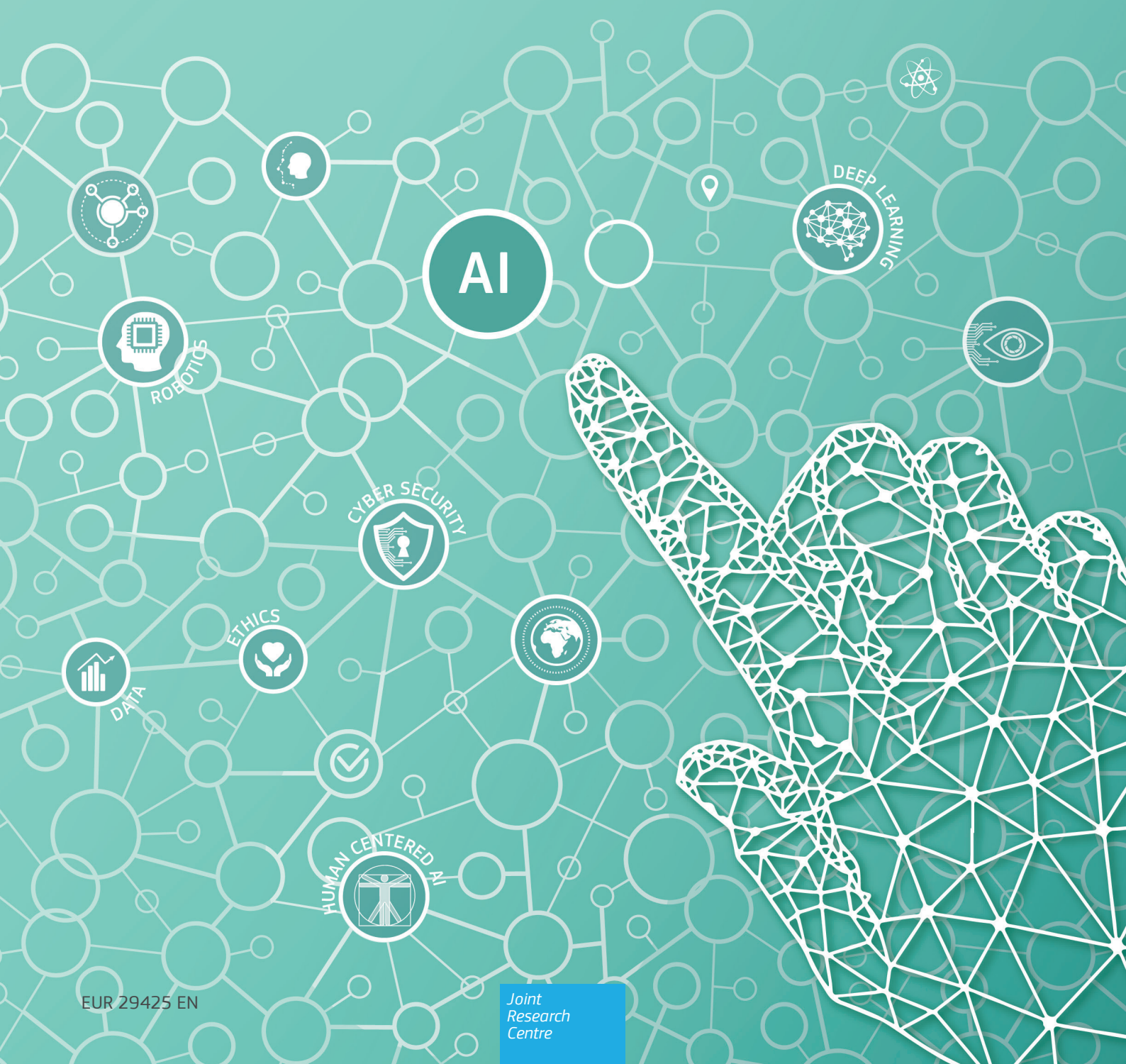




ARTIFICIAL INTELLIGENCE

A EUROPEAN PERSPECTIVE



This publication is a Science for Policy report by the Joint Research Centre (JRC), the European Commission's science and knowledge service. It aims to provide evidence-based scientific support to the European policymaking process. The scientific output expressed does not imply a policy position of the European Commission. Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use that might be made of this publication.

Contact information

Max Craglia (Ed.)

