As a result, policymakers should focus their attention on the subset of machine learning developments that will likely have the largest impact and represent some of the most difficult policy challenges. These challenges include unpacking the machines' decisions and making the decision-making process more transparent.

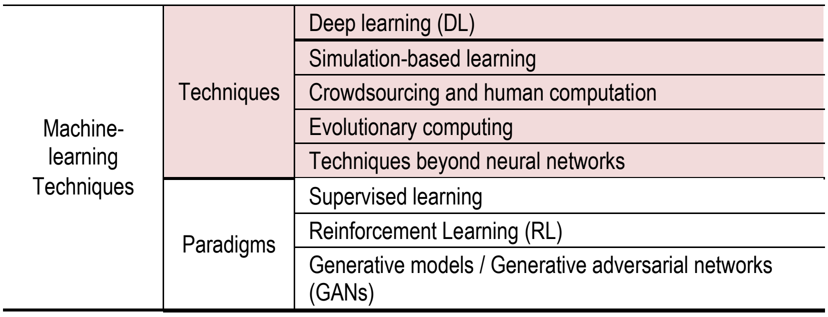
20. There is no widely agreed-upon taxonomy for AI research, nor for the subset of machine learning. The taxonomy proposed below represents 25 machine-learning research streams organised into 4 broad categories and 9 subcategories. It is important to

note that unlike traditional economic research traditions where researchers may focus on a narrow research area, AI researchers commonly work across multiple clusters simultaneously to solve open research problems.

1.1. Cluster 1: Machine learning techniques

21. The first broad category of machine learning research focuses on the techniques and paradigms used in machine learning. Similar to quantitative methods research in the social sciences, this line of research builds and supplies the technical tools and approaches that are used in machine learning applications (Table 1.1).

Table 1.1. Cluster 1: Machine learning techniques



Machine-learning Techniques	Techniques	Deep learning (DL)
		Simulation-based learning
		Crowdsourcing and human computation
		Evolutionary computing
		Techniques beyond neural networks
	Paradigms	Supervised learning
		Reinforcement Learning (RL)
		Generative models / Generative adversarial networks (GANs)

22. The category is dominated by neural networks (of which "deep learning" learning is a subcategory) and forms the basis for most machine learning today. Machine learning also includes various paradigms for helping the system learn. Reinforcement learning trains the system in a way that mimics the way humans learn - via trial and error. The algorithms are not provided explicit tasks, but rather learn by trying different options in rapid succession - receiving rewards or punishments as outcomes - and then adapting accordingly. This has been referred to as relentless experimentation (Knight, 2017).

23. Generative models, including generative adversarial networks (GANs) train a system to produce new data similar to an existing dataset. They are an exciting area of AI research because they pit two or more unsupervised neural networks against each other in a zero-sum game. In game theory terms, they function and learn as a set of rapidly repeated games. By setting the systems against each other at computationally high speeds, the systems are able to learn profitable strategies, particularly in structured environments with clear rules.

Cluster 1: Policy relevance

24. A number of relevant policy issues are linked to machine learning techniques. They include: investments in basic science, better training data sets, funding academic research, and implications for science, technology, engineering, and mathematics (STEM) and computing education. For example, research funding from the Canadian government supported the breakthroughs that led to the extraordinary success of modern neural networks (Allen, 2015).

1.2. Cluster 2: Ways of improving machine learning / optimisations

25. The second broad category of research focuses on ways to improve and optimise machine learning tools and breaks down research streams based on the time horizon for results (current, emerging and future) (Table 1.2). Short-term research is focusing on speeding up the deep learning process, either via better data collection or by using distributed computer systems to train the algorithm.

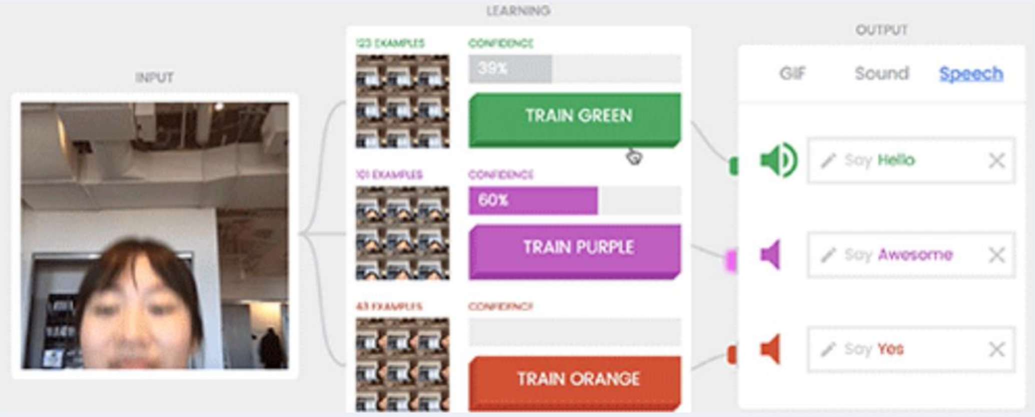
Table 1.2. Cluster 2: Ways of improving machine learning / optimisations

Ways of improving machine learning	Enabling factors (current)	Faster deep learning
		Better data collection
		Distributed training algorithms
	Enabling factors (emerging)	Performance on low-power devices
		Learning to learn / Meta learning
		AI developer tools
	Enabling factors (future)	Understanding neural networks
		One-shot learning

26. Researchers are also focusing on enabling machine learning on low-power devices such as mobile phones and Internet of Things (IoT) devices. Significant progress has been made on this front, with projects such as Google's Teachable Machine now offering open-source machine learning tools that are so lightweight that they run in a browser (Box 1.2). Teachable Machine is just one example of emerging AI development tools that are meant to expand the reach and efficiency of machine learning.

Box 1.2. Teachable Machine

Teachable Machine is a Google experiment that allows people to train a machine to detect different scenarios using a camera built into a phone or computer. The user takes a rapid series of pictures for three different scenarios (e.g. different facial expressions) to train the teachable machine. The machine then analyses all the photos in the training data set and can use them to detect different scenarios. For example, the machine can play a sound every time the person in a camera range smiles. What makes Teachable Machine stand out as a machine learning project is that the neural network runs exclusively in the user's browser without any need for outside computation or data storage.



Source: https://experiments.withgoogle.com/ai/teachable-machine

27. Machine learning research with a longer time horizon includes studying the mechanisms that allow neural networks to learn so effectively. Although neural networks have proven to be a powerful machine learning technique, understanding of how they operate is still limited. Understanding these processes would make it possible to engineer neural networks on a much deeper level than is currently feasible. Longer-term research is also looking at ways to train neural networks using much smaller sets of training data, sometimes referred to as "one-shot learning," and on generally making the training process much more efficient. Today, large models can take weeks or months to train and require hundreds of millions of training examples.

Cluster 2: Policy relevance

28. The policy relevance of the second cluster includes the implications of running machine learning on stand-alone devices, the potential for a reduction in energy use, and the need to develop better AI tools to expand its beneficial uses.

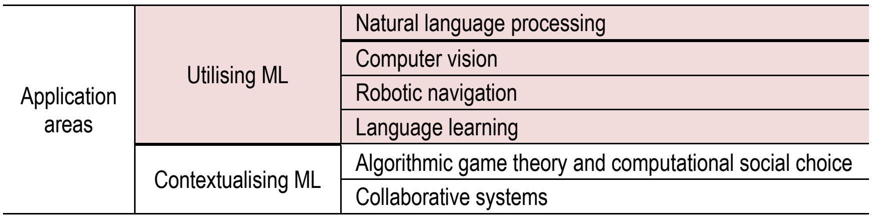
1.3. Cluster 3: Application areas

29. The third broad research category applies machine learning methods to solve various practical challenges in the economy and society. Examples of applied machine learning are emerging across fields in much the same way as Internet connectivity transformed certain industries first and then swept across the entire economy. The OECD Digital Economy Outlook (2017) provides a range of examples of machine learning

applications emerging across OECD countries. For this section, the research streams presented in Table 1.3 represent the largest areas of research linked to real-world application development (Table 1.3).

30. Core applied research areas that utilise machine learning include natural language processing, computer vision and robotic navigation. Each of these three research areas represents a rich and expansive research field. Research challenges may be confined to just one area, or can also span multiple research streams. For example, researchers in the United States are using a combination of natural language processing and computer vision to aid with breast cancer screening (Yala et al., 2017).

Table 1.3. Cluster 3 - Application areas



Application areas	Utilising ML	Natural language processing
		Computer vision
		Robotic navigation
		Language learning
	Contextualising ML	Algorithmic game theory and computational social choice
		Collaborative systems

31. Two research lines focus on ways to contextualise machine learning. Algorithmic game theory sits at the intersection of economics, game theory and computer science and uses algorithms to analyse and optimise multi-period games. Collaborative systems are an approach to large challenges where multiple machine learning systems combine to tackle different parts of complex problems.

Cluster 3: Policy relevance

32. Relevant policy issues linked to machine learning applications include discussions on the future of work, understanding the potential impact of AI, human capital and skills development, determining in which situations AI applications may or may not be appropriate in sensitive contexts, AI's impact on industry players and dynamics, government open data policies, regulations for robotic navigation, and privacy policies that govern the collection and use of data.

1.4. Cluster 4: Refining machine learning with context

33. The fourth broad research category examines machine learning context from technical, legal and social perspectives. As we increasingly rely on algorithms to make important decisions, it is important from a societal perspective to be able to unpack how algorithms come to specific decisions, and to work to eliminate bias from their outcomes. One of the most active research areas in machine learning deals with transparency and accountability of AI systems Table 1.4). Statistical approaches for AI have led to less human-comprehensible computation in algorithmic decisions that can have significant impacts on the lives of individuals - from bank loans to parole decisions (Angwin et al, 2016).

Table 1.4. Cluster 4 - Refining machine learning with context

Refining ML with context	Explainability	Transparency and accountability
		Explaining individual decisions
		Simplification into human comprehensible algorithms
		Fairness / bias
		Debug-ability
	Safety & Reliability	Adversarial examples
		Verification
		Other classes of attacks

34. Another category of contextual ML research involves steps to ensure the safety and integrity of these systems. Researchers' understanding of how neural networks arrive at decisions is still at an early stage, and neural networks can often be tricked using simple methods such as changing a few pixels in a picture (Ilyas et al, 2018). Researchers in these streams are trying to understand how to defend systems against adversarial attacks and how to verify the integrity of machine-learning systems.

Cluster 4: Policy relevance

35. A number of relevant policy issues are linked to the context surrounding machine learning including requirements for algorithmic accountability, combatting inherent bias, the rights of individuals who are affected by algorithms, product safety, liability and security.

Box 1.3. Predictions for the future of AI shared at the OECD conference "AI: Intelligent Machines, Smart Policies", October 2017

The OECD held the conference "AI: Intelligent Machines, Smart Policies" in Paris on 26 and 27 October, 2017 (http://oe.cd/ai). Speakers at the conference presented machine leaning as the best method for AI systems to learn about and adapt to complex interactions with the real world. However, machine learning still requires access to large amounts of curated and accurate data. It is also very difficult to explain machine learning decisions and to formally validate results.

Future machine learning systems will learn with less data

Experts predicted AI systems would evolve to learn from their interactions, as humans do, rather than just from data. One expert noted that good machine learning algorithms are transferable between applications because they leverage endemic characteristics of the learning process. Several participants cited AlphaGo Zero, which used reinforcement learning (Table 1.1) to learn to play Go superhumanly from scratch by playing Go against itself. They cautioned, however, that compared to the game of Go, the "rules of the game" in the real world are much more complex and not as easily replicated.

One researcher advocated creating a digital reality by modelling the real world to generate synthetic input data that could be used to train AI systems on specific situations or scenarios. Another research goal outlined by participants was for AI to evolve toward more proactive decision-making support for humans. Another focus of AI research was to ensure AI systems design goes from optimising a given objective to more nuanced but complex "provably beneficial AI" for humans, whereby reinforcement learning systems are designed that behave in such a way that humans overall are happy with the results.

Future machine learning systems will bridge the physical and digital worlds

Several participants presented visions of the future in which standalone AI systems would evolve into AI networks that communicate and interact. In this context, they highlighted the need for interoperability and communication between the AI systems of different companies and for data standardisation. For example, such standardisation would enable communication between the autonomous vehicles of different manufacturers.

Several experts said they believe that AI would become truly revolutionary when individual AI systems can connect and interoperate, understand each other, work together without humans teaching them, and act with "common sense." One speaker predicted AI would seamlessly interconnect the physical and digital worlds for humans by 2025-2035.

AI capabilities will have few limits; humans must tread carefully towards AGI

Some predictions pushed the discussion in unexpected territory. Artificial General Intelligence (AGI) and "singularity" were not on the conference agenda. But a few participants brought up the need to avoid assumptions on the upper limits of future AI capabilities. Instead, they said there is a need to plan carefully for the possible development of AGI, which is broadly defined as machines being able to perform all the same tasks as humans. [2][3][4][5][6]

Government involvement in AI research

At the Conference, there was a call to consider the appropriate level of government involvement in AI research, notably to resolve societal grand challenges. One speaker proposed the development of a "CERN for AI," a worldwide-recognised public research infrastructure that could share the use of curated databases and open-source programs, provide training, simulate critical scenarios and validate and test AI systems in a digital reality.

A return to models?

One participant expressed confidence that, from simply detecting correlations, the future would involve a return to models and causality links.

2. TRENDS IN DEVELOPMENT AND DIFFUSION OF AI

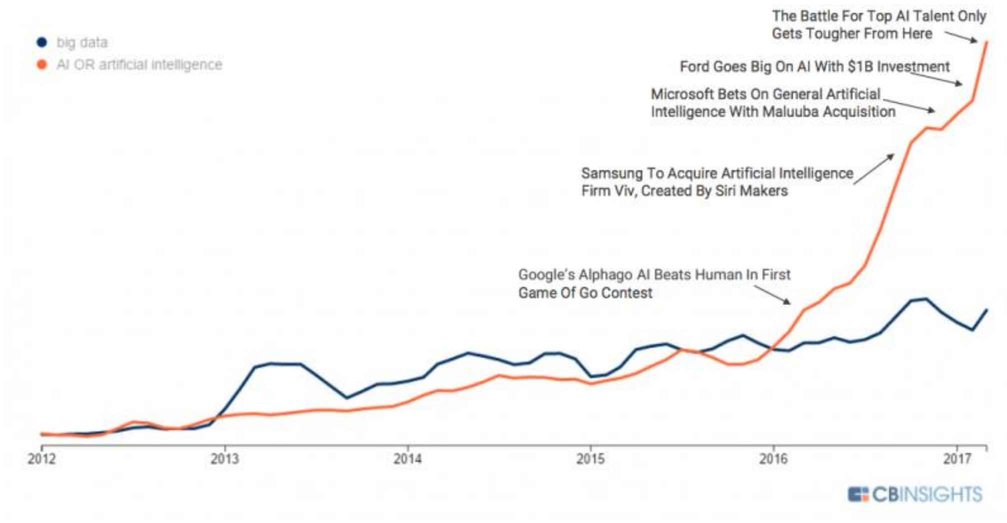
36. This section explores trends in tracking research, development and diffusion of AI based on available data on investment in AI research and applications in different countries, types of AI technologies, sectors, and organisation types. Phase 2 will aim to develop a definition and taxonomy of AI that can be used to derive trends in AI-related research and innovation as proxied by patents and scientific publications data (leveraging the STI Micro-data Lab).

2.1. Introduction

37. The principle of AI was conceptualised by John McCarthy, Alan Newell, Arthur Samuel, Herbert Simon and Marvin Minsky in the 1956 Dartmouth Summer Research Project that started AI. While AI research progressed steadily over the past 60 years, early promises were overly optimistic, leading to an "AI Winter" of reduced funding and interest in AI research during the 1970s.

38. As seen in Section 1, more recently, the availability of big data, cloud computing and recent breakthroughs in neural networks have dramatically increased the power, availability and impact of AI. As AI systems reach increasingly deeply into the global economy, tremendous media attention has accompanied the dramatic growth of investment in AI. For example, Asia was riveted when Google's AlphaGo AI program defeated the human Go champion Lee Sedol from Korea in 2016 – a feat that experts thought would take at least ten more years to accomplish (Figure 2.1).

Figure 2.1. AI and big data investments, timeline 2011-2017



Source: CBI (2018)

2.2. Investments in AI and in AI start-ups

39. AI investment as a whole is growing fast and AI already has significant business impact. MGI (2017) estimated that financial investment in AI worldwide represented USD 26 billion to 39 billion in AI in 2016, of which internal corporate investment represented about 70 percent, AI acquisitions some 10 percent and investment in AI start-ups some 20 percent.[7] The large majority of these investments (three quarters) were made by large technology companies. AI adoption outside of the technology sector is at an early stage and few firms have deployed AI solutions at scale. Large companies in other digitally mature sectors that have data that they can leverage, notably in the financial and automotive sectors are also adopting AI.

40. AI start-ups are being acquired at a rapid pace by large technology companies. According to CBI Insights (2018b), the companies that have acquired the most AI start-ups since 2010 include Google, Apple, Baidu, Facebook, Amazon, Intel, Microsoft, Twitter and Salesforce. Several AI cybersecurity start-ups were acquired in 2017 and early 2018, *e.g.* Amazon purchased Sqrrl and Oracle purchased Zenedge.

41. AI start-ups are also acquisition targets for companies in more traditional industries, notably automotive companies, healthcare companies such as Roche Holding or Athena Health and insurance and retail companies.

Box 2.1. The Crunchbase database

This section estimates investments in AI start-ups based on information contained in Crunchbase, a commercial database on innovative companies created in 2007 (further detailed in Annex 1). The version of the Crunchbase database used for this report was downloaded in April 2018 and contains information on over 500 000 entities in 199 different countries. Data contained in Crunchbase is provided by: global investment firms who provide monthly portfolio updates (some 3 000 investors); entrepreneurs and executives who update company profiles (some 500 000 companies), and by a team of analysts at Crunchbase that scans for anomalies and enriches profiles using machine learning algorithms.

Breschi, Lassebie, and Menon (2018) discuss the coverage and representativeness of the database. They find that Crunchbase covers investment deals and start-ups better than comparable data sources and that, with few exceptions, its coverage is sufficiently exhaustive across OECD and BRIICS countries. Companies are classified into 45 "sectors". The full database covers 230 000 investment transactions, 11 000 IPOs, and 34 000 acquisitions. It cover 50 000 investors, e.g., investment banks, business angels, incubators.

Caveats to using data in Crunchbase include: likely underreporting of company failures (a large majority of companies are reported as still active although failure rates are known to be high for innovative start-ups); a database scope that is not precisely defined; undisclosed investment amounts for many investment transactions (about a quarter of investment transactions in AI start-ups) and sample selection bias.