European Commission, Joint Research Centre, TP262, Via Fermi, 21027 Ispra (VA), ITALY

massimo.craglia@ec.europa.eu

Tel.: +39 0332 786 269

EU Science Hub

<https://ec.europa.eu/jrc>

JRC113826

EUR 29425 EN

PDF ISBN 978-92-79-97217-1 ISSN 1831-9424 doi:10.2760/11251

Print ISBN 978-92-79-97219-5 ISSN 1018-5593 doi:10.2760/936974

Luxembourg: Publications Office of the European Union, 2018

The reuse policy of the European Commission is implemented by Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Reuse is authorised, provided the source of the document is acknowledged and its original meaning or message is not distorted. The European Commission shall not be liable for any consequence stemming from the reuse. For any use or reproduction of photos or other material that is not owned by the EU, permission must be sought directly from the copyright holders.

All content © European Union, 2018 except: cover © European Union, 2018 - graphic elaboration from ©wvihr - AdobeStock; ©Oksancia - AdobeStock; ©Gstudio Group - AdobeStock; p. 11 ©metamorworks - AdobeStock; p. 18 ©zapp2photo - AdobeStock; p. 24 ©metamorworks - AdobeStock; p. 36 ©Jakub Jirsák - AdobeStock; p. 39 ©adam121 - AdobeStock; p. 42 ©peshkova - AdobeStock; p. 44 ©Christian Horz - AdobeStock; p. 44 ©bymandesigns - AdobeStock; p. 47 ©Gorodenkoff - AdobeStock; p. 54 ©fizkes - AdobeStock; p. 56 ©enterlinedesign - AdobeStock; p. 58 ©THANANIT - AdobeStock; p. 62 ©NicoElNino - AdobeStock; p. 67 ©nitsawan - AdobeStock; p. 70 ©zinkevych - AdobeStock; p. 76 ©phonlamaiphoto - AdobeStock; p. 83 ©metamorworks - AdobeStock; p. 85 ©JackF - AdobeStock; p. 86 ©Gorodenkoff - AdobeStock; p. 91 ©zapp2photo - AdobeStock; p. 92 ©vchalup - AdobeStock; p. 94 ©Gorodenkoff - AdobeStock; p. 99 ©seventyfour - AdobeStock; p. 101 ©Tierney - AdobeStock; p. 102 ©fgnopporn - AdobeStock; p. 114 ©orbital - AdobeStock; p. 120-122 ©Fabian- AdobeStock, and where otherwise stated.

How to cite this report: Craglia M. (Ed.), Annoni A., Benczur P., Bertoldi P., Delipetrev P., De Prato G., Feijoo C., Fernandez Macias E., Gomez E., Iglesias M., Junklewitz H., López Cobo M., Martens B., Nascimento S., Nativi S., Polvora A., Sanchez I., Tolan S., Tuomi I., Vesnic Alujevic L., *Artificial Intelligence - A European Perspective*, EUR 29425 EN, Publications Office, Luxembourg, 2018, ISBN 978-92-79-97217-1, doi:10.2760/11251, JRC113826.



ARTIFICIAL INTELLIGENCE

A EUROPEAN PERSPECTIVE

TABLE OF CONTENTS

Executive summary	8
Part 1: Introduction and AI international landscape	15
1 Motivation and objective of this report	16
2 About AI	19
2.1 A brief history of AI	19
2.2 Recent developments in machine learning	20
2.3 Recent developments in social robots	21
2.4 Current challenges	22
2.5 Summary and conclusions	23
3 EU in the AI competitive global landscape	25
3.1 The international policy context	25
3.1.1 USA	25
3.1.2 Other countries	27
3.2 Analysing the key features of the AI landscape	29
3.3 Overview	31
3.4 Technological capacity	34
3.5 Summary and conclusions	35
4 AI in the EU	37
4.1 Strategies and plans	37
4.1.1 The European Union	37
4.1.2 France	40
4.1.3 United Kingdom	41
4.1.4 Finland	43
4.1.5 Other EU countries	43
4.2 Summary and conclusions	43
5 The AI ecosystem in China	45
5.1 China's economic and policy context	45
5.2 Government policies and initiatives	46
5.3 Regional/local initiatives	49
5.4 The specificities of the AI innovation ecosystem in China	49
5.5 Summary and conclusions	51

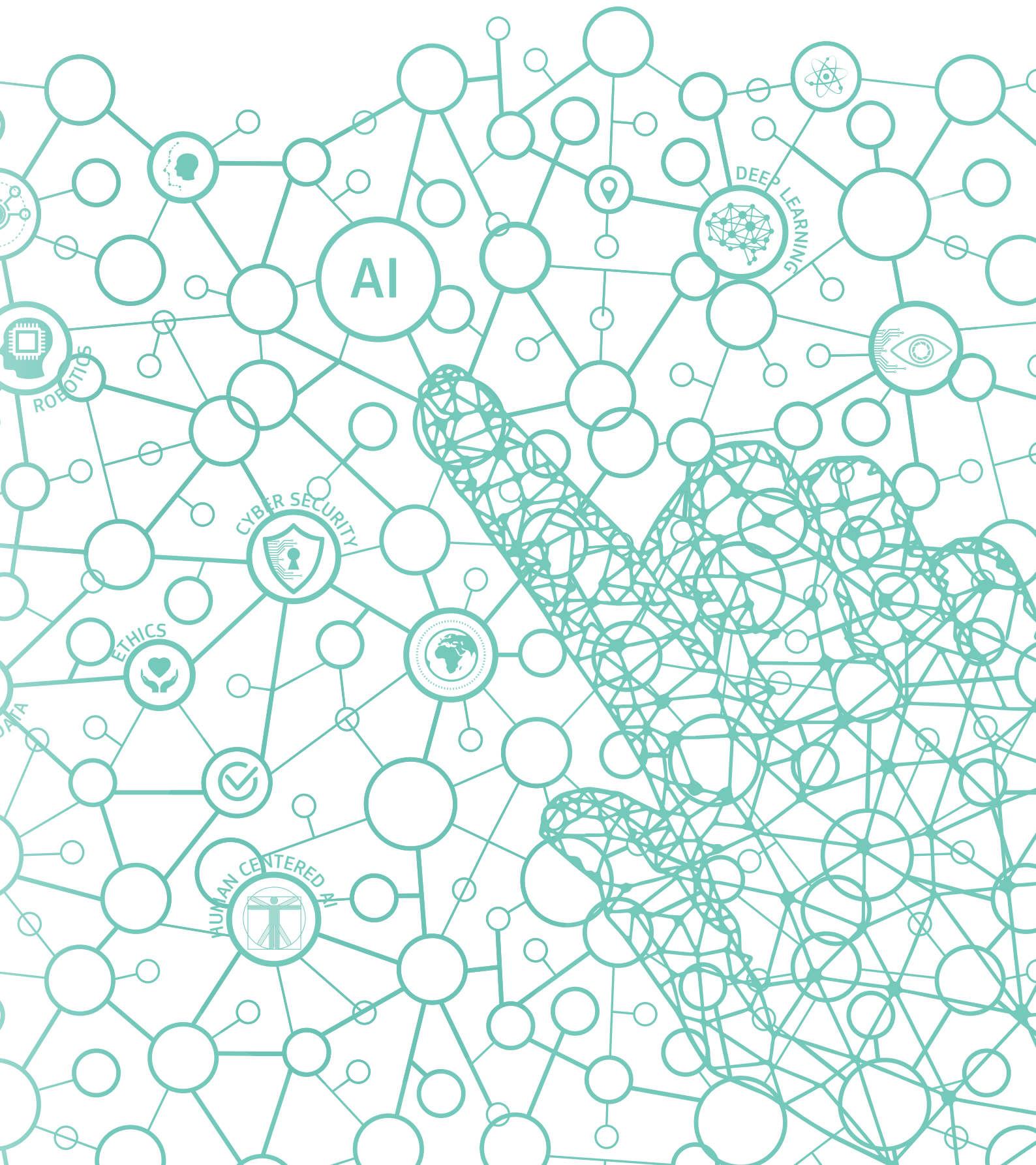


	Part 2: Multi-dimensional perspectives	53
6	Ethical and societal perspective	55
6.1	Introduction	55
6.2	Overview of individual and collective implications of AI	56
6.2.1	Challenges at individual level	56
6.2.2	Challenges at societal level	57
6.3	Summary and conclusions	60
7	Legal perspective	63
7.1	Main legal challenges identified in European AI strategies	63
7.2	Ownership, access and sharing of data	65
7.3	The protection of AI inventions/creations by intellectual property rights	66
7.4	Regulatory approach	68
7.5	Summary and conclusions	69
8	Educational perspective	71
8.1	Introduction	71
8.2	AI impact on skills demand, learning and teaching	72
8.2.1	Direct AI impact on advanced skills demand	72
8.2.2	Impact of AI on learning	72
8.2.3	Impact of AI on teaching	73
8.3	AI skills and academic supply	74
8.4	Summary and conclusions	75
9	Economic perspective	77
9.1	Potential impact of AI on jobs	77
9.2	Potential impact of AI on growth	80
9.3	Potential impact of AI on inequality	82
9.4	Summary and conclusions	84
10	Cybersecurity perspective	87
10.1	Background: AI and cybersecurity	87
10.2	Applications of AI in cybersecurity	88
10.3	Deterrence and fight against crime	88
10.4	Robustness of AI algorithms against malicious action	89
10.5	Summary and conclusions	91