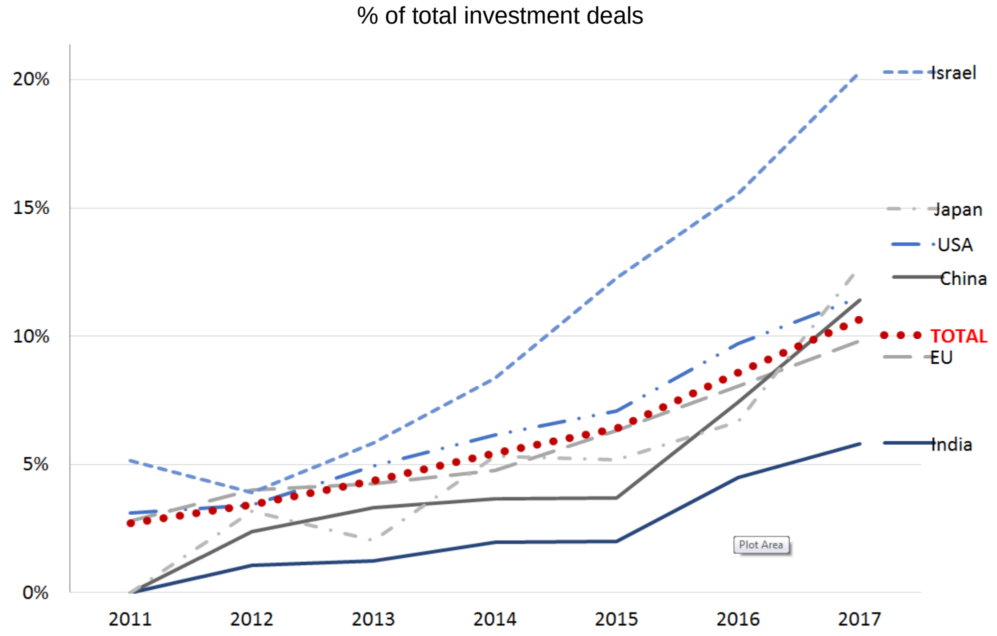
The companies considered to be AI start-ups in this report correspond to companies categorised in the 'artificial intelligence' sector of Crunchbase (2245 companies) and to companies that described used the keywords 'AI', 'artificial intelligence', 'machine learning', 'autonomous vehicle' and 'image recognition' in the company's short description (an additional 517 companies).

42. Private investments in AI start-ups -- venture capital and private equity financing, grants, and seed investments -- grew rapidly from 2011 to 2017, particularly in the United States, to a combined total of USD 8.5 billion in 2016 and USD 15 billion in 2017.

AI now represents over 10 percent of financial investments in start-ups

43. AI represents an increasing share of investments in all start-ups, from less than 3 percent in 2011 to more than 10 percent in 2017 (Figure 2.2).

Figure 2.2. AI as a share of financial investments in start-ups, 2011-2017



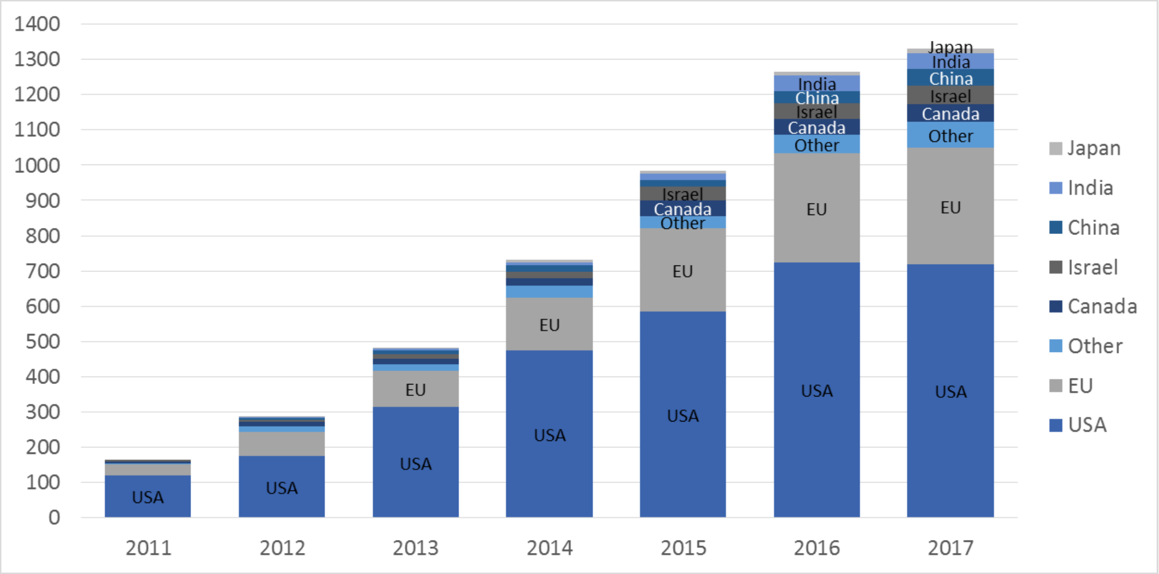
Source: OECD (est.), based on Crunchbase (April 2018), www.crunchbase.com

The number of 'deals' in AI start-ups grew by 35 percent annually from 2011 to 2017

44. The number of investment transactions, *i.e.* the number of 'deals' in AI start-ups grew globally by 35 percent compound annual growth rate from 2011 to 2017, from 150 investment transactions in 2011 to more than 1,300 in 2017 (Figure 2.3).

Figure 2.3. Number of financial investments in AI start-ups, 2011-2017

By start-up location



Source: OECD (est.), based on Crunchbase (April 2018), www.crunchbase.com

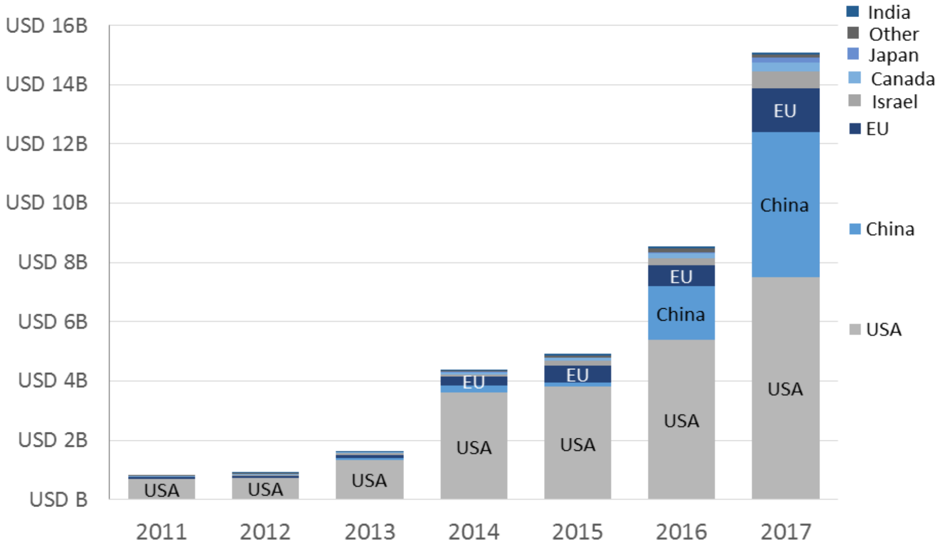
Start-ups based in the United States still attract more investment

45. In terms of both the number of investment transactions (Figure 2.3) and the amount of money raised (Figure 2.4), start-ups based in the United States attracted a significant portion of investments. In 2017, half of the dollar amount invested in AI start-ups was to companies based in the United States. Over the seven years from 2011 through 2017, more than USD 36 billion was invested in AI start-ups, two thirds of them based in the United States (Figure 2.4). This is consistent with research finding that the United States accounts for 70-80 percent of global venture capital investments across all technologies (Breschi, Lassebie, and Menon, 2018).

46. However, financial investments in other countries increased significantly: in 2017, about half of the number of investments made (Figure 2.3) and half of the money raised (Figure 2.4) was outside the US. From 2011 to 2015, the European Union was the second largest player and represented 10 percent of the value of investments over that period. Dramatic growth in investments in Chinese start-ups since 2016 has vaulted China into the second-largest AI powerhouse in terms of dollars. In 2017, China was responsible for 32 percent of the value of investments in AI start-ups (Figure 2.4).

Figure 2.4. Total estimated investments in AI start-ups (USD billion), 2011-2017

By start-up location



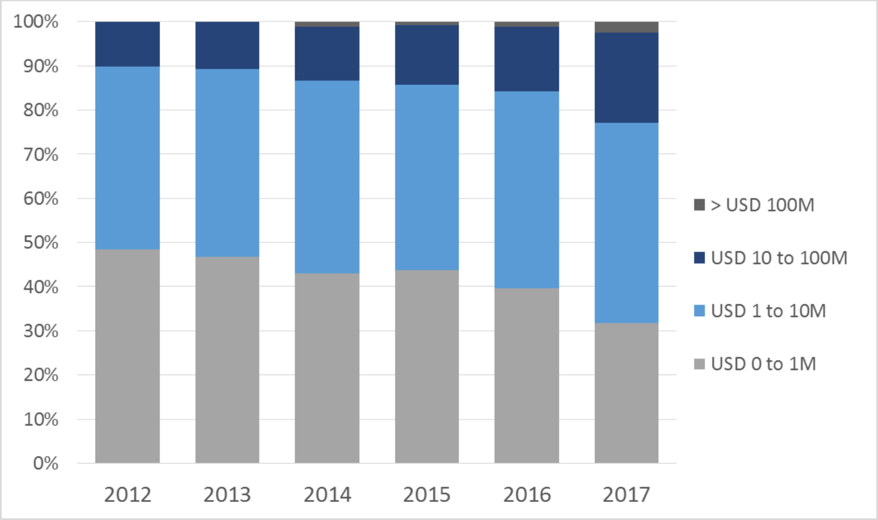
Source: OECD (est.), based on Crunchbase (April 2018), www.crunchbase.com

Financial investments in AI start-ups are larger

47. In 2012 and 2013, 90 percent of the investment transactions were worth less than USD 10 million; only 10 percent were between USD 10 and 100 million; and none were for more than USD 100 million. In 2017, one out of five of the investment deals was for more than USD 10 million. This could be partly due to start-ups becoming more mature and attracting larger, later stage investments. Another factor is likely to be increasing recognition of the role AI is playing in the global economy. In 2017, 22 companies attracted investments of more than USD 100 million for a total of USD 8 billion; which was nearly half of the total amount invested in AI start-ups that year. The company that attracted the largest investment was Chinese company Toutiao [8], a content recommendation system that provides personalised content to users in China through analysis of social networking data in which USD 3 billion was invested. The second largest investment was in Argo AI, an autonomous vehicles company that Ford acquired in 2017 and in which it committed to invest USD 1 billion.

Figure 2.5. Size of investment transactions

% of total number of investment deals



Source: OECD (est.), based on Crunchbase (April 2018), www.crunchbase.com

Table 2.1. Mega-deals (over USD 100 million) per country, 2014 to 2017

Year	Country	Company	Description	Amount raised
2014	**United States**	Cloudera	Cloudera is an enterprise software company that provides Apache Hadoop-based software and training to data-driven enterprises.	$900 M
		InsideSales.com	Insidesales.com offers sales acceleration platform built on a predictive and prescriptive self-learning engine.	$100 M
		Magic Leap	Magic leap is a proprietary wearable technology that enables users to interact with digital devices in a completely visually cinematic way.	$542 M
		Sentient Technologies	Sentient Technologies is the world's leading artificial intelligence company specialising in evolutionary intelligence.	$104 M
	China	Toutiao	Toutiao is a recommendation system product based on data mining, which recommends valuable, personalised information to users in china.	$100 M
	Total			**$1,746 M**
2015	**United States**	Banjo	Banjo provides real time content discovery by location across multiple social networks.	$100 M
		Dataminr	Dataminr's realtime AI platform discovers high impact events & critical breaking news far in advance of existing information systems	$130 M
		Udacity	Udacity is an e-learning platform that offers credential programs in artificial intelligence, machine learning, and robotics.	$105 M
		Wish	Wish gives everyone access to the most affordable, convenient, and effective shopping mall and affordable goods.	$500 M
		ZestFinance	Zestfinance's credit-decisioning platform helps lenders predict credit risk so they can increase revenues, reduce risk & ensure compliance.	$150 M
	United Kingdom	ACORN OakNorth Holdings Ltd.	A fintech platform that is unlocking the complex SME lending market globally by leveraging AI and machine learning	$100 M
	Total			**$1,085 M**
2016	**United States**	Cylance	Cylance is a global provider of cybersecurity products and services to solve security problems.	$100 M

		Kernel	Kernel is building advanced neural interfaces to treat disease, illuminate the mechanisms of intelligence, and extend cognition.	$100 M
		Magic Leap	Magic leap is a proprietary wearable technology that enables users to interact with digital devices in a completely visually cinematic way.	$794 M
		Velodyne LiDAR	Velodyne lidar, inc. is a leading developer, manufacturer and supplier of high performance LIDAR sensors (measures distance to a target).	$150 M
		Wish	Wish gives everyone access to the most affordable, convenient, and effective shopping mall and affordable goods.	$500 M
		Zoox	Zoox is a robotics company pioneering autonomous mobility as-a-service.	$250 M
	China	iCarbonX	Icarbonx is a china-based artificial intelligence platform for health data company.	$154 M
		Megvii	Megvii, short for mega vision, is a knowledge-intensive enterprise focusing on independent research and development of AI.	$100 M
		ROOBO	Roobo is a hardware and AI company that offers pre-orders of Domgy (pet robot) via crowdfunding sites in china.	$100 M
		WM Motor	Wm motor is an emerging Chinese auto maker.	$1,000 M
	Switzerland	MindMaze	Mindmaze builds human machine interfaces combining mixed reality, artificial intelligence, brain imaging, and neuroscience.	$100 M
	Total			**$3,348 M**
2017	**China**	Cambricon Technologies	Cambricon Technologies develops artificial intelligence chips.	$100 M
		Future Mobility	Future mobility is aiming to sell premium electric cars by 2020 on a worldwide scale.	$200 M
		LingoChamp (Liulishuo)	Through cutting-edge AI technology & innovative product design, we help users learn English more efficiently and communicate with the world.	$100 M
		Megvii	Megvii, short for mega vision, is a knowledge-intensive enterprise focusing on independent research and development of AI.	$460 M
		Mobvoi Inc.	An AI company that develops technologies in Chinese language speech recognition, natural language processing and vertical mobile search.	$180 M
		Toutiao	Toutiao is a recommendation system product based on data mining, which recommends valuable, personalised information to users in china.	$3,000 M
		WM Motor	Wm motor is an emerging Chinese auto maker.	$151 M
	United States	Argo AI	Argo AI: vehicles that are fully autonomous and able to navigate city streets.	$1,000 M
		Brain Corporation	Brain corporation is a developer of autonomous navigational technologies.	$114 M
		ClearMotion	Clearmotion is an automotive technology company	$100 M
		Grammarly	Grammarly's AI-powered products help people communicate more effectively.	$110 M
		i.am+	i.am+ designs and manufactures wearable products.	$117 M
		Lemonade	Lemonade offers homeowners and renters insurance powered by artificial intelligence and behavioral economics.	$120 M
		Magic Leap	Magic leap is a proprietary wearable technology that enables users to interact with digital devices in a completely visually cinematic way.	$502 M
		MIT-IBM Watson AI Lab	MIT-IBM Watson AI lab focuses on fundamental artificial intelligence research.	$240 M
		Nauto, Inc.	Nauto is an AI-powered autonomous vehicle technology company.	$159 M
		Plenty Inc.	Plenty is an agriculture technology company that develops plant sciences for crops to flourish in a pesticide- and GMO-free environment.	$200 M
		System1	System1 fuses technology and science to identify & unlock consumer intent.	$270 M
		Uptake	An SaaS-based product company partnering with global industry to improve productivity, safety, security and reliability.	$117 M
	United Kingdom	ACORN OakNorth Holdings Ltd.	A fintech platform that is unlocking the complex SME lending market globally by leveraging AI and machine learning	$331 M
	Canada	Element AI	Element AI is the platform that helps organisations embrace an AI-first world for today and tomorrow.	$102 M
		LeddarTech	LeddarTech develops led detection and ranging solutions for object recognition and distance measurement applications.	$107 M
	Total			**$7,780 M**

Source: OECD, based on Crunchbase (April 2018), www.crunchbase.com

China has fewer, larger investments

48. The profile of investments in Chinese start-ups seems different from other parts of the world. Investments in Chinese start-ups were worth an average of 150 million USD in 2017, twice the average amount invested in 2016. When investments above USD 100 million are excluded, the average investment in China in 2017 was USD 26 million, compared to USD 20 million in 2016 (Table 2.2). However, there were comparatively few investments in China (Figure 2.2). In other countries, average investments in 2017 were just a fraction of that amount. While the average amount per investment increased in all countries, three models can be distinguished: Chinese start-ups with few but very large investments; EU start-ups with a steadily increasing number of smaller investments (USD 4 million on average per investment in 2017) and the US with a steadily increasing number of larger investments (USD 8 million on average per investment in 2017). These differences in investment profiles are significant when the mega deals of over USD 100 million are excluded (Table 2.2) but also when mega deals are included (Table 2.3).

Table 2.2. Average amount raised per deal (USD million), for deals up to USD 100 million

	Canada	China	EU	Israel	Japan	United States
2015	$2 M	$12 M	$2 M	$4 M	$4 M	$6 M
2016	$4 M	$20 M	$3 M	$6 M	$5 M	$6 M
2017	$2 M	$26 M	$4 M	$12 M	$14 M	$8 M

Source: OECD (est.), based on Crunchbase (April 2018), www.crunchbase.com

Table 2.3. Average amount raised per deal (USD million), for all AI deals

	Canada	China	EU	Israel	Japan	United States
2015	$2 M	$12 M	$3 M	$4 M	$4 M	$8 M
2016	$4 M	$73 M	$3 M	$6 M	$5 M	$10 M
2017	$8 M	$147 M	$6 M	$12 M	$14 M	$14 M

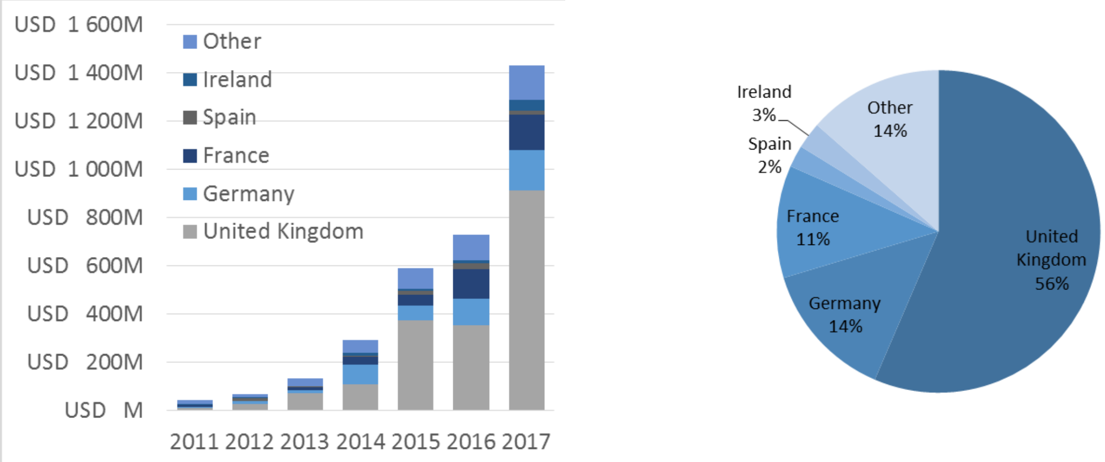
Source: OECD (est.), based on Crunchbase (April 2018), www.crunchbase.com

EU investment in AI start-ups is lead by the U.K.