11	Computer processing and energy perspective	95
11.1	Introduction	95
11.2	Assessment of data centre (DC) energy consumption	96
11.3	Options to improve the energy efficiency of the increasing demand for HPC.	96
11.3.1	CPU advancements: energy saving and parallelisation computing	97
11.3.2	Innovative and more efficient cooling systems/engineering solutions	98
11.3.3	Innovative infrastructures architectures and configurations	98
11.4	Summary and conclusions	100
12	Data perspective	103
12.1	The law and economics of access to data	103
12.1.1	The economic characteristics of data	103
12.1.2	Applying economic reasoning to data access	104
12.1.3	Policy intervention in the data market	105
12.2	Summary	108
12.3	Towards a data strategy for public administrations	108
12.3.1	Setting the scene	108
12.3.2	About platforms and APIs	109
12.3.3	From opening data to smart sharing	110
12.4	Summary and conclusions	113
13	Societal resilience perspective	115
13.1	The need for resilience	115
13.2	Measuring resilience to AI	116
13.3	Lessons from the analysis of AI for understanding resilience	119
14	Summary and conclusions	120
	Footnotes	122
	References	125
	List of boxes	136
	List of figures	137
	Acknowledgements	138





EXECUTIVE SUMMARY

This report presents a European view of **Artificial Intelligence (AI)** based on independent research and analysis by the European Commission Joint Research Centre to inform the debate at the European level.

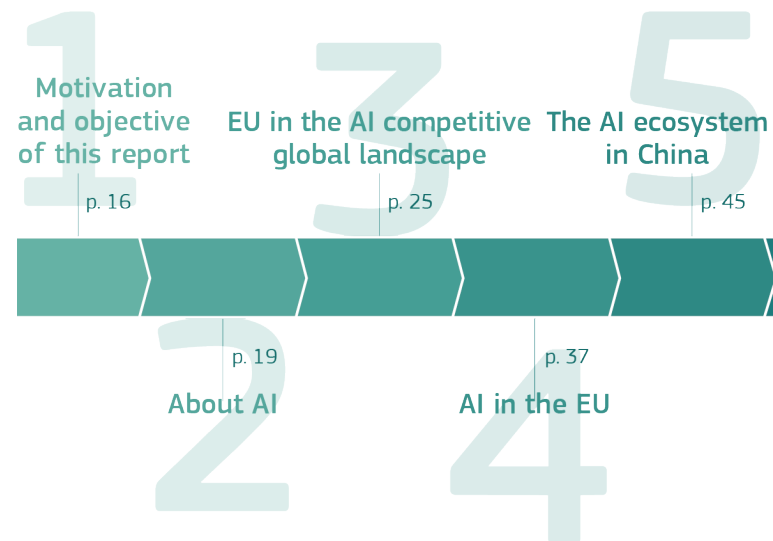
We first introduce AI as a generic term that refers to any machine or algorithm that is capable of observing its environment, learning, and based on the knowledge and experience gained, take intelligent actions or propose decisions. Autonomy of decision processes and interaction with other machines and humans are other dimensions that need to be considered.

Although many of the methodological developments in AI date back more than 50 years, the reason why we now pay so much attention to AI in general and machine learning (ML) in particular is that the recent advances in computing power, availability of data, and new algorithms have led to major breakthroughs in the last six to seven years. The many applications of AI/ML have started to enter into our everyday lives, from machine translations, to image recognition and music generation, and are increasingly being exploited in industry, government and commerce (*see Chapter 2*).

It is likely that we are only at the beginning of this process because the development of ubiquitous sensor networks, the IoT, will increase exponentially the sensing capabilities of AI, the volumes of data on which to train the algorithms, and their reach in society through decisions and actions.

The opportunities are many, and in some cases not yet foreseen. There are also many challenges,