49. Investments in AI start-ups based in the EU doubled between 2016 and 2017. The United Kingdom represented 56 percent of the investments in EU-based start-ups over the period 2011-2017, followed by Germany (14 percent) and France (11 percent) (Figure 2.6).

Figure 2.6. Investments in AI startups based in the EU (USD million), 2011-2017

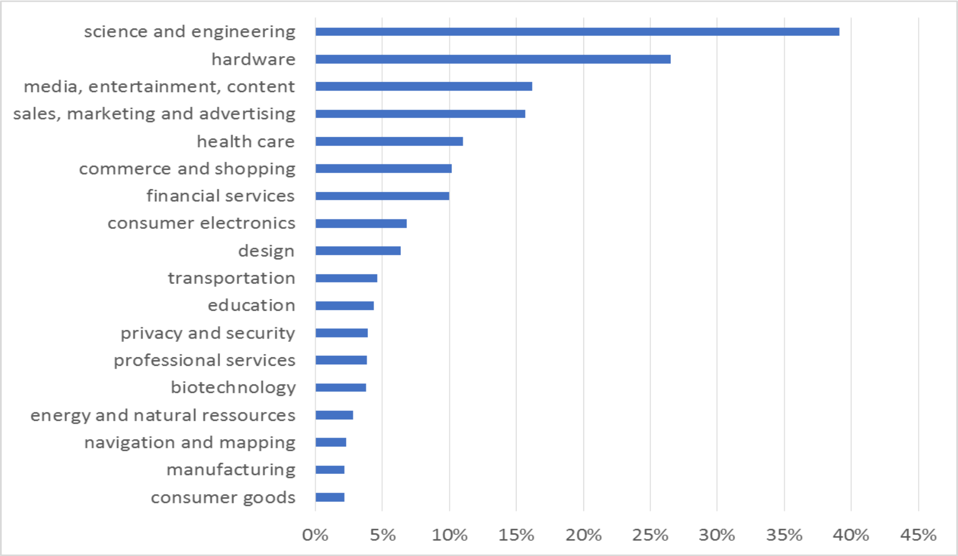
Percentage per country



Source: OECD (est.), based on Crunchbase (April 2018), www.crunchbase.com

50. AI start-ups operate in many sectors (Figure 2.7). 39 percent of AI start-ups define themselves partly or totally as science and engineering companies, including automotive. The next most common categories were: hardware (27 percent), media, entertainment and content (16 percent), and sales, marketing and advertising (16 percent). Healthcare, shopping and financial services each represent about 10 percent of AI start-ups. Consumer electronics, design, transportation, education, privacy and security, each represent less than 5 percent.

Figure 2.7. Top sectors of AI startups from 2011 to 2017 [9] (preliminary)



Source: OECD (est.), based on Crunchbase (April 2018), www.crunchbase.com

Profile of investors in AI start-ups (to be developed in Phase 2)

51. This sub-section will analyse Crunchbase data to try to characterise the profile of investors in AI start-ups.

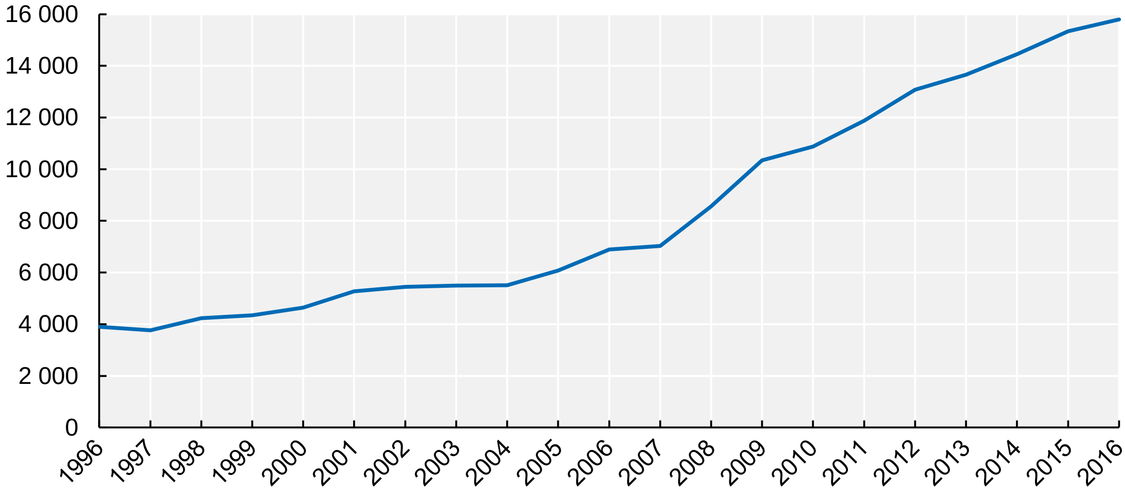
2.3. Insights from scientific publications [preliminary, to be developed in Phase 2]

52. This section will examine trends in AI research based on AI journals in the SCOPUS database (*e.g.*, growth, country of the institution the author(s) are affiliated with, and application areas/sectors).

53. Scientific publications related to AI are valuable to in identifying trends in AI research. In the field of machine learning, scientific publications are a popular way to share research results and seek peer review. Experts point out that the ML training data tends to be viewed as the defensible asset, rather than the ML technologies themselves.[10]

Figure 2.8. Total number scientific documents in AI journals, 1996-16

Based on journals assigned to the AI subject of the All Science *Journals Classification (ASJC)*

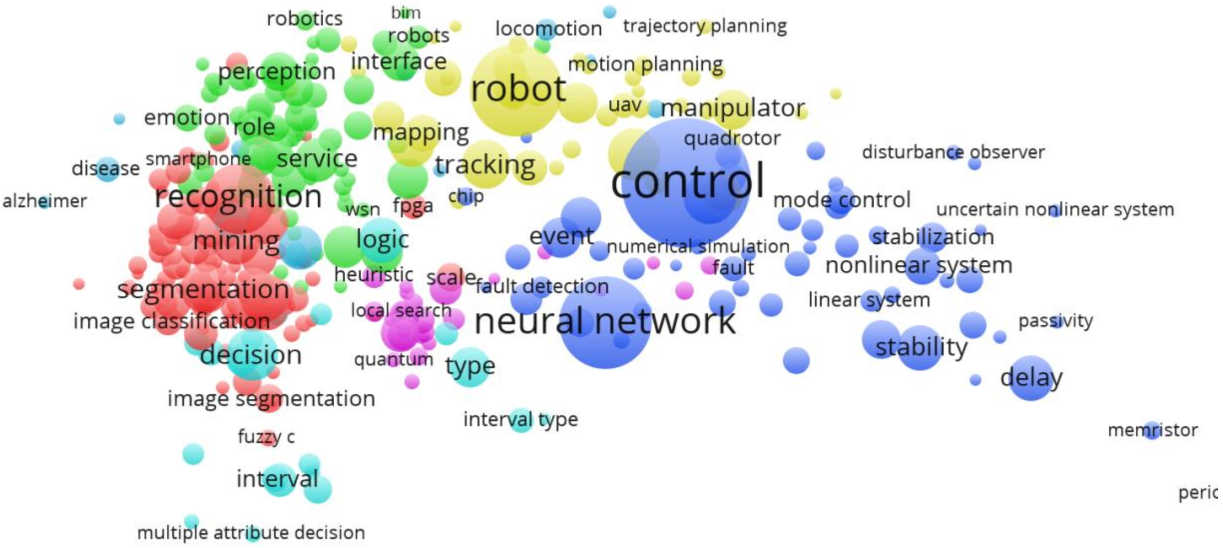


Source: OECD calculations based on Scopus Custom Data, Elsevier, Version 1.2018.

54. As Figure 2.8 shows, scientific advances in AI over the last 20 years have been growing steadily, as proxied by documents published in peer-reviewed journals assigned to the AI subject in the All Science Journals Classification (ASJC) in SCOPUS. The compound annual growth rate from 2006 to 2016, at close to 10 percent, was twice the annual growth from 1996 to 2006. Figure 2.9 clusters areas of AI research based on the titles of scientific documents used in “AI” journals and scientific publications.

55. This section is expected to build on work being undertaken by the OECD Working Party on Industry Analysis (WPIA) to refine the operational definition of AI in consultation with experts and companies, to identify trends in AI R&D and use.

Figure 2.9. Frequently-used keywords in titles of articles in "AI" journals and scientific publications, 2014-2017



Source: OECD calculations based on Scopus Custom Data, Elsevier, Version 1.2018.

2.4. Insights from patenting activity [to be developed in Phase 2]

56. This section will examine trends in AI-related patenting activity, such as growth, countries, active companies, sectors and penetration into different application domains. This section is expected to provide new metrics on AI-related inventions, building on existing methodological work to classify AI and machine learning patents, based on the International Patent Classification (IPC) codes listed in patent documents (Inaba and Squicciarini, 2017) and on keyword searches, on work conducted by RIETI, an external research institute of the Japanese Ministry of Economy, Trade and Industry (Fujii and Managi, 2017), and on other recent research (Cockburn, Henderson and Stern, 2017). The section will also investigate the patents filed by companies referenced in the Crunchbase dataset for further refinement of the OECD methodology to quantify AI-related patents.

2.5. Other possible indicators [to be developed in Phase 2]

57. This section will explore other possible indicators of AI activity such as access to advanced semiconductor, microprocessor and high-performance computing technologies per country. The rationale for looking at computing power is that it is part of the basic infrastructure necessary for institutions to be able to leverage data and AI.

58. It could also explore possible sources of data on AI R&D expenditure such as the OECD Main Science and Technology Indicators (MSTI) Database.

3. AI APPLICATIONS AND CASE STUDIES [to be developed in Phase 2]

59. This section will review AI applications in sectors seeing widespread or rapid uptake of AI technologies, to illustrate policy opportunities and challenges in these specific areas.[11] The section will build on the presentations given at the OECD October 2017 Conference "*AI: Intelligent Machines, Smart Policies*"[12]; on the Digital Economy Outlook (2017) and on work being undertaken throughout the OECD *e.g.* by the International Transport Forum (ITF), the Committee for Scientific and Technological Policy (CSTP), the Committee on Public governance (PGC) and by working parties of the Committee on Digital Economy Policy.

3.1. Autonomous vehicles

60. Investments and research efforts in the area of autonomous vehicles are significant. Companies and research institutes working on autonomous vehicles include Argo AI (Ford Motor company), BMW, General Motors, Tesla, Uber, Waymo (Google), Aurora, and MIT. They are developing virtual driver systems with sensors including cameras, radar, light detection, and ranging radar known as LIDAR; software and computer platform, as well as high-definition maps.

61. Autonomous vehicles like driverless cars and trucks promise cost, safety, quality of life and environmental benefits. [13] They also raise policy questions related to safety, security, privacy, ethical choices in case of an accident, accountability and liability, new infrastructure requirements, but also employment and skills since driving is a major source of jobs. This section is expected to draw on work by the ITF[14] and the OECD Working Party on Communications Infrastructures and Services Policy (CISP).[15]

3.2. Marketing and advertising

62. In marketing and advertising, machine learning algorithms mine data on consumer and user behaviour to target and personalise content, advertising, goods and services, recommendations and prices.

63. They offer benefits including convenience, personalisation, personal satisfaction, and efficiency. They also raise policy questions including their impact on fundamental rights and values, privacy, consumer protection, discrimination, autonomy, self-determination and choice. This section is expected to draw on work by the OECD Committee on Consumer Policy (CCP).[16]

3.3. Health

64. AI applications in healthcare and pharmaceuticals can help detect health conditions early, deliver preventative services and discover new treatments and medications. They also power self-monitoring tools, applications, and trackers and facilitate personalised healthcare.

65. AI in healthcare offers benefits for quality of care and health, life expectancy, cost, and elderly care. Policy questions include access to health data, privacy, and non-discrimination. This section is expected to draw on relevant OECD work, including the Digital Economy Outlook (2017).

3.4. Scientific discovery and space

66. In the space sector, machine learning applications analyse satellite data to make predictions, detect a variety of threats and monitor entire sectors.[17] [18] [19] In science, machine learning is being used to curate data and analyse data sets and scientific literature that exceed human comprehension,[20] with machine learning algorithms using ‘data as the model’ when traditional models cannot account for complex interacting factors.[21] Machine learning robots in laboratories have been able to formulate scientific hypotheses, devise and conduct experiments and analyse test results. Machine learning is being employed to enhance research efficiency by optimising experimental designs.[22]

67. AI applications in space and in the process of science offer benefits including enhanced research productivity and new scientific breakthroughs, more accurate environmental monitoring and disaster prediction and more efficient disaster recovery. Policy questions concern issues of data access and open data, how best to support and incentivise open science and knowledge-sharing, the allocation and scale of public R&D, and scientific education relevant to AI. This section is expected to draw on work being conducted by the OECD Committee for Scientific and Technological Policy (CSTP).

Box 3.1. Enhancing Discovery – The Role of AI in Science (preliminary)

AI promises to improve research productivity at a time when evidence suggests that discoveries in some fields of knowledge are becoming harder to achieve, pressure on public research budgets is increasing, and global challenges – from climate change to disease threats - require scientific breakthroughs. Using AI in science is becoming indispensable in a context where scientific insight depends on being able to draw understanding from vast amounts of scientific data. Furthermore, AI will be a necessary complement to human scientists because the volume of scientific papers is vast and growing, and scientists may have reached ‘peak reading’.

Even if sporadically, forms of AI have been applied to scientific discovery for some time. The AI programme DENDRAL was used in the 1960s to help identify chemical structures, and in the 1970s an AI known as Automated Mathematician assisted mathematical research. Today, the combination of more capable forms of AI, broader developments in scientific instrumentation and the associated generation of enormous volumes of scientific data have increased AI’s utility in science. AI has already been deployed in functions that range from the analysis of large datasets, to hypothesis generation, to comprehension and analysis of scientific literature, to facilitating data gathering, experimental design and experimentation itself.

Key issues

AI is being used across many fields of research. As described in Science (2017), AI is now a frequently used technique in particle physics, which depends on finding complex spatial patterns in vast streams of data yielded by particle detectors. With data gleaned from social media, AI is providing evidence on relationships between language use, psychology and health, social and economic outcomes. AI is tackling complex computational problems in genetics, improving the quality of imaging in astronomy, and helping discover the rules of chemical synthesis, among other uses. The range and frequency of such applications is likely to grow as companies like Data Robot, and others, work to automate the machine learning process, so that scientists, businesses and other users can more readily employ this technology.

Some progress has also occurred in AI-enabled hypothesis generation. IBM has produced a prototype system, Knit, that mines information contained in scientific literature, represents it explicitly in a queryable network, and then reasons on these data to generate new and testable hypotheses. Knit has text mined published literature to successfully identify new kinases that phosphorylate a protein tumor suppressor.

AI is likewise assisting in the review, comprehension and analysis of scientific literature. Natural language-processing has advanced to the point at which it can automatically extract not only relationships but also context from scientific papers. Stanford-based Manning and Gupta present a framework for extracting information from scientific articles, such as main contributions, tools and techniques used, and domain problems addressed, matching semantic extraction patterns in dependency trees. This work is aimed at studying the dynamics of research communities and measuring the influence of one research community on another. But other approaches seek to advance research processes. For example, the Knit system referenced above involves automated hypothesis generation based on text mining of scientific literature. Iris.AI (https://iris.ai/) is a start-up which offers a free tool that extracts key concepts from research abstracts, presents the concepts visually (such that the user can see cross-disciplinary relationships) and gathers relevant papers from a library of over 66 million open access papers. Semantic Scholar (https://www.semanticscholar.org) aims to help scholars explore peer-reviewed literature, and can assess which among a paper's listed references have been most influential in the paper's findings.

AI is likewise assisting in large-scale data collection. One example of this is in citizen science, where Apps that use artificial intelligence can help users to identify unknown animal and plant specimens.

Combined with developments in hardware – such as robotics and the technology of lab-on-a-chip experimentation – AI is central to efforts to automate some areas of science. Automation of discovery requires the effective incorporation of machines in observation, hypothesis generation and experimentation, with all three steps being efficiently linked. The fact that AI can assist in observation and hypothesis generation has been noted above. Progress has also been seen in experimentation. For instance, operating from a laboratory at the University of Aberystwyth in Wales, 'Adam', a robot using artificial intelligence techniques to automatically perform cycles of scientific experimentation, has been described as the first machine to independently discover new scientific knowledge (discovering a compound, Triclosan, which works against wild-type and drug resistant *Plasmodium falciparum*, and *Plasmodium vivax*). And companies such as Transcriptic and Emerald Cloud Lab, in California, are building systems to automate most physical tasks performed by biomedical scientists.

Sparkes et al (2010), suggest that automation of some scientific processes, using closed-loop computational learning systems, could have a number of advantages over human scientists, because 'their biases are explicit, they can produce full records of their reasoning processes, they can incorporate large volumes of explicit background knowledge, they can incorporate explicit complex models, they can anlyse data much faster, and they do not need to rest.'

The problem of intelligibility

In a 2006 article in the Times Higher Education supplement the mathematician Brian Davies worried that "…we can no longer survey the entire proofs of an increasing number of important theorems as we once could, and we have to accept our computers' word that they have carried out our instructions and obtained the result that we suspected. Their own caculations are literally unreadable." This 'black box' problem – the inscrutability of the processes of machine learning - is commonly cited in discussion of AI. The problem is perhaps particularly salient in science, which relies on verification of proofs, on understanding models and on incremental accumulation of knowledge. Research is underway in a number of laboratories aimed at creating AI's which explain their own output. DARPA, for example, is funding 13 different research groups, working on a range of approaches to making AI more explainable. But fundamental limits to human intuition may always exist regarding high-dimensional problems.

Source: Contribution by OECD's Alistair Nolan (2017)

3.5. Surveillance and digital security

68. AI is already broadly used for surveillance and for digital security applications such as network security, anomaly detection and threat detection (OECD, 2017). At the same time, malicious use of AI, such as brute-force attacks against security loopholes is expected to increase with AI expanding existing threats, introducing new threats, and changing threat character.

69. This section is expected to draw on the February 2018 “Malicious AI Report” published by a group of AI researchers (https://maliciousaireport.com) and on the outcomes of the OECD Workshop on Digital Security and Resilience in Critical Infrastructure and Essential Services of 15-16 February 2018.[23]

Box 3.2. Surveillance with ‘smart’ cameras

The French CEA, in partnership with Thales, uses deep learning to automatically analyse and interpret videos for security applications. A Violent Event Detection module automatically detects violent interactions such as a fight, aggression, or an altercation captured by CCTV cameras and alerts operators in real-time.[24] Another module helps to locate the perpetrator(s) on the camera network. These applications are being evaluated by French public transportation bodies RATP and SNCF in the Châtelet-Les Halles and Gare du Nord subway stations, two of Paris' busiest train stations. The city of Toulouse, France, also uses smart cameras to signal unusual behavior and spot abandoned luggage in public places. Similar projects are being trialed in Berlin, Rotterdam and Shanghai.

Source: Demonstrations and information provided to the OECD by CEA Tech and Thales (2018).

3.6. Public services

70. Digital government services leverage AI to facilitate natural interaction with citizens. Virtual legislative assistants synthesise citizen feedback for lawmakers. AI-powered fraud detection systems monitor government expenditures in real-time. Smart cities manage urban green areas. [25] [26]

71. More broadly, benefits of AI in public service applications include its potential to improve public sector efficiency and responsiveness, services' quality and citizens' access. Policy questions include the treatment of data with public interest ramifications (*e.g.* transportation, traffic, energy consumption) owned by private companies, privacy and digital government. This section is expected to draw on work being conducted by the OECD Public Governance Committee (PGC) under the auspices of the E-Leaders group.[27]

3.7. Augmented and virtual reality (AR/VR)

72. Companies are using AI technology and high-level visual recognition tasks such as image classification and object detection to develop augmented reality (AR) and virtual reality (VR) hardware and software. Benefits include immersive experiences to training and education, helping people with disabilities, and entertainment. Policy questions include addiction and health issues.

3.8. Criminal justice

73. Box 3.3 reviews how AI is being used in the judicial system. AI holds the potential to improve access to justice and advance effective, speedy and impartial adjudication of justice. But concerns are raised about AI systems' potential challenge to citizen participation, transparency, embedded bias, dignity, privacy and liberty.

Box 3.3. AI in criminal justice [preliminary]

AI is increasingly used as a support tool for judges and police around the world. Although many applications are still experimental, a number of advanced products are in use, with influence on justice provision and law enforcement. The section is expected to draw on work that is being conducted by the Berkman Klein Center (BKC) on AI in criminal justice to develop a public database of risk assessment tools in use across the United States.

Assessing risk of recidivism

Several courts in the United States use AI-based tools to assess risk of recidivism in presentencing investigation reports and such use was approved by the US Supreme Court (Wisconsin v. Loomis, 2017). Popular tools include Correctional Offender Management and Profiling for Alternative Sanctions (COMPAS), Public Safety Assessment (PSA), and Level of Service Inventory (LSI-R). COMPAS, developed by Northpointe, assesses risk using five main factors: criminal involvement, relationship and lifestyle, personality/behavior and family and social exclusion. LSI-R, created by Canada-based Multi-Health Systems, uses data including criminal history and personality patterns. PSA, developed by the Laura and John Arnold Foundation, uses narrower data on the offender's age and criminal history.

In the United Kingdom, Durham Constabulary has developed the Harm Assessment Risk Tool (HART) to evaluate the risk of convicts reoffending based on a person's past offending history, age, postcode and other background characteristics through algorithms that then classify them as a low, medium or high risk.

Predictive crime mapping

Police forces also increasingly use predictive algorithmic tools to map when and where crime is likely to occur based on past crime data. Initiatives are being trialed in cities including Manchester, Durham, Bogota, London, Madrid, Copenhagen and Singapore.

In the United Kingdom, the Greater Manchester Police developed a predictive crime mapping system in 2012 and since 2013 the Kent Police has been using a system called PredPol. These systems to estimate the likelihood of crime in particular locations during a window of time are based on an algorithm that was originally used to predict earthquakes.

Based on predictions, the Data-Pop Alliance in Colombia uses crime and transportation data to predict criminal hotspots in Bogota. Polices forces are re-deployed to places and times where risk of crime is higher.

Algorithmic tools and reproducing biases

Concerns have been raised about the societal and legal ramifications of using algorithmic tools in policing, notably in terms of reproducing biases. Concepts of 'experimental' proportionality have been proposed to allow unproven algorithms to be used in the public sector in controlled and time-limited ways (West, 2013). One of the approaches to improve algorithmic transparency is a framework called 'ALGO-CARE' to ensure that police using algorithmic risk assessment tools consider key legal and practical elements.

Source: draws on contribution by James Haillot-O'Connor, Institut des Hautes Etudes sur la Justice (2018).

4. PUBLIC POLICY ISSUES [preliminary, to be finalised in Phase 2]

4.1. Introduction

74. This section will focus on a few broad types of policy challenges raised by AI in the areas of: human rights and fundamental values; employment and skills; safety, responsibility and liability; transparency and accountability; access to data and personal data protection; fairness and non-discrimination; and open and inclusive development and diffusion of AI.

75. The section builds on the presentations given at the OECD October 2017 Conference "*AI: Intelligent Machines, Smart Policies*"[28] and is expected to build on ongoing work streams in the Committee on Consumer Policy (CCP) and its working party on product safety (WPCPS); the Employment, Labour and Social Affairs Committee, the Education Policy Committee, and the CDEP Working Party on Security and Privacy in the Digital Economy (SPDE).

4.2. Human rights and fundamental values [to be developed in Phase 2]

76. This section will describe the opportunities and challenges that AI systems provide in the areas of human rights and fundamental values, which are a significant focus of AI policy discussions.

4.3. Employment and skills [preliminary, to be finalised in Phase 2]

77. Recent OECD measurement work has found that AI currently has a literacy level as good as or better than 89 percent of adults in OECD countries (Elliot, 2017).[29] AI matches or exceeds human performance in a growing number of domains. In addition, it does not need rest, processes data beyond human-scale, has more information than humans and will be able to look further into the future. AI excels at narrow tasks and can reason over vast amounts of data, turning video, speech, and sounds into text with image recognition. And the tasks that AI and robots cannot do are shrinking rapidly.

78. Although no one can say yet exactly how far AI-enabled machines will go in augmenting or replacing humans in the workplace, the most recent OECD estimates indicate that, based on existing technologies, 14% of jobs in OECD countries are at high risk of automation and that many other jobs will change significantly (Nedelkoska and Quintini, 2018; OECD, 2018). People in lower-skilled occupations involving significant amounts of repetition and youth are thought to be at highest risk of losing their jobs and automation is thought to be pressuring wages downwards. Some also suggest a link between AI and job polarisation, populism and increasing inequality.[30]

79. Some argue that demand in new types of jobs such as data scientists is unlikely to provide jobs for large numbers of people. In addition, displaced routine workers may not be able to find a job requiring equivalent skills since many sectors could be disrupted

simultaneously by AI. For example, the International Transportation Forum report on driverless trucks estimated that vehicle automation would reduce the costs of trucking so much (*e.g.*, by 30 percent) that a rapid deployment could rapidly drive out of business any non-adopters and estimated that 3.4 to 4.4 million truck drivers could lose their jobs over a fairly short period of time. [31]

80. At the same time, AI complements and requires people. AI is still far from acquiring the language skills needed for advanced intellectual activities. Posing dilemmas[32], interpreting situations[33] or extracting meaning from text requires people with qualities such as judgment, fairness and empathy. In science, for example, AI can complement humans in charge of the conceptual thinking necessary to build the research framework and to set the context for specific experiments. Robots such as those of Softbank are designed to assist, complement and obey humans.[34]

81. Many jobs remain outside the reach of AI, at least for now. AI can automate many repetitive tasks -- some 40 percent of tasks according to MGI.[35] But AI cannot readily automate entire jobs that consist of many tasks. AI will also make goods and services less expensive to produce and improve their quality, thereby increasing demand for them and, hence, increasing the demand for labour (Bessen, 2015). As with previous technological disruptions, AI is expected to generate whole new types of activities. In addition, AI could help improve health and safety at work and help automate routine, repetitive tasks, allowing more interesting and flexible work and possibly better work-life balance.[36]

82. Significant adjustments may be needed as people discover how best to participate in society and take advantage of their humanity, individuality and creativity.[37] Some think it is likely that humans will increasingly focus on work to improve each other's lives, such as childcare and care for the terminally ill[38] and that human creativity and ingenuity can leverage increasingly powerful computation, data and algorithm resources to create new tasks and directions that require human creativity.[39]

83. But there is also a sense that labour policy needs to respond to the new workplace reality, to focus on job creation and how AI will change tasks in the workplace. Labour Unions emphasise that policies should not place sole responsibility on displaced workers to deal with unemployment, job changes and wage reduction, especially for people at lower skills level. These policies should also facilitate job transitions and plan appropriate financing and governance.[40]

84. As the jobs change, so must education and training. Education policy is expected to require adjustments to expand continuous education, training and skills development. More AI theory and practice will be needed at university, as well as programmes on AI ethics. In science, there is a call for higher education to focus more on conceptual thinking, which machines cannot do and on rethinking scientific dissemination to ensure reproducibility.[41]

85. In addition, new tools and incentives to promote adult skills and skills policies combined with social protection and social dialogue may be needed. Furthermore, a world with robots and AI and potentially less human labour calls for sustainable social security systems.[42] It should be noted that AI also promises to help to deliver personalised education and tutoring tools that can provide new skills and quality education at scale. [43]

86. There are concerns about a shortage of skilled workers in machine learning, especially for SMEs and public universities and research centres that compete for talent with dominant firms but cannot provide the quantity of data, computation power and attractive salaries available at large companies. There are related concerns about the flow

of researchers and engineers from Europe to other regions and from the public sector to the private sector.

87. This section draws on work by the OECD ELSA Committee and could draw on the new CERI project following up on the report *Computers and the Future of Skills Demand* (Elliot, 2017).[44]

4.4. Safety, responsibility and liability [preliminary, to be finalised in Phase 2]

4.4.1. Accountability and responsibility

88. The diffusion of AI-embedded systems also raises important responsibility, liability and product safety considerations. Alongside dramatic benefits in areas such as safety and convenience, AI-embedded products raise a number of practical and legal issues. A range of new products and services learn from their environment throughout their life cycle based on consumer behaviour and other data and their functionality and performance can be customised and improved after purchase. In addition, consumer products are increasingly 'autonomous' or semi-autonomous, meaning they make and execute decisions with no or little human input.

89. The range of autonomous products is expanding rapidly, from robotics and driverless cars to everyday life consumer products and services, such as smart appliances and smart home security systems. Different types of AI applications are expected to call for different policy, legal and regulatory responses.[45] But in broad terms, learning and autonomous systems that make decisions call for reconsidering: *i)* who should be liable and to what extent for harm caused by an AI system, based on which parties can contribute to the safety of autonomous machines, *e.g.* users, product and sensor manufacturers, software producers, designers, infrastructure providers, data analytics companies; *ii)* what liability principle(s) should be used, *e.g.* strict liability, fault-based liability and the role of insurance. The opacity of some AI algorithms compound the issue of liability; *iii)* how can the law be enforced, what is a 'defect' in an AI product, what is the burden of proof and what remedies are available.

90. For example, the European Union's Product Liability Directive (Directive 85/374/EEC) of 1985 establishes the principle of 'liability without fault' or 'strict liability' according to which, if a defective product causes damage to a consumer, the producer is liable even without negligence or fault. The EC is reviewing this Directive. While preliminary conclusions find the model to be broadly appropriate[46], current and foreseeable AI technologies do impact the concepts of "product", "safety", "defect", and "damage" and make the burden of proof more difficult.

91. In the automotive sector, current fault-based liability schemes need to be rethought for driverless vehicles. There are calls for strict liability to be placed on manufacturers of autonomous cars, based on the controllability of risk and on the fact that a mere passenger of a driverless car cannot be at fault or have breached a duty of care. Legal experts put forward that even a "registered keeper" concept would not work because the keeper must be able to control the risk.[47] They suggest that an insurance system could cover the risk of damage by autonomous systems by classifying registered autonomous machines based on risk assessments.

4.4.2. Product safety and security

92. AI products also impact product safety (OECD, 2017). On the one hand, the safety benefits of AI products are significant and a major driver of adoption. But safety standards tend to regulate "finished" hardware products rather than software.[48] This section will also explain the larger issues around ensuring that the goals of AI systems are aligned with those of humans.

93. In addition, issues of safety, liability and security of AI are interlinked. This section will explain how digital security can affect product safety if connected products such as driverless cars are not sufficiently secure and for instance hackers can take control of them and change settings at a distance.

4.5. Transparency and accountability [to be developed in Phase 2]

94. A key focus of policy discussions and of technical research on AI is on transparency in AI decision-making to ensure fairness and accountability, particularly for AI-powered decisions that impact people's lives. Transparency in AI systems is generally described as allowing people to understand how an AI system is developed, trained and deployed and the factors that impact a specific decision or recommendation, rather than sharing underlying code or data. [49]

95. There is consensus that the need for explanations and for a human-machine interrogation process is particularly acute in high-stakes applications like criminal justice, driverless vehicles, personal finance and healthcare. For instance, people need to understand the decision-making behind whether a driverless car faced with an accident chooses to hit a bicyclist or hit a pedestrian or how AI was used in deciding whether to hire a specific job applicant. Articles 13-15 of the European Union's (EU's) new General Data Protection Regulation (GDPR) mandate that data subjects receive meaningful information about the logic involved, the significance and the envisaged consequences of automated decision-making systems and includes, in Article 22, the "right not to be subject to automated decision making".

96. There is discussion on differentiating applications or decisions that could rely on machines only and those that should require a human to be "in the loop". Some emphasise that AI-based scores should not be the sole factor used in decisions such as a sentencing that impact people's lives. For example, the GDPR stipulates that a human must be "in-the-loop" if a decision has a significant impact on people's life. *e.g.* if AI used to make credit determinations, grant educational opportunities, job screening, or to sentence criminals. [50]

97. Beyond high stakes decisions, understanding AI reasoning processes is viewed as important for the technology to become commonly accepted and useful. This section is expected to draw on research by a group of computer scientists, cognitive scientists, and legal scholars at Harvard University in the Berkman Klein Center *Working Group on Explanation and the Law* who identify trade-offs with each of the three core tools to provide AI accountability: *i)* explanation, *ii)* empirical evidence, and *iii)* theoretical guarantees (Table 4.1).

98. Explanation focuses on the factors that impact specific decisions of an AI system in order to prevent errors, increase trust, and verify whether criteria were used appropriately or inappropriately in case of a dispute (Doshi-Velez, 2017). However, designing explainable AI systems is difficult. In addition to technical difficulties and

resource requirements, concerns raised include the risk of stifling innovation, revealing trade secrets, or reducing performance measures such as accuracy.

99. Experts note that machine learning algorithms that are based on complicated neural network, or genetic algorithms produced by directed evolution may be extremely difficult to explain and to understand, while machine learning based on decision trees or Bayesian networks may be much simpler to understand (Hastie, Tibshirani, and Friedman 2001).

100. Machine learning systems are used to make sense of very large volumes of data in ways "beyond human scale", such as with AlphaGo, typically when tasks are not well defined.[51] Making their explanations understandable to humans might require a reduction in performance because it might require rejecting a solution that cannot be reduced to a human-understandable set of factors. Sensitivity analysis of critical variables could be provided. The permissible error rate would likely vary depending on the application[52] and the level of uncertainty of an AI system could be communicated to human decision-makers. For example, the permissible error rate for a translation tool may not be acceptable for autonomous driving or medical examinations.

101. This section will further detail the state of reflections and research on transparency and explainability in Phase 2, including research by Doshi-Velez (2017) that finds that an explanation should be able to answer the following questions about a particular decision: *i)* what main factor(s) went into a decision (ideally ordered by significance); *ii)* which specific factor(s) determined the outcome, (*i.e.* changing the factor(s) would change the outcome); and; *iii)* which factor(s) caused a different outcome in another decision.

Table 4.1. Considerations for Different Approaches to AI Accountability

Approach	Description	Well-suited Contexts	Poorly-suited Contexts
Theoretical Guarantees	In some situations, it is possible to give theoretical guarantees about a system backed by proof - explanations or evidence are not required. For example, a system could be designed to provably follow agreed-upon processes for voting and vote counting.	Situations in which both the problem and the solution can be fully formalised (perfect accountability, for such cases). Such cleanly specified contexts often do not hold in real-world settings.	Any situation that cannot be sufficiently formalised (most cases)
Statistical evidence	Empirical evidence measures a system's overall performance, demonstrating the value or harm of the system, without providing an explanation for specific decisions. For example, an autonomous aircraft landing system may have fewer safety incidents than human pilots, or a clinical diagnostic support tool may reduce mortality. Questions of bias or discrimination can be ascertained statistically: for example, a loan approval system might demonstrate its bias by approving more loans for men than women when other factors are controlled for.	Problems in which outcomes can be completely formalised, and a strict liability view is taken; Problems for which it is acceptable to wait to see negative outcomes happen to measure them. Certain types of subtle errors or discrimination may only be visible in aggregate.	Situations where the objective cannot be fully formalised in advance Statistical evidence cannot be used to assign blame or innocence surrounding a particular decision.
Explanation		Problems that are incompletely specified, where the objectives are not clear and inputs might be erroneous	Situations in which other forms of accountability are not possible

Source: based on Doshi-Velez, 2017

102. Over time, technical advances are expected to help to meet explanation challenges. Promising research is underway on the interpretability, accuracy and balances of algorithms in groups such as Fairness, Accountability, and Transparency in Machine Learning (FATML) that regard AI systems as part of the solution. Technical research in also ongoing in the area of driverless cars is exploring techniques to enable autonomous systems to explain themselves by generating representations of the relevant antecedents of significant events in the course of driving.

103. The US Defense Advanced Research Projects Agency (DARPA) is overseeing a program on "Explainable Artificial Intelligence" that funds academic and industry projects to provide a rationale for machine-learning systems' outputs. Under one method led by Professor Carlos Guestrin at the University of Washington, a computer automatically finds a few examples from a data set and highlights examples in a short explanation. For example, a system designed to classify e-mail messages from terrorists could highlight certain keywords found in a message, based on the millions of messages in training and decision-making data. For image recognition systems to hint at their reasoning, they can highlight the parts of an image that were most significant. One drawback to this approach is that the explanations provided are simplified, meaning some important information could be lost.

4.6. Access to data and personal data protection [to be developed in Phase 2]

4.6.1. Enhanced Access to data

104. Current machine learning technologies require curated and accurate data to enable companies, research institutes and the public sector to create innovative products and services. Access to data can accelerate or slow down AI development. For example, the significant innovation taking place in the satellite sector is attributed to the open data policies for satellite data of public entities like NASA or Copernicus.[53] Data protection, data regulation, and the data economy have legal, cultural, and technical ramifications. This section is expected to draw on ongoing work by the CDEP's Working Party on Security and Privacy in the Digital Economy (SPDE) on Enhanced Access to Data that is looking at policies and practices to improve access to data while protecting legitimate interests, including personal data.

4.6.2. Privacy and data protection

105. This section is expected to draw on ongoing work by the OECD's Working Party on Security and Privacy in the Digital Economy (SPDE). Some of the themes that the section is expected to discuss are included below.

106. Organisations leverage personal data profiles and AI to micro-target audiences with advertising and information. There is broad agreement that profiling in the AI context requires attention and scrutiny and a societal debate to determine the boundaries of acceptable versus non acceptable profiling.[54] Issues of privacy have been linked to broader sovereignty issue and to risks to self-determination, democracy and of agency of individuals over their data.[55]

107. Companies reported that the exponential rate at which AI gains access to, analyses and uses data is difficult to reconcile with traditional data protection principles on purpose specification, data minimisation and use limitation (e.g. OECD, 2013). They also report difficulty to benefit from data-driven AI technologies and business models while abiding by traditional data protection principles (Cellarius, 2018). Algorithmic

correlation weakens the distinction between personal data and other data, and non-personal data can increasingly be used to make inferences on and possibly re-identify individuals.[56] For instance, geolocation data from sensors might in some context be considered as personal data.