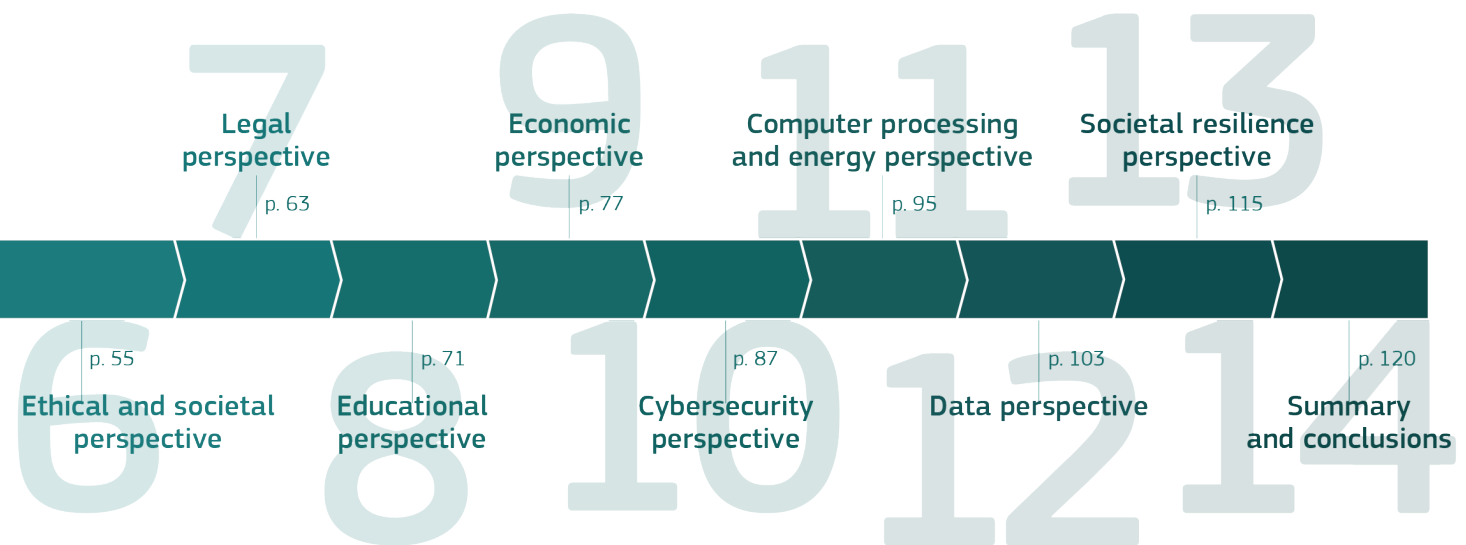
however. Among them, current ML algorithms display some of the characteristics of a black box: we access the inputs and outputs but do not understand fully what happens in-between, and how certain outputs, including decisions and actions, are derived. This calls for a greater effort to understand their theory and to develop explainable and accountable algorithms. We also need strong evalua-



tion frameworks that can assess not only the performance but also the quality of AI and build trust in this disruptive technology (*Chapters 2, 6 and 10*). The overview of the global and European AI landscape shows that there is an intense competition on AI taking place world-wide with three main leaders: the USA, Europe, and China (*Chapter 3*). Each region has about one quarter of all key players¹ in the AI field, including both research and industry, but has a distinctive different mix of players: while Europe is

well balanced in the number of research and non-research players, the USA has approximately three times as many industrial/corporate players as research ones, and China has about six times as many research players as industrial ones. The strength of the corporate world in the USA is also indicated by its dominance in the number of start-ups (almost half of the total worldwide) and venture capital (more than one third of the total). China, on the other hand, is making a strong effort to turn research research into patents, and accounts for almost 60 % of the world total. It has also put in place a strongly coordinated approach to AI, including government policy, industrial applications and research with the objective of becoming the world leader in AI by 2030. This is an ambitious but achievable target ([Chapter 5](#)).

guide the development of AI, with shared concerns over the need for an agreed ethical framework and applications that clearly benefit European society and uphold the European values enshrined in the Treaties ([Chapter 4](#)). The High-Level Expert Group established in 2018 by the European Commission is elaborating a framework for subsequent developments linked to these values. We also note the high level of awareness that data is crucial to the development of AI, with policy documents at both national and European level putting strong emphasis on the need to share data better among all the stakeholders: the public sector, industry, and the public. Finding the right way of doing it is challenging, as analysed in [Chapter 12](#) of the report.



Europe is currently well positioned in the quality of its research production, with more than 30 % of all papers on AI published in top scientific journals, just behind the USA (33 %), and is considerably ahead of China. Key areas of strength in Europe on which to build upon are automated and connected vehicles and robotics.

We note that many European countries as well as the EC are developing strategies and programmes to

After the overview and analysis of the AI global and European landscapes we discuss AI from multiple perspectives to add some depth and explore synergies.

As we highlight at multiple stages in this report, a key characteristic of the European way to AI has to be a strong ethical framework. There is consensus on this principle, and we report ([Chapter 4](#)) on many initiatives at both national and European levels to develop ethical guidelines to frame the

development of AI. We review the main dimensions that such guidelines should consider and highlight the potential implications of AI at the level of individuals and society. We conclude that to build and retain trust in AI we need a multi-layered approach that includes the critical engagement of civil society to discuss the values guiding and being embedded into AI, public debates in different fora to translate these values into strategies and guidelines, and responsible design practices that encode these values and guidelines into AI systems, so that they are ethical-by-design (*Chapter 6*).