4.7. Fairness and non-discrimination [to be developed in Phase 2]

108. This section will detail concerns that machine learning algorithms reproduce the biases implicit in the training data they used, for instance, racial biases and stereotyped associations. This section will also provide an overview of approaches proposed to mitigate discrimination in AI, including awareness building, organisational diversity policies and practices, standards, technical solutions to detect and correct algorithmic bias and self-regulatory or regulatory approaches. For example in predictive policing systems, some propose algorithmic impact assessments or statements that would require police departments to evaluate the efficacy and potential discriminatory effects of all available choices for predictive policing technologies (Selbst, 2017).

4.8. Open and inclusive development and diffusion of AI [to be developed in Phase 2]

109. This section will describe other policy considerations highlighted at OECD conferences on AI in 2016 and 2017, including:

- The concentration of technology and financial resources in a few companies and nations flagged as an important issue in conferences on AI that the OECD held in 2016 and in 2017.
- Concerns about AI exacerbating inequality or increasing the divide between the haves and have-nots and between developed and developing countries.
- The role of policies to enable SMEs and other entities to access the technology, data and skills to navigate the AI transition, adopt and benefit from AI. For example, national AI strategies promote investments in promising vertical sectors where SMEs may not be able to invest without incentives or assistance.
- Another important theme is openness and common norms for machine learning, including among private sector actors. For example, Google shares curated training datasets (e.g. photos and video, speech commands, online discussion, audio effects, and crowdsourced drawings); tools (TensorFlow; FACETS) and training tools in the public domain to diffuse AI.

5. THE AI POLICY LANDSCAPE

110. AI is at the top of policy agendas for governmental institutions at both national and international levels. National government initiatives focus on using AI for productivity and competitiveness in countries including Canada, China, Estonia, Finland, France, Japan, Korea, the United Kingdom, and the United States. All strategies aim to increase AI researchers and skilled graduates; to strengthen national AI research capacity and translate AI research into public and private-sector applications. In considering the economic, social, ethical, policy and legal implications of AI advances, the national initiatives reflect differences in national cultures, legal systems, country size and level of AI adoption. This section also examines recent developments in regulations and policies related to AI.

111. AI is also a priority at the international level, such as at the G7, G20 and EU levels. The European Commission emphasises AI-driven efficiency and flexibility, interaction and cooperation, productivity, competitiveness and growth, and quality of citizens' life. Following the G7 ICT Ministers' Meeting in Japan in April 2016, the G7 ICT and Industry Ministers Meeting in Turin, Italy in September 2017 shared a vision of "human centric" AI and agreed to lead international cooperation and multi-stakeholder dialogue on AI, supported by the OECD.[57] Ongoing G20 and B20 attention to AI can also be underlined.[58]

112. Stakeholder groups are actively engaged in discussions on how to steer AI development and deployment to serve all of society. The Institute for Electrical and Electronics Engineers (IEEE) Standards Association launched its "Global Initiative on Ethics of Autonomous and Intelligent Systems" launched in April 2016 and recently published the version 2 of its Ethically Aligned Design (EAD) principles. The "Partnership on Artificial Intelligence to Benefit People and Society" started in September 2016 and plans to develop principles. The "Asilomar AI Principles" are a set of research, ethics and values, and longer-term issues and principles for the safe and socially beneficial development of AI in the near and longer term. The "AI Initiative" civic debate aims to bring together experts, practitioners and citizens globally to build common understanding of concepts such as AI explainability.

5.1. Governmental entities

5.1.1. G7

113. At the April 2016 G7 ICT Ministerial Meeting of Takamatsu (Japan), the Japanese Minister of Internal Affairs and Communications proposed that G7 countries lead international discussions on a non-binding international framework for AI development, building on the set of AI R&D Principles developed by her Ministry. The Japanese Ministry of Internal Affairs and Communications (MIC) then created an advisory group of experts (the 'Conference toward AI Network Society') in October 2016 to develop 'AI R&D Guidelines' for international discussion.

114. The G7 ICT and Industry Ministerial held in Turin in September 2017 under the Italian presidency issued a Ministerial Declaration in which G7 countries acknowledged the tremendous potential benefits of AI but also its uncertain impact on society and economy and agreed to take a 'human centric' approach to AI. They committed to *i)* gain understanding of the cultural, ethical, regulatory, and legal impact of AI, *ii)* explore both the positive and controversial impact of AI, notably on growth, job creation, accountability, privacy and security, *iii)* pursue a multi-stakeholder approach to address policy and regulatory issues, and *iv)* work towards common understanding of how to benefit from AI for an equitable society, while underlining that regulation must not hinder the development of technology and industry.

115. Under Canada's 2018 G7 Presidency, G7 Innovation Ministers convened in Montréal in March 2018 expressed a vision of human-centric AI that could help G7 countries to stimulate inclusive and sustainable growth and remove barriers to labour force participation. Ministers noted the importance of government policy in "stimulating innovation through investing in collaborative innovation ecosystems; improving access to capital and adoption of technology for SMEs; supporting significant investments in R&D; enabling firms to tap into global talent pools; streamlining government programs; developing online platforms to support entrepreneurship; using government procurement to foster Micro, Small and Medium Enterprise (MSME) innovation; refocusing investment in science, research and technology; promoting cyber-resiliency in value chains (particularly among MSMEs); and, especially, labour force training and skills development".

116. To advance their shared understanding of how to seize the opportunities presented by AI, G7 Innovation Ministers decided to convene a multi-stakeholder conference on AI, to be hosted by Canada in the fall of 2018. The conference will discuss future economic, legal, social, and ethical issues relating to the development and deployment of AI and how to harness its potential to break down barriers to labour force participation. In 2019, France will hold the G7 Presidency.

5.1.2. United Nations

117. In September 2017, the United Nations Interregional Crime and Justice Research Institute (UNICRI) signed the Host Country Agreement to open a Centre on Artificial Intelligence and Robotics within the United Nations system in The Hague, The Netherlands. The goal of the Centre is to serve as an international resource on matters related to AI and robotics.[59]

118. The International Telecommunications Union (ITU) is spearheading an initiative called "AI for Good" alongside over 25 sister United Nations agencies, in partnership with the XPRIZE Foundation and the Association for Computing Machinery (ACM). Following a first "AI for Good" Summit in June 2017 the ITU will hold a second Summit in Geneva in May 2018.[60]

5.1.3. European institutions: EC, CoE, EESC

119. On 25 April 2018, the European Commission (EC) issued a Communication on AI in Europe that aims to ensure that Europe is competitive in the AI landscape, that no one is left behind in the digital transformation and that new technologies are based on values. The EC's approach is three-pronged: increasing investments in the development of AI in Europe; preparing societies and economies for socio-economic changes brought about by AI and ensuring an appropriate legal and ethical framework for AI by setting up

a European AI Alliance and developing ethical guidelines (Table 5.1). The EC also plans to issue a Communication on the future of connected and automated mobility in Europe and a Communication on the future research and innovation ambitions for Europe. AI is planned to be a key element of these initiatives.

Table 5.1. EC Communication - 'Artificial Intelligence for Europe'

Goal	EU actions
1) Boost the EU's technological and industrial capacity and AI uptake	Funding fundamental research in AI and bringing more innovations to market (European Innovation Council pilot); supporting AI research excellence centres by Members; supporting AI development and uptake notably by SMEs (through a thematic platform in the European Fund for Strategic Investments or EFSI); funding innovative start-ups (with the €2.1 billion VentureEU Pan-European Venture Capital Fund-of-Funds and facilitating experimentation (with partly EC-funded Digital Innovation Hubs) and industrial data platforms offering high quality datasets
2) Prepare for socio-economic changes brought about by AI	Investments in digital skills by the European Social Fund 2014-2020 (€2.3 billion) and support from the private sector; EU funding as part of the Blueprint for Sectoral cooperation on skills in sectors from automotive to green tech (€50 million); and national digital skills strategies
3) Ensure an appropriate ethical and legal framework for AI	Clarifying the existing Directives on Products Liability for AI-related safety and liability; based on the Union's values and in line with the Charter of Fundamental Rights of the EU; drafting AI ethics guidelines to address issues related to the future of work, fairness, safety, social inclusion, algorithmic transparency, and impact on fundamental rights -- privacy, dignity, consumer protection and non-discrimination, following stakeholder consultation under the European AI Alliance to be set-up by July 2018.

Source: EC (2018)

120. Early 2018, the **Council of Europe** formed a committee composed of government and independent experts to explore the impacts of algorithmic decision-making processes and future technologies including AI on human rights. The "Committee of experts on human rights dimensions of automated data processing and different forms of artificial intelligence", that is under the supervision of the Steering Committee on Media and Information Society, held its first meeting on 6-7 March 2018 in Strasbourg.[61] The group is working towards an instrument on human rights and automatic data processing techniques.

121. In May 2017 the **European Economic and Social Committee** (EESC) adopted an opinion on the societal impact of AI. The opinion called on EU stakeholders to ensure that AI development, deployment and use work for society and social well-being. The EESC said humans should keep control over when and how AI is used in daily lives and identified 12 areas where AI raises societal concerns, including ethics, safety, transparency, privacy, standards, labour, education, access, laws and regulations, governance, democracy, but also warfare and superintelligence. The opinion called for pan-European standards for AI ethics, adapted labour strategies, and a European AI infrastructure with open-source learning environments (Muller, 2017). An EESC Permanent Study group on AI will be set up in April 2019.

5.1.4. Canada

122. Canada is positioning itself as an AI leader notably with the Pan-Canadian AI Strategy launched in March 2017. The strategy is led by the nonprofit Canadian Institute for Advanced Research (CIFAR) and backed with government funding of USD 100 million (CAD 125 million) over the next 5 years for programs to expand Canada's human capital, support AI research in Canada, and translate AI research into public and private-sector applications. The objectives of the Pan-Canadian AI Strategy are:

1. *Goal 1*. To increase AI researchers and skilled graduates in Canada;
2. *Goal 2*. To establish interconnected nodes of scientific excellence in Canada's three major AI institutes: in Edmonton Amii (Alberta Machine Intelligence Institute), Montreal MILA (Montreal Institute for Learning Algorithms), and Toronto Vector (Vector Institute for Artificial Intelligence);
3. *Goal 3*. To develop a global program on AI in Society and global thought leadership on the economic, social, ethical, policy and legal implications of advances in AI
4. *Goal 4*. To support a national research community on artificial intelligence.

123. In addition to this federal funding, the Quebec government is allocating USD 80 million (CAD 100 million) to the AI community in Montreal, Ontario USD 40 million (CAD 50 million) for the Vector Institute for Artificial Intelligence. In 2016, the Canada First Research Excellence Fund allocated USD 75 million (CAD 93.6 million) to three universities for cutting-edge research in deep learning; the Université de Montréal, Polytechnique Montréal and HEC Montréal. Facebook and other dynamic private companies like ElementAI are active in Canada.

124. The Quebec government plans to create a world observatory on the social impacts of artificial intelligence (AI) and digital technologies (Fonds de recherche du Québec, 2018). A workshop late March 2018 began to consider the observatory's mandate and potential model, governance mode, funding and international co-operation, as well as sectors and issues of focus. The Québec government has tabled USD 3.7 million (CAD 5 million) in funding to facilitate the implementation the observatory.

5.1.5. China

125. In May 2016, the Chinese government published a three-year national AI plan formulated jointly by the National Development and Reform Commission (NDRC), the Ministry of Science and Technology (MOST), the Ministry of Industry and Information Technology (MIIT), and the Cyberspace Administration of China (CAC). China's 'Three-year Guidance for Internet Plus Artificial Intelligence Plan (2016-2018)' focuses on: *i)* enhancing AI hardware capacity, *ii)* strong platform ecosystems, *iii)* AI applications in important socioeconomic areas, and *iv)* AI's impact on society. In it the Chinese government envisioned creating a USD 15 billion market by 2018 by investing in research and supporting the development of the Chinese AI industry.

126. Mid 2017, China's State Council released the 'Guideline on AI Development', which provides China's long-term perspective on AI with industrial goals for each period: *i)* AI-driven economic growth in China by 2020, *ii)* major breakthroughs in basic theories by 2025 and in building an intelligent society, and *iii)* for China to be a global AI innovation center by 2030 and build up an AI industry of USD 150 billion (1 trillion RMB). The plan's implementation seems to be advancing throughout government and China has been developing leadership in AI with state support and private company dynamism. China's State Council set objectives for "new-generation information technology" as a strategic industry targeted to account for 15 percent of GDP by 2020.

127. In its 13th Five-Year Plan timeframe (2016-2020), China ambitions to transform itself into a science and technology leader, with 16 "Science and Technology Innovation 2030 Megaprojects", including 'AI 2.0'. The plan has provided impetus for action in the public sector (Kania, 2018). The plan asks companies to accelerate AI hardware and

software R&D, including in AI-based vision, voice and bio-metric recognition, man-machine interfaces and smart controls.

128. On January 18, 2018, China established a national AI standardisation group and a national AI expert advisory group. At the same time, the National Standardisation Management Committee Second Ministry of Industry, helped in editing by the China Electronic Standardisation Institute (CESI, a division under the Ministry of Industry and Information Technology), also released a white paper on AI standardisation (CESI, 2018).

129. It could be noted that Chinese private companies' efforts predate the more recent government support, with significant efforts and investments by Chinese companies such as Baidu, Alibaba and Tencent ("BAT"). Chinese industry has focused on applications development and data integration. The central government focuses on research in basic algorithms, open data, and conceptual work. City governments focus on the use of applications and open data at a city level.

5.1.6. Estonia

130. Estonia is planning the next step of its e-governance system powered by AI to save costs and improve efficiency and experimenting with e-healthcare and situational awareness. Estonia's focus is on improving lives and cities and supporting human values. On the enforcement side, Estonia focuses on core values of ethics, liability, integrity, and accountability, rather than on rapidly evolving technology and building an enforcement system based on blockchain that mitigates integrity and accountability risks, with a pilot project planned in 2018.

131. With 'StreetLEGAL' self-driving cars can be tested on Estonian public roads since March 2017. Estonia is also the first and only government discussing the legalisation of AI; *i.e.* giving representative rights, and responsibilities, to algorithms to buy and sell services on their owners' behalf. The government is considering four options and aims to have a bill ready by March-April 2019.

5.1.7. Finland

132. Finland aims to develop a safe and democratic society with AI, to provide the best public services in the world and for AI to bring new prosperity, growth, and productivity to citizens. The AI strategy "Finland's Age of Artificial Intelligence", published in October 2017, is a roadmap for Finland to leverage its educated population, advanced digitalisation and public sector data resources, while building international links in research and investment and encouraging private investments. Finland hopes to double its national economic growth by 2035 thanks to AI. Eight key actions for AI-enabled growth, productivity and well-being are: *i)* Enhancing companies' competitiveness , *ii)* Utilising data in all sectors, *iii)* Speeding up and simplifying AI adoption, *iv)* Ensuring top-level expertise, *v)* Making bold decisions and investments, *vi)* Building the world's best public services, *vii)* Establishing new cooperation models, and *viii)* Making Finland a trendsetter in the age of AI. The report highlights using AI to improve public services. For example, the Finnish Immigration Service uses the national customer service robot network called 'Aurora' to provide multi-lingual communication.

133. In February 2018, the government also created a funding entity for AI research and commercial projects. The entity will allocate EUR 200 million in grants and incentives to the private sector, including SMEs. Finland reports some 250 companies

working on AI development, for example in Finnish healthcare industry organisations, the healthcare system, by professionals and patients and associated in-depth reforms (Sivonen, 2017). The role of Finnish state-funded technical innovation agency (VTT) and funding agency for innovation (TEKES) will also be expanded.

5.1.8. France

134. French President Emmanuel Macron announced France's AI strategy on 29 March 2018 that allocates EUR 1.5 billion of public funding into artificial intelligence by 2022 to help France become an AI research and innovation leader. The measures are largely based on recommendations laid out in the report developed by MP Cédric Villani (Villani, 2018). The strategy calls for public research, education, building world-class research hubs linked to industry through public-private partnerships, and attracting foreign and French elite AI researchers working abroad. To develop the AI ecosystem in France, the strategy's approach is to 'upgrade' existing industries**.** Starting from applications in health, environment, transport and defence, the idea is to help bring AI into existing industry practices to renew existing industries. It proposes to prioritise access to data by creating "data commons" between private and public actors; adapting copyright law to facilitate data mining; and opening public sector data such as health to industry partners.

135. The strategy also outlines measures to begin planning for AI-induced disruptions, taking a firm stance on data transfers out of Europe (Thompson, 2018). It would create a central data agency with a team of about 30 advisory experts on AI applications across government. The ethical and philosophical boundaries articulated in the strategy include algorithm transparency as a core principle (for example, algorithms developed by the French government or with public funding will reportedly be open) and respect for privacy and other human rights "by design". The strategy also develops vocational training in professions threatened by AI. It calls for policy experimentation in the labour market and for dialogue on how to share AI-generated value added across the value chain. A French report on AI and work was also released late March (Benamou and Janin, 2018).

136. A number of international AI companies are already finding reasons to open up AI laboratories in France, e.g. Google DeepMind, Fujitsu, SAP, Samsung. The French AI strategy will be linked to the Joint European Disruption Initiative (JEDI), a European "DARPA" that will finance Moonshots / risky disruptive innovations (Fouquet, 2018).

5.1.9. Germany

137. In Germany, the Federal Ministry for Economic Affairs and Energy (BMWi) supports the development of innovative technologies, regulatory instruments and dialogue processes with regard to AI. BMWi provides public funding for flagship projects to deploy intelligent big data technologies and develop AI-based applications, in co-operation with other funding programs that target SMEs. BMWi plans to introduce 'regulatory test beds' to allow new technologies and business models to be tested within a limited area and limited period of time. Regulatory instruments can then be adjusted based on the results of these test beds.

138. In June 2017, the Federal Ministry of Transport and Digital Infrastructure developed ethical guidelines for self-driving cars. The guidelines, developed by the Ethics Commission of the Ministry, stipulate 15 rules for programmed decisions embedded in self-driving cars. The Commission considered ethical questions in depth, including whose

life to prioritise (known as the “Trolley problem”). The guidelines provide that self-driving cars should be programmed to consider all human lives as equal. If a choice is needed between people, self-driving cars should choose to hit whichever person would be hurt less, regardless of age, race, or gender. The Commission also makes clear that no obligation should be imposed on individuals to sacrifice themselves for others.

5.1.10. Japan

139. The Japanese Cabinet office established a ‘Strategic Council for AI Technology’ in April 2016 to promote AI technology R&D and business applications. The Council published an ‘Artificial Intelligence Technology Strategy’ in March 2017 that identified issues that Japan should address notably by: *i)* increasing investment in AI by both public and private sectors, *ii)* facilitating the use and access to data, *iii)* increasing the numbers of AI researchers and engineers. The Strategy also provided focused measures on strategic areas in which AI could bring significant benefits: productivity; health, medical care and welfare; mobility; and information security. The Council consolidates R&D capabilities in Japan and coordinates industrial policies with industry, academia and relevant government agencies. The Council began multi-stakeholder dialogue to facilitate the deployment and related businesses in spring 2018, including discussing the initiatives in academia, business and government agencies.

140. At the G7 ICT Ministerial meeting held in Japan in April 2016, Japan proposed the formulation of shared principles for AI research and development. The Japanese Ministry of Internal Affairs and Communications (MIC) published “AI R&D Guidelines” in July 2018, developed by a group of advisory experts (the ‘Conference toward AI Network Society’). The objectives of the guidelines are to achieve a human-centered society, to balance benefits and risks of AI networks and to ensure technological neutrality while avoiding placing excessive burden on developers. The guidelines consist of nine principles that researchers and developers of AI systems should pay attention to. In October 2017, the advisory group (‘Conference’) began to study the use of AI in society. Table 5.2**Error! Reference source not found.** provides the abstract of the guidelines.

Table 5.2. R&D Principles provided in the AI R&D Guidelines

Principle of:	*Developers should:*
I. Collaboration	Pay attention to the interconnectivity and interoperability of AI systems.
II. Transparency	Pay attention to the verifiability of inputs/outputs of AI systems and explainability of their decisions.
III. Controllability	Pay attention to the controllability of AI systems.
IV. Safety	Ensure that AI systems not harm the life, body, or property of users or third parties through actuators or other devices.
V. Security	Pay attention to the security of AI systems.
VI. Privacy	Take into consideration that AI systems will not infringe the privacy of users or third parties.
VII. Ethics	Respect human dignity and individual autonomy in R&D of AI systems.
VIII. User Assistance	Take into consideration that AI systems will support users and make it possible to give them opportunities for choice in appropriate manners.
IX. Accountability	Make efforts to fulfill their accountability to stakeholders including users of AI systems.

Source: Japanese MIC Conference toward AI Network Society (2017)

5.1.11. Korea

141. The Korean government published the “Intelligent Information Industry Development Strategy” in March 2016 and announced public investment of 940 million

USD (1 trillion KRW) by 2020 in the field of AI and related information technologies such as IoT and cloud computing. This strategy aimed to create a new intelligent information industry ecosystem and to encourage USD 2.3 billion (2.5 trillion KRW) of private investment by 2020. Under the strategy, the government planned to: *i)* launch AI development flagship projects, for example in the areas of language-visual-space-emotional intelligence technology, *ii)* strengthen AI-related workforce skills, and *iii)* promote the access and use of data by government, companies and research institutes.

142. In December 2016, the Korean government published the "Mid- to Long-Term Master Plan in Preparation for the Intelligence Information Society". The plan contains national policies to respond to the changes and challenges of the 4th Industrial Revolution. It contains a vision of a 'human-centric intelligent society' and aims to establish the foundations for world-class intelligent IT that can be applied across industries and be used to upgrade social policies. To implement the plan, the government is creating large-scale test beds to facilitate the development of new services and products, including better public services.

5.1.12. United Kingdom (U.K.)

143. The Digital Strategy published in March 2017 recognises AI as key to help grow the United Kingdom's digital economy (U.K. Government, 2017a). The strategy identifies AI as a key field to grow the United Kingdom's digital economy, and includes USD 22.3 million (17.3 million pounds) in funding for UK universities to develop AI and robotics technologies. The government has increased investment in AI R&D research and development by USD 6.6 billion (4.7 billion pounds) over the next 4 years, partly through its Industrial Strategy Challenge Fund.

144. In October 2017, the U.K. Government published an industry-led review on the U.K.'s AI industry in the UK. The report identifies the UK as an international centre of AI expertise, in part as a result of pioneering computer scientists such as Alan Turing. The U.K. government estimated that AI could add USD 814 billion to the domestic economy (U.K. Government, 2017). Existing AI tools used in the U.K. include an AI personal health guide (Your.MD), an AI chatbot developed for bank customers, an AI platform to help children learn and teachers provide personalised education programmes. The report provided 18 recommendations that include: *i)* improving access to data and data sharing by developing Data Trusts, *ii)* improving the AI skills supply through industry-sponsored Masters in AI, *iii)* maximising AI research by coordinating the demand for computing capacity for AI research among relevant institutions, *iv)* supporting the uptake of AI by establishing a "U.K. AI Council" and *v)* developing a framework to improve transparency and accountability of AI-driven decisions.

145. The UK Government published its industrial strategy in November 2017. The strategy identifies putting the U.K. at the forefront of the artificial intelligence and data revolution as one of the country's four Grand Challenges (U.K. Government 2017c).

5.1.13. United States

146. Building upon an inter-agency initiative and a series of public-outreach activities, the White House Office of Science and Technology Policy (OSTP) published a report on AI "Preparing for the Future of Artificial Intelligence" in October 2016, which reviewed existing and potential applications of AI and raised questions for society and public policy. This report also made recommendations for specific further actions by Federal agencies and other actors. This report was accompanied by a "National Artificial

Intelligence Research and Development Strategic Plan" published in October 2016, which identified priorities and established objectives and priorities for federally-funded AI research. The White House published a report on AI-driven automation (White House, 2016) in December 2016 that proposed policy responses in education, training, and safeguards for people to manage job transitions and unemployment.

147. The U.S. Congress has also been active, with Congressman John K. Delaney launching the bipartisan Artificial Intelligence (AI) Caucus in May 2017 co-chaired by Congressman Pete Olson. The caucus brings together experts from academia, government and the private sector to discuss the implications of AI technologies. Delaney introduced a bill entitled "Fundamentally Understanding The Usability and Realistic Evolution of Artificial Intelligence Act of 2017" (H.R. 4625, "FUTURE of Artificial Intelligence Act of 2017") in December 2017, which was also introduced in the Senate by Senator Maria Cantwell. The bill would create a federal advisory committee in the Department of Commerce with 19 voting members from research, industry, civil society and labour organisations. The committee would examine AI issues and make recommendations on AI investment, workforce, ethics and privacy issues, among others.

148. For autonomous motor vehicles, the House of Representatives passed the "Safely Ensuring Lives Future Deployment and Research In Vehicle Evolution Act" (H.R. 3388, "SELF DRIVE Act") in September 2017. In the Senate, the "American Vision for Safer Transportation through Advancement of Revolutionary Technologies Act" (S. 1885, "AV START Act") was approved in the Senate Committee on Commerce, Science, and Transportation in October 2017. Both bills would shift authority in the area of autonomous motor vehicles -- notably for developing safety standards including on cybersecurity -- from States to the National Highway Traffic Safety Administration (Canis 2017).

5.2. Private Stakeholder Initiatives

149. Several partnerships and initiatives have been formed to promote ethical AI and try to prevent adverse effects of AI. While many of these initiatives are multi-stakeholder in nature, this section describes a few of these initiatives, categorised based on whether stakeholders are mostly from: the technical community (including the IEEE and the Future of Life Institute); the private sector (including the PAI, ITI, Deepmind, Microsoft); labour (UNI); or academia (the 'AI Initiative'). This list is not exhaustive.

5.2.1. Technical community

150. The Institute for Electrical and Electronics Engineers (IEEE) Standards Association launched its "Global Initiative on Ethics of Autonomous and Intelligent Systems" in April 2016. The initiative aims to advance public discussion on the implementation of AI technologies and to define values and ethics that should be prioritised. The IEEE published version 2 of its Ethically Aligned Design (EAD) principles in December 2017, invited comments from the public and plans to publish the final version of the design guidelines in 2019 (Table 5.3).

Table 5.3. General Principles contained in IEEE’s Ethically Aligned Design Version 2

Principles	Objectives
Human Rights	Ensure autonomous and intelligent systems (AIS) do not infringe on internationally recognised human rights
Prioritising Well-being	Prioritise metrics of well-being in the design and use of AIS because traditional metrics of prosperity do not take into account the full effect of AI systems technologies on human well-being
Accountability	Ensure that designers and operators of AIS are responsible and accountable
Transparency	Ensure AIS operate in a transparent manner
AIS Technology Misuse and Awareness of It	Minimise the risks of misuse of AIS technology

Source: IEEE (2017)

151. The “Asilomar Principles” are a set of 23 principles for the safe and socially beneficial development of AI in the near and longer term that resulted from the Future Life Institute’s conference of January 2017.[62] The Asilomar conference extracted core principles from discussions, reflections and documents produced by the IEEE, academia and non-profit organisations.

152. The issues are grouped into: *i) research issues*, with a call for research funding for beneficial AI that include difficult questions in computer science; economics, law and social studies’; a constructive ‘science-policy link’; and a technical research culture of cooperation, trust and transparency; *ii) ethics and values*, with a call for AI systems’ design and operation to be safe and secure, transparent and accountable, protective of individuals’ liberty, privacy, human dignity, rights and cultural diversity, broad empowerment and shared benefits; and *iii) longer-term issues*, notably avoiding strong assumptions on the upper limits of future AI capabilities and planning carefully for the possible development of artificial general intelligence (AGI) (FLI, 2017). Table 5.4 provides the list of Asilomar AI Principles.

Table 5.4. Asilomar AI Principles (excerpt titles of principles)

	Research Issues	*Ethics and Values*	*Longer-term Issues*
Titles of Principles	- Research Goal - Research Funding - Research Funding - Science-Policy Link - Research Culture - Race Avoidance	- Safety - Failure Transparency - Judicial Transparency - Responsibility - Value Alignment - Human Values - Personal Privacy - Liberty and Privacy - Shared Benefit - Shared Prosperity - Human Control	- Capability Caution - Importance - Risks - Recursive Self-Improvement - Common Good

Source: FLI (2017)

153. The non-profit AI research company OpenAI was founded late 2015 and now employs 60 full-time researchers with the mission to “build safe AGI, and ensure AGI’s benefits are as widely and evenly distributed as possible”.

5.2.2. Private sector initiatives

154. In September 2016, Amazon, DeepMind/Google, Facebook, IBM and Microsoft launched the "Partnership on Artificial Intelligence to Benefit People and Society" (PAI) to advance public understanding of AI technologies and formulate best practices on its challenges and opportunities.[63] After PAI's creation, other companies, non-profit and academic organisations joined the partnership. PAI has become a multi-disciplinary stakeholder community with over 50 partners.

155. The Information Technology Industry Council (ITI) is a business association of technology companies based in Washington D.C. with more than 60 members. ITI published "AI Policy Principles" in October 2017 (Table 5.5). The principles identified industry's responsibility in areas including ensuring safety and controllability, robust and representative data, and interoperability in the deployment and use of AI. The principles also call for government support of AI research and for public-private partnerships, *inter alia* to facilitate STEM education and skills training.

Table 5.5. ITI AI Policy Principles

Responsibility: *Promoting Responsible Development and Use*	*Opportunity for Governments: Investing and Enabling the AI Ecosystem*	*Opportunity for Public-Private Partnerships: Promoting Lifespan Education and Diversity*
- Responsible Design and Deployment - Safety and Controllability - Robust and Representative Data - Interpretability - Liability of AI Systems Due to Autonomy	- Investment in AI Research and Development - Flexible Regulatory Approach - Promoting Innovation and the Security of the Internet - Cybersecurity and Privacy - Global Standards and Best Practices	- Democratising Access and Creating Equality of Opportunity - Science, Technology, Engineering and Math (STEM) Education - Workforce - Public Private Partnership

Source: ITI (2017)

156. Companies are taking action individually too. For example, Google-owned DeepMind also created a DeepMind Ethics & Society (DMES) unit in October 2017. The unit's goal is to help technologists understand the ethical implications of their work and help society decide how AI can be beneficial. The unit will also fund external research on algorithmic bias, the future of work, lethal autonomous weapons and more.

157. Microsoft's AI vision is to 'amplify human ingenuity with intelligent technology' and the company has launched projects to ensure inclusive and sustainable development. For example, 'Seeing AI' is a free mobile application to help the visually impaired: the application identifies situations and provides information orally. Microsoft is investing USD 2 million in qualified initiatives to tackle sustainability challenges such as biodiversity and climate change.

5.2.3. Labour organisation

158. UNI Global Union represents over 20 million workers from over 150 countries in skills and services sectors. A Future World of Work that empowers workers and provides decent work is a key UNI Global Union priorities. UNI Global Union has identified ten key principles for Ethical AI that ensures workers' rights to be realised by unions, shop stewards and global alliances, in collective agreements, global framework agreements and multinational alliances (Table 5.6).

Table 5.6. Top 10 Principles for Ethical Artificial Intelligence (UNI Global Union)

1. AI systems must be transparent:	Workers should have the right to demand transparency in the decisions and outcomes of AI systems as well as their underlying algorithms. They must also be consulted on AI systems' implementation, development and deployment.
2. AI systems must be equipped with an "ethical black box":	"Ethical black box" should not only contain relevant data to ensure system transparency and accountability, but also clear data and information on the ethical considerations built into the system.
3. AI must serve people and planet:	Codes of ethics for the development, application and use of AI are needed so that throughout their entire operational process, AI systems remain compatible and increase the principles of human dignity, integrity, freedom, privacy and cultural and gender diversity, as well as fundamental human rights.
4. Adopt a human-in-command approach:	The development of AI must be responsible, safe and useful, where machines maintain the legal status of tools, and legal persons retain control over, and responsibility for, these machines at all times.
5. Ensure a genderless, unbiased AI:	In the design and maintenance of AI and artificial systems (AS), it is vital that the system is controlled for negative or harmful human-bias, and that any bias–be it gender, race, sexual orientation, age–is identified and is not propagated by the system.
6. Share the benefits of AI systems:	The economic prosperity created by AI should be distributed broadly and equally, to benefit all of humanity. Global as well as national policies aimed at bridging the economic, technological and social digital divide are therefore necessary.
7. Secure a just transition and ensure support for fundamental freedoms and rights:	As AI systems develop and augmented realities are formed, workers and work tasks will be displaced. It is vital that policies are put in place that ensure a just transition to the digital reality, including specific governmental measures to help displaced workers find new employment.
8. Establish global governance mechanism:	Establish multi-stakeholder Decent Work and Ethical AI governance bodies on global and regional levels. The bodies should include AI designers, manufacturers, owners, developers, researchers, employers, lawyers, CSOs and trade unions.
9. Ban the attribution of responsibility to robots:	Robots should be designed and operated as far as is practicable to comply with existing laws, and fundamental rights and freedoms, including privacy.
10. Ban AI arms race:	Lethal autonomous weapons, including cyber warfare, should be banned. UNI Global Union calls for a global convention on ethical AI that will help address, and work to prevent, the unintended negative consequences of AI while accentuating its benefits to workers and society. We underline that humans and corporations are the responsible agents.

Source: Colclouh C. (2018)

5.2.4. Academia

159. "The AI Initiative" was created by the Future Society of Kennedy School of Government, Harvard University in 2015 to help shape the global AI policy framework. It launched a civic debate on "Governing the Rise of Artificial Intelligence" in September 2017. This is an online platform for multi-disciplinary discussion on AI among experts, practitioners, policy makers and citizens. The objective of the platform is to help understand the dynamics, benefits and risks of AI technology. The result of the civic debate will inform policy recommendations.

5.3. Existing sets of principles for AI

160. A number of initiatives have developed valuable sets of principles to guide AI development (Table 5.7), many of which focus on the technical communities who conduct research and development of AI systems. While many of these principles were developed in multi-stakeholder processes, they can be categorised broadly as echoing considerations from: a technical community angle (IEEE, JSAI, FATML, ACM, Asilomar); a private sector focus (PAI, ITI, Nadella); a government focus (MIC, COMEST, EPSRC); an academic focus (Montreal, Economou); and a labour focus (UNI). This section will be updated in Phase 2 to examine common elements of existing sets of principles and to include sets of principles developed by other groups.

5.4. Next steps for the OECD

Council Recommendation on Artificial Intelligence in Society

161. As discussed at the CDEP November 2017 meeting, the proposed next step is to develop a draft Council Recommendation on AI in society. The work will aim to promote a human-centric approach to AI and be conducted in collaboration with stakeholders and other OECD committees. It will draw on common themes that emerge from existing initiatives and sets of guidelines that have been developed by different stakeholders. It could include focus on: human values and rights; non-discrimination; awareness and control; access to data; privacy and control; safety and security; skills; transparency and explainability; accountability and responsibility; whole of society dialogue; and measurement.

Table 5.7. Selection of existing sets of guidelines for AI developed by stakeholders

Reference	Existing sets of guidelines for AI developed by stakeholders
ACM	Association for Computing Machinery US Public Policy Council (2017) “Statement on Algorithmic Transparency and Accountability” https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf
Asilomar	Future of Life Institute (FLI) (2017) “Asilomar AI Principles” https://futureoflife.org/ai-principles/
COMEST	World Commission on the Ethics of Scientific Knowledge and Technology (COMEST) (2017) “Report of COMEST on Robotics Ethics” http://unesdoc.unesco.org/images/0025/002539/253952E.pdf
Economou	Economou, N (2017) “A ‘principled’ artificial intelligence could improve justice” http://www.abajournal.com/legalrebels/article/a_principled_artificial_intelligence_could_improve_justice
EPSRC	Engineering and Physical Sciences Research Council (EPSRC) (2010) “Principles of robotics” https://epsrc.ukri.org/research/ourportfolio/themes/engineering/activities/principlesofrobotics/
FATML	Fairness, Accountability, and Transparency in Machine Learning (FATML) (2016) “Principles for Accountable Algorithms and a Social Impact Statement for Algorithms” https://www.fatml.org/resources/principles-for-accountable-algorithms
IEEE	Institute of Electrical and Electronics Engineers (IEEE) (2017), Global Initiative on Ethics of Autonomous and Intelligent Systems, “Ethically Aligned Design Version 2”, http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf
ITI	Information Technology Industry Council (ITI) (2017) “AI Policy Principles” https://www.itic.org/resources/AI-Policy-Principles-FullReport2.pdf
JSAI	The Japanese Society for Artificial Intelligence (JSAI) (2017) “The Japanese Society for Artificial Intelligence Ethical Guidelines” http://ai-elsi.org/wp-content/uploads/2017/05/JSAI-Ethical-Guidelines-1.pdf
MIC	Japanese Ministry of Internal Affairs and Communication (MIC), the Conference toward AI Network Society (2017) “Draft AI R&D Guidelines for International Discussions” (Tentative Translation) http://www.soumu.go.jp/main_content/000507517.pdf
Montreal	University of Montreal (2017), “The Montreal Declaration for a Responsible Development of Artificial Intelligence” https://www.montrealdeclaration-responsibleai.com/
Nadella	Nadella, S (2017) “The Partnership of the Future” http://www.slate.com/articles/technology/future_tense/2016/06/microsoft_ceo_satya_nadella_humans_and_a_i_can_work_together_to_solve_society.html
PAI	Partnership on AI to benefit people and society (2016) “TENETS” https://www.partnershiponai.org/tenets/
UNI	UNI Global Union (2017) “Top 10 Principles for ethical artificial intelligence” http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf

REFERENCES

Allen, Kate, (2015) "How a Toronto professor's research revolutionized artificial intelligence", The Toronto Star, 17 April 2015, https://www.thestar.com/news/world/2015/04/17/how-a-toronto-professors-research-revolutionized-artificial-intelligence.html

AI Initiative (2017) "Governing the rise of Artificial Intelligence – A Global Civic Debate" https://assembl-civic.bluenove.com/ai-consultation/home

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016). Machine bias: There's software used across the country to predict future criminals. and it's biased against blacks. ProPublica, May, 23. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

Association for Computing Machinery, US Public Policy Council (USACM) (2017) "Statement on Algorithmic Transparency and Accountability" https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

Benhamou, S. and Janin, L. (2018), Intelligence Artificielle et Travail, France Strategie, 28 mars 2018, http://www.strategie.gouv.fr/publications/intelligence-artificielle-travail

Bessen, J.E., 2015, How Computer Automation Affects Occupations: Technology, Jobs and Skills, Boston University School of Law, Law and Economics Research Paper No. 15 - 49. http://www.bu.edu/law/faculty/scholarship/workingpapers/2015.html.