In reviewing the European legal framework for AI, particularly with respect to the fundamental rights, data ownership, and intellectual property, we note the tensions between protecting rights of individuals and firms and encouraging innovation with trying to maximise openness and transparency. We conclude however that Europe is well placed to establish a distinctive form of AI that is ethically robust and protects the rights of individuals, firms, and society at large. For example, the General Data Protection Regulation, opposed by many during preparation, is now perceived as a European asset and is inspiring similar approaches outside Europe. Extending this notion, we should consider the high standards of the European legal and regulatory landscape in a similar way to those in environmental quality which are an asset for Europeans and their future generations to build upon, not a barrier (*Chapter 7*).

From an educational perspective, we observe (*Chapter 8*) that AI has potential positive impacts on shortages of skills, learning, and teaching. Three crucial points stand out from the review: firstly, the need to understand better how the interaction with AI impacts human intelligence in cognitive capacities in both adults, and even more importantly, children. Secondly, we need to think beyond current needs and practices, and consider how AI is likely to change the relationship between education and work, and human development. Thirdly, we highlight possible risks related to AI in education, particularly privacy and ethical ones. In this chapter, we also provide an initial overview

and geographic distribution of the academic offer of study topics related to AI. This is relevant to the discussion in *chapters 12 and 13* on possible strategies to start preparing society, and the most vulnerable regions, to the challenges that AI will bring.

“The digital transformation of society has just begun: *AI is central to this change and offers major opportunities to improve our lives.*”

The potential impact of AI on the labour market and inequality raises concerns in the media, research and public debates. We analyse the literature and evidence available to date on the potential impacts, both positive and negative, with respect to work, growth, and inequality. In relation to work, we find that neither theory nor evidence are very conclusive at the present time. AI could complement and enhance human activity, replace an increasing number of routine tasks, or both. Studies measuring the share of jobs at high risk of automation exhib-

it a high variety in their findings depending on the definition and level of granularity at which tasks and occupations are defined (*Chapter 9*).

Another area of uncertainty is the extent to which AI has the potential to spur economic growth. When considered as a general-purpose technology,



AI could spread across many jobs and industrial sectors, boosting productivity, and yielding strong positive growth. To the extent that ML generates new insights from data, it may also contribute to the production of new ideas and innovation itself. Economic growth models are starting to explore various scenarios but there is no empirical evidence yet that favours one or the other.

When it comes to inequality, we find that AI can affect unfavourably the distribution of income

through many channels. The most discussed concern job polarisation, (i.e. increased demand for high-level, highly paid jobs on the one hand and low-level, poorly paid jobs on the other), reduction in job quality at the lower-skilled end, and also greater difficulty for lower-skilled workers to adjust to change and find new jobs, with longer periods of unemployment than those with higher qualifications and skills. These potentially negative consequences on the labour market have a geographical footprint, as regions and subregions that are already experiencing greater difficulties in terms of unemployment and low level of skills are likely to be the ones suffering most, if no action is taken. We clearly need to monitor closely and research the multiple impacts of AI on the economy in the coming years. We also need to consider a more proactive strategy to build the resilience of regions across Europe with a particular regard to the most vulnerable ones. We return to this in *Chapters 12 and 13*.

When it comes to cybersecurity, AI is a double-edged sword: it can be greatly beneficial to increase the security of devices, systems and applications, but can also empower those who seek to attack systems and networks and thus become an advanced tool in the arsenal for cyber-attacks. Moreover, the robustness of AI against malicious action itself becomes an issue, posing the most immediate danger for the security of cyber-physical systems, in which AI will be increasingly deployed. There are lines of research focusing on understanding the specific vulnerabilities of AI and related attacks, and ways of increasing AI robustness and interpretability and safety by design. We also need shared, large, high-quality datasets to train and test algorithms, and agreed frameworks to evaluate them (*Chapter 10*).

We discuss at several points in the report that the major progress we are seeing in the development of AI is linked to the rapid advances in computing and in data availability. With respect to computing, we are starting to see a paradigm shift due to the recognition that the increasing energy consumption of data centres and data transfers will become unus-

tainable in the era of the Internet of Things and 5G networks. Data traffic and processing loads are likely to be unprecedented when billions of additional devices will be connected to the internet, sensing the environment, and constantly sending and receiving data. With this in mind, we are seeing a trend towards more decentralised frameworks of edge and fog computing where processing is done closer to the sensors capturing and displaying data, including mobile phones and tablets. Industry has started moving in this direction but there is still a window of opportunity for European investment, regulatory frameworks, and standards to shape these developments so that they benefit Europe (*Chapter 11*).