Bloomberg Technology, Telecom and Internet Blog (2018) "U.S. Needs Sharper Focus on Artificial Intelligence Policy: Lawmaker" https://www.bna.com/us-needs-sharper-b57982089177/ (accessed on Apr 13 2018)

Bostrom, N. and Yudkowsky, E. (2011), "The Ethics of Artificial Intelligence", In Cambridge Handbook of Artificial Intelligence, edited by Keith Frankish and William Ramsey, New York: Cambridge University Press, https://intelligence.org/files/EthicsofAI.pdf

Canadian government (2018) "G7 ministerial meeting: Preparing for jobs of the future", 27-28 March 2018, https://g7.gc.ca/en/g7-presidency/themes/preparing-jobs-future/

Canis, B. (2017) "Issues in Autonomous Vehicle Deployment" https://www.hsdl.org/?view&did=804009

CB Insights (2017a), "The State of Artificial Intelligence", https://www.cbinsights.com/research/report/artificial-intelligence-trends/

CB Insights (2017b), "The 2016 AI Recap: Startups See Record High in Deals and Funding", https://www.cbinsights.com/research/artificial-intelligence-startup-funding/.

CB Insights (2018a), "Artificial Intelligence Trends To Watch In 2018"

CB Insights (2018b), “The Race For AI: Google, Intel, Apple In A Rush To Grab Artificial Intelligence Startups”, February 27, 2018, https://www.cbinsights.com/research/top-acquirers-ai-startups-ma-timeline/

Cellarius, M. (2017), Artificial intelligence and the right to informational self-determination, OECD Forum Network, https://www.oecd-forum.org/users/75927-mathias-cellarius/posts/28608-artificial-intelligence-and-the-right-to-informational-self-determination

Chinese Government, “AI Standardisation White Paper” (2018), released 28 January 2018, translation by Jeffrey Ding, researcher in the Future of Humanity’s Governance of AI Program. https://baijia.baidu.com/s?id=1589996219403096393

Chinese Government, State Council (2017), “Guideline on Next Generation AI Development Plan”, July, http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm - translated into English by: Rogier Creemers, Leiden Asia Centre; Graham Webster, Yale Law School Paul

Chinese Ministry of Industry and Information Technology (MIIT) (2016), “Three-Year Action Plan for Promoting Development of a New Generation Artificial Intelligence Industry (2018-2020)”, May 2016, http://www.miit.gov.cn/n1146290/n1146392/c4808445/content.html

Cockburn, I., Henderson, R. and S. Stern (2017) The Impact of Artificial Intelligence on Innovation, Paper prepared for the NBER Conference on Research Issues in Artificial Intelligence, Toronto, September 2017, NBER, http://www.nber.org/chapters/c14006.pdf.

Colclough, C. (2017) “Ethical artificial intelligence - 10 essential ingredients” OECD the Forum Network, January 24, 2018, https://www.oecd-forum.org/channels/722-digitalisation/posts/29527-10-principles-for-ethical-artificial-intelligence

Dalle, J., M. den Besten and C. Menon (2017), “Using Crunchbase for economic and managerial research”, OECD Science, Technology and Industry Working Papers, No. 2017/08, OECD Publishing, Paris, http://dx.doi.org/10.1787/6c418d60-en.

Doshi-Velez, Finale and Kortz, Mason and Budish, Ryan and Bavitz, Christopher and Gershman, Samuel J. and O'Brien, David and Shieber, Stuart and Waldo, Jim and Weinberger, David and Wood, Alexandra, “Accountability of AI Under the Law: The Role of Explanation” (November 3, 2017). Berkman Center Research Publication Forthcoming. Available at SSRN: https://ssrn.com/abstract=3064761 or http://dx.doi.org/10.2139/ssrn.3064761.

EC (2018), 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 on Artificial Intelligence for Europe, 25 April 2018, {SWD(2018) 137 final}, https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe

Economou, N (2017) “A ‘principled’ artificial intelligence could improve justice” http://www.abajournal.com/legalrebels/article/a_principled_artificial_intelligence_could_improve_justice

Elliott, S. (2017), Computers and the Future of Skill Demand, Educational Research and Innovation, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264284395-en.

Finnish Ministry of Economic Affairs and Employment (2017) “Finland’s Age of Artificial Intelligence” http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf?sequence=1&isAllowed=y

Fonds de recherche du Québec (2018), Québec lays the groundwork for a world observatory on the social impacts of artificial intelligence and digital technologies, 29 March 2018, https://www.newswire.ca/news-releases/quebec-lays-the-groundwork-for-a-world-observatory-on-the-social-impacts-of-artificial-intelligence-and-digital-technologies-678316673.html

Fouquet H., Nussbaum A. and Mawad M. (2018), European tech industry calls on JEDI to fight off U.S., China, 28 March 2018, https://www.information-management.com/articles/european-tech-industry-calls-on-jedi-to-fight-off-us-china

FUJII H., MANAGI S. (2017), Trends and Priority Shifts in Artificial Intelligence Technology, Invention: A global patent analysis, RIETI Discussion Paper Series 17-E-066, https://www.rieti.go.jp/en/rieti_report/203.html.

Future of Life Institute (FLI) (2017) “Asilomar AI Principles” https://futureoflife.org/ai-principles/

German Federal Ministry of Federal Ministry for Economic Affairs and Energy (BMWi) (2017) “Input to the OECD conference: paper on AI – Germany” (Position paper submitted to the OECD conference “AI: Intelligent machines, Smart policies”)

German Federal Ministry of Transport and Digital Infrastructure, Ethics Commission (2017) “Automated and connected driving” https://www.bmvi.de/SharedDocs/EN/publications/report-ethics-commission.pdf?__blob=publicationFile

Gershgorn, D (2017) “Germany’s self-driving car ethicists: All lives matter” https://qz.com/1061476/germanys-new-regulations-on-self-driving-cars-means-autonomous-vehicles-wont-compare-human-lives/ (accessed on Apr 16 2018)

Gupta, S., and C.D. Manning (date unknown), “Analyzing the Dynamics of Research by Extracting Key Aspects of Scientific Papers”, https://nlp.stanford.edu/pubs/gupta-manning-ijcnlp11.pdf

Heiner, D and Nguyen, C (2018) “Shaping human-centered artificial intellilgence” OECD the Forum Network, February 27, 2018, https://www.oecd-forum.org/users/86008-david-heiner-and-carolyn-nguyen/posts/30653-shaping-human-centered-artificial-intelligence

Hirano, S. (2017) “AI R&D Guidelines” Presented at the OECD Conference “AI: Intelligent Machines, Smart Policies”) October 25-26, 2018 http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-hirano.pdf

https://www.thestar.com/news/world/2015/04/17/how-a-toronto-professors-research-revolutionized-artificial-intelligence.html

Inaba, T. and M. Squicciarini (2017), “ICT: A new taxonomy based on the international patent classification”, OECD Science, Technology and Industry Working Papers, No. 2017/01, OECD Publishing, Paris, http://dx.doi.org/10.1787/ab16c396-en.

Information Technology Industry Council (ITI) (2017) “AI Policy Principles” https://www.itic.org/resources/AI-Policy-Principles-FullReport2.pdf

Institute of Electrical and Electronics Engineers (IEEE) (2017), Global Initiative on Ethics of Autonomous and Intelligent Systems, “Ethically Aligned Design Version 2”, http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf

Italian government (2017), “G7 ICT and Industry Ministers’ Declaration - Making the Next Production Revolution Inclusive, Open And Secure”, Torino, 25-26 September 2017 http://www.g7italy.it/sites/default/files/documents/G7%20ICT_Industry_Ministers_Declaration_%20Italy-26%20Sept_2017final_0.pdf and Annex 2 – Artificial Intelligence: http://www.g7italy.it/sites/default/files/documents/ANNEX2-Artificial_Intelligence_0.pdf

ITF (International Transport Forum) (2017), “Managing the Transition to Driverless Road Freight Transport”, International Transport Forum Policy Papers, No. 32, OECD Publishing, Paris, http://dx.doi.org/10.1787/0f240722-en.

Japanese Ministry of Internal Affairs and Communication (MIC), the Conference toward AI Network Society (2017) “Draft AI R&D Guidelines for International Discussions” (Tentative Translation) http://www.soumu.go.jp/main_content/000507517.pdf

Japanese New Energy and Industrial Technology Development Organisation (NEDO) (2017) “Artificial Intelligence Technology Strategy” http://www.nedo.go.jp/content/100865202.pdf

Kaevats, M. (2017) “Estonia's ideas on legalising AI” Presented at the OECD Conference “AI: Intelligent Machines, Smart Policies” October 25-26, 2018 https://prezi.com/yabrlekhmcj4/oecd-6-7min-paris/

Kania, E. (2018), “China’s AI Agenda Advances”, The Diplomat, 14 February 2018, https://thediplomat.com/2018/02/chinas-ai-agenda-advances/

Knight, Will (2017), “5 Big Predictions for Artificial Intelligence in 2017”, MIT Technology Review, 4 Jan 2017, at: https://www.technologyreview.com/s/603216/5-big-predictions-for-artificial-intelligence-in-2017/

Knight, Will (2018). “Here’s how the US needs to prepare for the age of artificial intelligence”, MIT Technology Review, 6 April 2018 at: https://www.technologyreview.com/s/610379/heres-how-the-us-needs-to-prepare-for-the-age-of-artificial-intelligence/?source=download-metered-content

Korean Government, Interdepartmental Exercise (2016) “Mid- to Long-Term Master Plan in Preparation for the Intelligent Information Society” http://english.msip.go.kr/cms/english/pl/policies2/__icsFiles/afieldfile/2017/07/20/Master%20Plan%20for%20the%20intelligent%20information%20society.pdf

Korean Ministry of Science and ICT (2017) “Science, Technology & ICT Newsletter (No.25)” http://english.msit.go.kr/english/msipContents/contentsView.do?cateId=msse44&artId=1325782 (accessed on Apr 13 2018)

MGI (2017), Artificial Intelligence - The Next Digital Frontier?, McKinsey Global Institute Discussion Paper June 2017, https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx

Miailhe, N (2017) “Position Paper” submitted to the OECD Conference “AI: Intelligent Machines, Smart Policies” October 25-26, 2018

Muller, C. (2017), European Economic and Social Committee Opinion on the societal impact of AI, 31 May 2017, https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/artificial-intelligence

Nadella, S (2017) “The Partnership of the Future” http://www.slate.com/articles/technology/future_tense/2016/06/microsoft_ceo_satya_nadella_humans_and_a_i_can_work_together_to_solve_society.html

Ilyas, A., Engstrom, L., Athalye, A., Lin J. (2018), Black-box Adversarial Attacks with Limited Queries and Information, 23 Apr 2018, arXiv:1804.08598

Nedelkoska, L. and G. Quintini (2018), "Automation, skills use and training", OECD Social, Employment and Migration Working Papers, No. 202, OECD Publishing, Paris, http://dx.doi.org/10.1787/2e2f4eea-en.

Nilsson, N. (2010), The Quest for Artificial Intelligence: A History of Ideas and Achievements, Cambridge University Press, Cambridge, United Kingdom.

O’Dwyer, G. “Finnish government backs national AI development strategy” (2018) https://www.computerweekly.com/news/252438764/Finnish-government-backs-national-AI-development-strategy (accessed on April 16 2018)

OECD (2013), Privacy Guidelines of 1980 (revised in 2013)

OECD (2015), Data-Driven Innovation: Big Data for Growth and Well-Being, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264229358-en.

OECD (2016), Recommendation on Consumer Protection in E-commerce, https://www.oecd.org/sti/consumer/ECommerce-Recommendation-2016.pdf.

OECD (2016a), “Summary of the CDEP Technology Foresight Forum: Economic and Social Implications of Artificial Intelligence”, OECD, Paris, http://oe.cd/ai2016.

OECD (2016b), OECD Science, Technology and Innovation Outlook 2016, OECD Publishing, Paris, http://dx.doi.org/10.1787/sti_in_outlook-2016-en.

OECD, (2017). OECD Digital Economy Outlook 2017. OECD Publishing. https://read.oecd-ilibrary.org/science-and-technology/oecd-digital-economy-outlook-2017_9789264276284-en#page7

OECD (2017a), The Next Production Revolution: Implications for Governments and Business, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264271036-en.

OECD (2017b), Consumer Product Safety in an Era of Technology-Driven Products and Supply Chains, [DSTI/CP/CPS(2017)1DSTI/CP/CPS(2017)1].DSTI/CP/CPS(2017)1].

OECD (2017c), “Basic income as a policy option: Can it add up?” Policy Brief on the Future of Work, OECD Publishing, Paris, www.oecd.org/employment/emp/Basic-Income-Policy-Option-2017.pdf.

OECD (2018), "Consumer product safety in the Internet of Things", OECD Digital Economy Papers, No. 267, OECD Publishing, Paris, http://dx.doi.org/10.1787/7c45fa66-en.

OECD (forthcoming), “AI: Intelligent Machines, Smart Policies – Conference Summary”, OECD, Paris, forthcoming, http://oe.cd/ai2017.

Oswald, Marion and Grace, Jamie and Urwin, Sheena and Barnes, Geoffrey, Algorithmic Risk Assessment Policing Models: Lessons from the Durham HART Model and 'Experimental' Proportionality (August 31, 2017). Information & Communications Technology Law, Forthcoming. Available at SSRN: https://ssrn.com/abstract=3029345

Pan-Canadian AI Strategy (2017), https://www.cifar.ca/assets/pan-canadian-artificial-intelligence-strategy-overview/

Park, L. (2016) "The South Korean government will invest 1 trillion won ($1 Billion USD) for the development of the intelligence information industry over the next five years" http://seoulspace.com/2016/03/22/the-south-korean-government-will-invest-1-trillion-won-1-billion-usd-for-the-development-of-the-intelligence-information-industry-over-the-next-five-years/ (Accessed on Apr 13 2018)

Partnership on AI to Benefit People and Society (PAI) (2017) "Partnership on AI Strengthens Its Network of Partners and Announces First Initiatives" (press release) https://www.partnershiponai.org/2017/05/pai-announces-new-partners-and-initiatives/

Partnership on AI to Benefit People and Society (PAI) (2016) "Tenets" https://www.partnershiponai.org/tenets/

Quid (2017), "Quid special report: The New Wave of Artificial Intelligence"https://quid.com/feed/quid-special-report-the-new-wave-of-artificial-intelligence.

Quid (2018), Artificial Intelligence companies landscape overview, 5 October 2018, https://vimeo.com/236989979.

Rayo, E-A., (2017) «AI in Law and Legal Practice – A Comprehensive view of 35 Current Applications», https://www.techemergence.com/ai-in-law-legal-practice-current-applications.

Salter, S., Thompson, D., (2017) « Public-Centered Civil Justice Redesign: a case study of the British Columbia Civil Resolution Tribunal », Mc Gill Journal of Dispute Resolution, Vol 3, 113 pp 114 – 136

Science (2017), "AI is changing how we do science", Science, July 5th 2017. http://www.sciencemag.org/news/2017/07/ai-changing-how-we-do-science-get-glimpse

Selbst, A. (2017), Disparate Impact in Big Data Policing (February 25, 2017). 52 Georgia Law Review 109 (2017). Available at SSRN: https://ssrn.com/abstract=2819182

Sivonen, P. (2017) "Ambitious development program enabling rapid growth of AI and platform economy in Finland" (Presented at OECD Conference "AI: Intelligent Machines, Smart Policies") http://www.oecd.org/going-digital/ai-intelligent-machines-smart-policies/conference-agenda/ai-intelligent-machines-smart-policies-sivonen.pdf

Sparkes, A.,W.Aubrey, E.Byrne, A.Clare, N.K.Muhammed, M.Liakata, M.Markham, J.Rowland, L.N.Soldatova, K.E.Whelan, M.Young and R.D.King (2010), "Towards Robot Scientists for autonomous scientific discovery", https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2813846/

Temperton, J. (2017) , "DeepMind's new AI ethics unit is the company's next big move", Wednesday 4 October 2017, http://www.wired.co.uk/article/deepmind-ethics-and-society-artificial-intelligence

The Japanese Society for Artificial Intelligence (JSAI) (2017) “The Japanese Society for Artificial Intelligence Ethical Guidelines” http://ai-elsi.org/wp-content/uploads/2017/05/JSAI-Ethical-Guidelines-1.pdf

Thompson, N., (2018), “Emmanuel Macron Talks to WIRED About France's AI Strategy”, Wired Magazine, 31 March 2018, https://www.wired.com/story/emmanuel-macron-talks-to-wired-about-frances-ai-strategy

Tractica, “Artificial Intelligence Market Forecasts”, 3Q 2016, https://www.tractica.com/wp-content/uploads/2016/08/MD-AIMF-3Q16-Executive-Summary.pdf.

Tsai China Center; Paul Triolo, Eurasia Group; and Elsa Kania https://na-production.s3.amazonaws.com/documents/translation-fulltext-8.1.17.pdf

UK Government (2017a), UK Digital Strategy, Published 1 March 2017 https://www.gov.uk/government/publications/uk-digital-strategy/uk-digital-strategy

UK Government (2017b), AI industry review [Press release] https://www.gov.uk/government/news/industry-led-review-details-plans-to-supercharge-uk-artificial-intelligence-ai-industry
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf

UK Government (2017c), industry strategy “Building a Britain fit for the future” https://www.gov.uk/government/publications/industrial-strategy-building-a-britain-fit-for-the-future

UNI Global Union (2017) “Top 10 Principles for ethical artificial intelligence” http://www.thefutureworldofwork.org/media/35420/uni_ethical_ai.pdf

United Nations Educational, Scientific and Cultural Organisation (UNESCO) and World Commission on the Ethics of Scientific Knowledge and Technology (COMEST) (2017) “Report of COMEST on robotics ethics” http://unesdoc.unesco.org/images/0025/002539/253952E.pdf

United States Supreme Court (2017), Wisconsin v. Loomis, June 2017 16-6387.

University of Montreal (2017), “The Montreal Declaration for a Responsible Development of Artificial Intelligence” https://www.montrealdeclaration-responsibleai.com/

US CONGRESS AI Caucus Press Release (2017) “Delaney Launches Bipartisan Artificial Intelligence (AI) Caucus for 115th Congress” https://artificialintelligencecaucus-delaney.house.gov/media-center/press-releases/delaney-launches-ai-caucus (accessed on Apr 13 2018)

US CONGRESS.GOV (2017) “AV START Act” https://www.congress.gov/bill/115th-congress/senate-bill/1885/all-info https://www.congress.gov/115/bills/s1885/BILLS-115s1885rs.pdf

US CONGRESS.GOV (2017) “H.R. 4625 – FUTURE of Artificial Intelligence Act of 2017”https://www.congress.gov/bill/115th-congress/house-bill/4625
https://www.congress.gov/115/bills/hr4625/BILLS-115hr4625ih.pdf

US CONGRESS.GOV (2017) “H.R.3388 – SELF DRIVE Act” https://www.congress.gov/bill/115th-congress/house-bill/3388
https://www.congress.gov/115/bills/hr3388/BILLS-115hr3388rfs.pdf

US CONGRESS.GOV (2017) "S. 2217 – FUTURE of Artificial Intelligence Act of 2017" https://www.congress.gov/bill/115th-congress/senate-bill/2217 https://www.congress.gov/115/bills/s2217/BILLS-115s2217is.pdf

US White House Executive Office of the President (2016), "Artificial Intelligence, Automation, and the Economy" https://obamawhitehouse.archives.gov/sites/whitehouse.gov/files/documents/Artificial-Intelligence-Automation-Economy.pdf

US White House Office of Science and Technology Policy (2016), "National Artificial Intelligence Research and Development Strategic Plan" https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/national_ai_rd_strategic_plan.pdf

US White House Office of Science and Technology Policy (2016), "Preparing for the future of artificial intelligence" https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

US White House Office of Science and Technology Policy (2018) "Science and Technology Highlights" https://www.whitehouse.gov/wp-content/uploads/2018/03/Administration-2017-ST-Highlights.pdf

Villani, C. (2018), "For a Meaningful Artificial Intelligence - Towards a French and European Strategy", https://www.aiforhumanity.fr/

Wachter, S., Mittelstadt, B. and Russell, C.. 2017, "Counterfactual explanations without opening the black box: Automated decisions and the GDPR", arXiv preprint arXiv:1711.00399.

West, D., Bernstein, D., (2017), « Benefits and Best Practices of Safe City Innovation », Center for Technology Innovation, Brookings, p19

Yala, A., Barzilay, R., Salama, L., Griffin, M., Sollender, G., Bardia, A., ... & Garber, J. E. (2017). Using machine learning to parse breast pathology reports. Breast cancer research and treatment, 161(2), 203-211.

Annex A. INFORMATION ON CRUNCHBASE

Crunchbase is a commercial database on innovative companies maintained by Crunchbase Inc., an innovative start-up in itself, located in California, US. The database was created in 2007 but its scope and coverage has increased significantly over the past few years. As reported by Kaufmann Foundation, the database is increasingly used by the venture capital industry as a "the premier data asset on the tech/startup world".[1] Dalle, den Besten, & Menon (2017) present a detailed discussion of the database and its potential for economic, managerial, and policy-oriented research. Compared to commercial databases covering similar information and frequently used for economic research (see e.g. Da Rin, Hellman, & Puri, 2011, for an overview of available data sources), Crunchbase has major advantages. It is partially crowd-sourced, i.e., users can add and revise contents, which add to the comprehensiveness and timeliness of the database; is updated on a daily basis; contains cross-linked information on companies, their funders, and their staff; and it is structured in an accessible way. Furthermore, it lists both companies that have received VC and start-ups that have not been funded yet but that are presumably actively looking for funding, and thus permits a meaningful comparison between both types of firms.

Consequently, academic interest in Crunchbase has recently grown and research using this database has been published in major journals. Examples include (but are not restricted to) Alexy, Block, Sandner, & Ter Wal (2012), Bertoni & Tykvovà (2015), and Block, Fisch, Hahn, & Sandner (2015). For a more detailed literature review, see Dalle, den Besten, & Menon (2017), who discuss more than 80 academic studies in the field of economic, managerial, and entrepreneurship research based on the Crunchbase data.

In the version used for this report, downloaded in April 2018, the database contains information on more than 500 000 distinct entities located in 199 different countries. The raw data are obtained through two main channels: a large investor network and community contributors. More than 3000 global investment firms submit monthly portfolio updates to Crunchbase, in exchange for free data access. In addition, around 500,000 executives, entrepreneurs, and investors contribute to update and revise Crunchbase company profile pages. This wealth of data is processed by the Crunchbase analyst team with the support of artificial intelligence (AI) and machine learning algorithms, in order to ensure accuracy and scan for anomalies. Additionally, algorithms continuously search the web and thousands of news publications for information to enrich profiles.

Breschi, Lassebie, and Menon (2018) discuss the coverage and representativeness of the database, compared to some benchmark data sources that are more commonly used in the literature. The general message of the benchmarking exercise is that Crunchbase has a better coverage of VC deals and start-ups than comparable data sources. The country-year comparison with aggregated sources on VC investments also suggests that the coverage of Crunchbase is sufficiently exhaustive across OECD member countries and other large emerging economies (Brazil, China, India, Russia, South Africa), with few exceptions. It

[1] http://www.kauffman.org/microsites/state-of-the-field/topics/finance/equity/venture-capital accessed on September 11th, 2017.

is recognised that databases on VC all suffer from some sample selection issues (Da Rin, Hellman, & Puri, 2011) and CrunchBase is not an exception. However, Crunchbase is quickly becoming a reference for professionals seeking to invest in start-ups, and young firms have strong incentives to appear in the database. Therefore, the results presented in this report can be generalised to start-ups that are actively looking for funding opportunities.

Companies are classified into 45 different economic activity groups, which henceforth are referred to as "sectors". This classification does not appear to follow any major industry classification system, but rather to be especially customised to the start-up world and to emerging sectors. It is however possible to benchmark the Crunchbase classification against the NACE rev. 2 industry classification in the sample of 27 thousand companies for which it is possible to find a unique correspondent in the ORBIS database with the same name and country code. The comparison shows that Crunchbase categories are overall meaningful, as the distribution of 2-digit NACE codes within categories show a fair degree of concentration. At the same time, the benchmarking also shows that in 37 out of 45 categories the relative majority of companies is classified either in the "Computer programming, consultancy and related activities" NACE sector (code 62) or in the "Manufacture of computer, electronic and optical products" one (code 26). This may suggest that the NACE classification may not fully capture the technological diversification of new start-ups, at least at 2-digit level.

Crunchbase also contains around 580 thousand records on people who are connected to at least one company listed in the database. The following variables are reported: the full name, location (city and region), gender, job title, and the dates on which the record was created and updated, respectively. Most people are classified as founder, co-founder, or CEO. In addition, the database also contains two linked tables reporting the education and the employment history, respectively, of the listed individuals. Whenever feasible, this information has been complemented and cross-validated with data taken from Breschi et al. (2017). While education and employment history is not available for the full sample of listed individuals, the data allow to analyse the "curriculum vitae" of approximately 130 thousand people listed as founders or managers of more than 25 thousand start-ups.

Furthermore, the database covers 230 thousand VC deals, 11 thousands IPOs, and 34 thousand acquisitions. The table on investors cover 50 thousand entities, which are classified by their country code and their "type" (e.g., investment bank, business angel, incubators, etc.). This latter classification enables to identify government venture capital (GVC) funds, incubators, and other public investors. The list of GVC investors has been further refined using several additional sources: the 2012 OECD Financing Questionnaire (Wilson & Silva, 2013), the membership list of InvestEurope (the European association of VC investors), and a list of government VC funds compiled manually by the authors. These three lists have been combined and matched to the main database with a fuzzy matching procedure on investors' names in each country.[2]

Caveats

- As the large majority of companies are reported as still active, the historical dimension of the database is mainly limited to the snapshot of companies that are still operating today. i.e., it is

[2] This procedure compares for each country investors names in CrunchBase and names in the list of government VC described above, computes a similarity score for each pair of names and matches for each investor in CrunchBase its closest counterpart in the list of government VC, given that they are located in the same country. An investor is then classified as public if the similarity score exceeds a certain threshold. From the list of investors in CrunchBase, the different procedures described above allow for the identification of approximately 1,000 public investors (2 percent of the sample, while 35 percent remain unclassified).

possible to track the funding history of companies that are still operating today (at least nominally), while it is likely that past investments in companies that became insolvent and closed down definitively are underreported. For the same reason, the coverage for the most recent years is likely to be more comprehensive than for earlier periods.

- The scope of the database is not precisely defined. Their website states that “Crunchbase is the leading destination for millions of users to discover industry trends, investments, and news about global companies—from startups to the Fortune 1000. Crunchbase was founded to be the master record of data on the world’s most innovative companies.”

- A significant number of investment transactions (one quarter of the AI-related transactions) do not disclose the dollar amount of the investment. The assumption: these deals are assumed to be small or medium sized investments (less than USD 10 million). For the United States and for Europe, averages for the undisclosed transactions. companies with Average were calculated per country for Estimates of the amounts

Source: *adapted from* *https://www.oecd-ilibrary.org/science-and-technology/a-portrait-of-innovative-start-ups-across-countries_f9ff02f4-en*

NOTES

[1] OECD Conference “AI: Intelligent Machines, Smart Policies” held in Paris on 26-27 October 2017, presentations at http://oe.cd/ai2017.

[2] Francesca Rossi, Research Scientist, IBM Watson and Professor of Computer Science, University of Padova, Italy (*Technical & Ethical Challenges to Human-AI Collaboration)*

[3] Stephen Roberts, Professor of Machine Learning in Information Engineering, University of Oxford, United Kingdom – “*21st century science: the age of intelligent algorithms* **”**

[4] Philipp Slusallek, Scientific Director at DFKI, Germany, focused on how to ensure AI systems interact safely and reliably with the very complex real world ***(“****Artificial intelligence and digital reality: Do we need a "CERN for AI"?****)***

[5] Francesca Rossi, Research Scientist, IBM Watson and Professor of Computer Science, University of Padova, Italy (*Technical & Ethical Challenges to Human-AI Collaboration)*

[6] Stuart Russell, Professor of Computer Science, University of California, Berkeley, USA (*Human-compatible artificial intelligence***)**

[7] Valerio Dilda, Partner, Paris, McKinsey & Company, *AI: perspectives and opportunities*.

[8] https://www.crunchbase.com/organization/toutiao

[9] Note: Start-ups generally selected more than one sector. Excludes start-ups for which the sector identified was “artificial intelligence” or “AI” (11 percent of AI start-ups). Percentage based on the number of AI start-ups of which the description included the sector, divided by total number of AI start-ups for which a meaningful sector was identified. Excludes sectors for which the percentage was lower than 2 percent. Note 2: The science and engineering category includes companies such as driverless car activities like ArgoAI, autonomous mobility company zoox.com or MIT-IBM Watson AI Lab.

[10] According to Google patent experts, publications may focus more on fundamental machine learning inventions than on applied inventions. One way to identify fundamental research in machine learning will be to review papers submitted to the three largest machine learning conferences: International Conference on Machine Learning (ICML), the Conference and Workshop on Neural Information Processing Systems (NIPS), and the International Conference on Learning Representations (ICLR). This is expected to be done for Phase 2 of the report.

[11] The OECD Digital Economy Outlook (2017) identified broad types of benefits of AI: i) improving efficiency, saving costs and enabling better resource allocation; ii) helping to identify suspicious activity, people or information; iii) generating a new wave of productivity gains, and; iv) helping address complex challenges in areas like health, transport and security.

[12] OECD Conference “AI: Intelligent Machines, Smart Policies” held in Paris on 26-27 October 2017, presentations at http://oe.cd/ai2017.

[13] Reinhard Stolle, Department of Artificial Intelligence at BMW AG, Munich – ***(****AI as a driver of the automotive industry****)***

[14] The International Transport Forum (2017) found that driverless trucks could reduce the costs of trucking by 30 percent. Young Tae Kim, Secretary General, International Transport Forum (ITF) – New transport for the new digital age.

[15] CISP is examining infrastructure requirements for the Internet of things, including driverless cars [DSTI/CDEP/CISP/MADE(2017)1].

[16] OECD Committee on Consumer Policy (CCP) on behavioural advertising.

[17] Bryan Yates, Director of Sales - EMEA region, Orbital Insight, Mountain View, California *(New geoanalytics: tracking economies from space)*, *http://oe.cd/ai2017*

[18] Thanh-Long Huynh, CEO, Quantcube Technology, Paris – Big data analytics for strategic intelligence, *http://oe.cd/ai2017*

[19] Tugdual Ceillier, Lead Data Scientist, EarthCube, Toulouse (Artificial intelligence and remote sensing: new capabilities to monitor infrastructure), *http://oe.cd/ai2017*

[20] Stephen Roberts, Professor of Machine Learning in Information Engineering, University of Oxford, United Kingdom (*21st century science: the age of intelligent algorithms*)

[21] Ross King, Professor of Machine Intelligence, Manchester University School of Computer Science, United Kingdom (*The automation of science*)

[22] Hiroaki Kitano, President and CEO of Sony Computer Science Laboratories, Japan – The Nobel Turing Challenge: creating the engine of scientific discovery

[23] http://www.oecd.org/going-digital/digital-security-in-critical-infrastructure/.

[24] See more information on the VOIE project at http://www.gouvernement.fr/demonstrateur-plateforme-voie-3230 and on Violent Event Detection videos http://www.kalisteo.eu/ressources/videos/computer-vision-violent-event-detection-real-train.mp4

[25] Max Yuan, founder and chairman, Xiaoi Robot Technology, Shanghai ***(****AI empowers government and enterprises****)***

[26] Bahaa Alhaddad, Space Business Development, Starlab Space, Harwell Oxford, United Kingdom– Neurosciences and space data: a new big bang

[27] “The Digital Transformation of the Public Sector: Helping Governments Respond to the Needs of Networked Societies” (GOV/PGC/EGOV(2016)5/REV1).

OECD Conference “AI: Intelligent Machines, Smart Policies” held in Paris on 26-27 October 2017, presentations at http://oe.cd/ai2017.

[29] Stuart Elliot, from the US National Academies of Sciences, Engineering and Medicine ***(****AI and the future of skill demand****)***.

[30] Frank Levy, Rose Professor Emeritus at the Massachusetts Institute of Technology, added to these concerns *(Computers and populism)*. Frank Levy, Rose Professor Emeritus at the Massachusetts Institute of Technology, added to these concerns*(Computers and populism)*.

[31] Young Tae Kim, Secretary General of the International Transport Forum ***(****New transport for the new digital age****)***

[32] Konstantinos Karachalios, Managing Director of the IEEE-Standards Association, *The Role Of Technical Communities in making Intelligent Technologies Work for the Benefit of Humanity*

[33] Tugdual Ceillier, Lead Data Scientist, EarthCube, Toulouse *(Artificial intelligence and remote sensing: new capabilities to monitor infrastructure)*

[34] Rodolphe Gelin, Robotics Software Engineering Lead, SoftBank Robotics, Paris **(***Robots, man's best friend***)**

[35] Valerio Dilda, Partner, Paris, McKinsey & Company (*AI: perspectives and opportunities)*

[36] Mark Keese, Head of OECD Division on Skills and Employability

[37] Joanna Bryson, Reader at University of Bath, and Affiliate, Center for Information Technology Policy at Princeton University (*Current and Potential Impacts of Artificial Intelligence and Autonomous Systems on Society)*

[38] Stuart Russell, Professor of Computer Science, University of California, Berkeley, USA (*Human-compatible artificial intelligence***)**

[39] Garry Kasparov, Former World Chess Champion and author of 'Deep Thinking'

[40] Christina Colclough

[41] Jonathan McLoone**,** Technical Director, Wolfram Research Europe (*Preparing science for AI: rethinking education, research and publication***)**

[42] Osamu Sudoh, Professor, University of Tokyo Interfaculty Initiative in Information Studies, Japan – **(***Towards AI network society - addressing social, economic, ethical and legal issues***)**

[43] James Hairston, Head of Public Policy, Oculus VR, Facebook **(***AI, employment, and general purpose technologies***)**

[44] "The Future of Skills: Understanding the Educational Implications of AI and Robotics" (EDU/CERI/CD(2018)5).

[45] Rod Freeman, international products lawyer, Partner at Cooley, UK (*Evolution or revolution? The future of regulation and liability for AI*)

[46] Hans Ingels, Head of Unit, Single Market Policy, Mutual Recognition and Surveillance, European Commission, DG GROW (*Artificial intelligence and EU product liability law)*

[47] Georg Borges, Professor, Faculty of Law, Saarland University, Germany (*Liability for machine-made decisions: gaps and potential solutions)*

[48] Pierre Chalançon, Chair of the BIAC Consumer Task Force and Vice President Regulatory Affairs, Vorwerk & Co KG, Representation to the EU – *Science-Fiction is not a Sound Basis for Legislation*

[49] which would require advanced technical understanding.

[50] Marc Rotenberg, representative of OECD Civil Society Information Society Advisory Council (CSISAC) and President, Electronic Privacy Information Center (EPIC)

[51] Francesca Rossi, Research Scientist, IBM Watson and Professor of Computer Science, University of Padova, Italy (*Technical & Ethical Challenges to Human-AI Collaboration)*

[52] Stuart Russell, Professor of Computer Science, University of California, Berkeley, USA (*Human-compatible artificial intelligence***)**,

[53] Alexander Cooke, Counsellor, Department of Industry, Innovation and Science, Australia (*Digital Earth Australia)*

54 For example, Peter Fleischer, Global Privacy Counsel, Google *(Privacy and AI: designing machine learning systems to respect privacy)*

55 Konstantinos Karachalios, Managing Director of the IEEE-Standards Association (*The Role Of Technical Communities in making Intelligent Technologies Work for the Benefit of Humanity)*

[56] Mathias Cellarius, Data Protection and Privacy Officer, SAP – *AI: Challenges and Opportunities for Data Protection*

57 Benedetta Arese Lucini (Italy) ***(****G7 Italy: towards a human-centric AI****)***

58 Nicole Primmer, Senior Policy Director, Business at OECD (BIAC)

59 http://www.unicri.it/news/article/2017-09-07_Establishment_of_the_UNICRI

60 https://www.itu.int/en/ITU-T/AI/

61 https://www.coe.int/en/web/freedom-expression/-/first-meeting-of-the-msi-aut-on-algorithms-and-artificial-intelligence

62 Seán Ó hÉigeartaigh, Executive Director of Cambridge's Centre for the Study of Existential Risk (*Asilomar Principles)*

63 Since that time, new partners have joined the partnership, including for-profit companies (eBay, Intel, McKinsey & Company, Salesforce, SAP, Sony, Zalando, and Cogitai), and non-profits (Allen Institute for Artificial Intelligence, AI Forum of New Zealand, Center for Democracy & Technology, Centre for Internet and Society – India, Data & Society Research Institute, Digital Asia Hub, Electronic Frontier Foundation, Future of Humanity Institute, Future of Privacy Forum, Human Rights Watch, Leverhulme Centre for the Future of Intelligence, UNICEF, Upturn, and the XPRIZE Foundation). They join the founding companies and existing non-profit Partners (AAAI, ACLU and OpenAI). The partnership's tenets include a commitment to open research and dialog on the ethical, social, economic and legal implications of AI and to developing AI research and technology that is robust, reliable, trustworthy and operates within secure constraints.