With respect to digital data, we review its key economic characteristics including economies of scale, of scope, and non-rivalry. The first two point to an increase in the concentration of data, and hence information and power, in the hands of a few actors in the internet economy, while non-rivalry creates potential tensions between opening access to the data so that society benefits the most, and restricting access so that the data holder benefits most. These tensions are reflected in the legal framework, as discussed earlier, and make the development of data strategies in an open and globalised environment a particular challenge. We argue, however, that if we apply the lessons of successful internet companies to the European public sector and develop ecosystems based around public platforms, it is possible to create large pools of shared data that are semantically well structured and labelled, and can fuel new AI applications in different domains. In this way we can open access to the data, develop the market, serve the public, and enrich the data commons at the same time (*Chapter 12*).

We conclude our multi-perspective analysis of AI focusing on the concept of resilience which is useful to frame a European approach to AI addressing different phases: prevention, anticipation, preparation, but also adaptation and transformation to bounce back from the effects of a shock. Seen from this perspective, AI could not only be an engine for growth and change in Europe but also become

an opportunity to bootstrap social and economic development in peripheral regions, leveraging the richness of European diversity and culture to the full (*Chapter 13*).

We put forward some elements of a possible approach centred on the network of Digital Innovation Hubs. There are already several hundred hubs which are expanding rapidly with a target of one for each region in Europe. The Communication on AI (EC, 2018a) foresees a specialised subset of these hubs on AI to facilitate access to technology and know-how to public administrations and firms, SMEs in particular, in combination with the planned AI-on-demand-platform. We suggest putting these hubs at the centre of local ecosystems comprising public administrations, local enterprises, educational and training establishments, and civil society.

These ecosystems could create local pools of shared data among the key actors so that the AI skills developed/upgraded locally could be put to good use in developing algorithms and solutions based on local data to address local needs. Each European region has its own specific mixture of priorities in relation to its environment, economy, demography, health, and so on. Therefore, this approach could provide an excellent opportunity to harness local creativity, culture, and knowledge of the territory to create socially relevant and people-centred AI and develop diverse and inclusive AI systems (*Chapters 12 and 13*).

From this multi-disciplinary analysis, we draw the following conclusions:

We are only at the beginning of a rapid period of transformation of our economy and society due to the convergence of many digital technologies. AI is central to this change and offers major opportunities to improve our lives.

There is strong global competition on AI among the USA, China and Europe. The USA leads for now but China is catching up fast and aims to lead by 2030. For the EU, it is not so much a question of winning or losing a race but of finding the way of embracing

the opportunities offered by AI in a way that is human-centred, ethical, secure, and true to our core values.

The EU Member States and the European Commission are developing coordinated national and European strategies, recognising that we can only succeed together. We can build on our areas of strength including excellent research, leadership in some industrial sectors like automotive and robotics, a solid legal and regulatory framework, and very rich cultural diversity at both regional and subregional level.

It is generally recognised that AI can flourish only if supported by a robust computing infrastructure and good quality data:

- With respect to computing, we identified a window of opportunity for Europe to invest in the emerging new paradigm of computing distributed towards the edges of the network in addition to centralised facilities. This will also support the future deployment of 5G and the IoT.
- With respect to data, we argue in favour of learning from successful internet companies, opening access to data and developing interactivity with the users rather than just broadcasting data. In this way, we can develop ecosystems of public administrations, firms, and civil society enriching the data to make it fit for AI applications responding to European needs.

We should embrace the opportunities afforded by AI but not uncritically. The black box characteristics of current ML make them obscure even to specialists. AI systems are currently limited to narrow and well-defined tasks, and their technologies inherit imperfections from their human creators, such as the well-recognised bias effect present in data. We should challenge the shortcomings of AI and work towards strong evaluation strategies, transparent and reliable systems, and good human-AI interactions.

Ethical and secure-by-design algorithms are crucial to build trust in this disruptive technology, but we also need a broader engagement of civil society on the values to be embedded in AI and the directions for future development.

This social engagement should be part of the effort to strengthen our resilience at all levels from local, to national and European, across institutions, industry and civil society. Developing local ecosystems of skills, computing, data, and applications can foster the engagement of local communities, respond to their needs, harness local creativity and knowledge, and build a human-centred and socially driven AI.

We still know very little about how AI will impact the way we think, make decisions, relate to each other, and how it will affect our jobs. This uncertainty can be a source of concern but is also a sign of opportunity. The future is not yet written. We can shape it based on our collective vision of what future we would like to have. But we need to act together and act fast.

“Uncertainty about the impact of AI can be a concern but is also an opportunity: the future is not yet written. *We can, and should, shape it.*”



**PART 1:
INTRODUCTION AND
INTERNATIONAL AI
LANDSCAPE**

MOTIVATION AND OBJECTIVE OF THIS REPORT

Artificial Intelligence (AI) will bring many opportunities...

AI is a technological field that powers many applications we use daily, often without knowing it. As it develops further, it has the potential to bring many changes to both economy and society. This might unfold in a similar way to the invention of the World Wide Web 25 years ago which, combined with the subsequent diffusion of mobile phones, opened the way to the massive increase the use of the internet across the world.

Looking back over these 25 years, we can see how early views about the death of distance (*Economist*, 1995), and the demise of cities (Gilder in Karlgaard and Malone, 1995) as a result of the internet were vastly exaggerated: cities have continued to grow and in fact increased their role as the hubs of innovation. The internet has profoundly changed the way we work, learn, and socialise, created entirely new products, services, and business models, and reshaped international relations and markets. We have seen new challenges emerging, including cyber-threats, misinformation, loss of privacy, and the displacement or loss of jobs in many sectors, but we have also seen many benefits such as ubiquitous access to information and services.

The recent developments in AI are the result of increased availability of processing power, improvements in algorithms and the exponential growth in the volume and variety of digital data, and increased funding. We use AI today in many

contexts, e.g. to guide us to a destination, help us find the next song to listen to, filter our spam e-mails, or help us texting messages. In the coming years, AI-supported medical diagnostics may become increasingly common. Evaluating job candidates or predicting reoffending could be supported by the responsible use of transparent algorithms helping to overcome personal or institutional biases (see, for example, Faliagka et al., 2012 and Tollenaar et al., 2013). Robots able to understand the complexity of social and human behaviour have the potential to improve human cognitive performance (Belpaeme et al., 2018), and support the development of skills, e.g. in children with autism (Scassellati et al., 2018).

...and many challenges

AI can solve a wide range of problems, but we still know very little about how it will impact the way we think, make decisions, or relate to each other (Gomez et al., 2018). AI is still not well understood even by specialists. AI systems are currently limited to narrow and well-defined tasks, and their technologies inherit imperfections from their human creators, such as the well-recognised bias effect present in data and algorithms. We should therefore be aware of the potential shortcomings of AI and work towards agreed ways to evaluate algorithms, build transparent and reliable systems, and design good human-AI interactions.

Why this report...

Hundreds of reports are published each year on AI and related technologies, in addition to scientific papers. This clearly shows the high level of interest in this subject, but do we really need another report?

Most of these publications come from industrial or national interests. Therefore, an independent European perspective underpinned by original research and not focused on a single policy domain is missing. Hence the value of this report from the Joint Research Centre (JRC), the science and knowledge management service of the European Commission (EC). This report is also timely because many European countries are developing national strategies for AI (see Chapter 4) to embrace this technological field and prepare society for the changes it will bring. The EC is working with the Member States on a coordinated Action Plan for AI. A High-Level Expert Group on AI has also been established by the EC to make recommendations on policy and investment, and set guidelines on the ethical development of AI. The latter are crucial as there is a strong consensus in Europe that our AI should be human-centric, with ethics and security by design, and supporting the core values of the European Union: human dignity, freedom, democracy, equality, rule of law and human rights.

...and its objective

To provide a balanced assessment of opportunities and challenges for AI from a European perspective, and support the development of European action in the global AI context.

What is in the report

We have organised the report into two sections: in **Part 1** we give a brief introduction to AI, and analyse the EU positioning in the global AI landscape, including key features and leading players, research and technological capacity, and areas of strength. We then consider a few examples of Member States' policies and strategies to assess points of synergy and where European action can best add value. This is followed by an analysis of AI in China since this country has set out, and could

well achieve the objective of becoming the dominant AI player in the world by 2030, overtaking the USA. Whilst we are more familiar with USA companies, technologies and applications, we know little about developments in China and for this reason we pay comparatively more attention to China in this report, taking advantage of bespoke research done on this subject for the JRC.

In **Part 2** we look at AI from multiple perspectives: ethical, legal, educational, economic, cybersecurity, computing and energy, data-related, and societal resilience. We discuss each perspective in turn but they are intertwined in multiple ways and should be considered all together in framing a European way to the development and use of AI.

What is not in the report

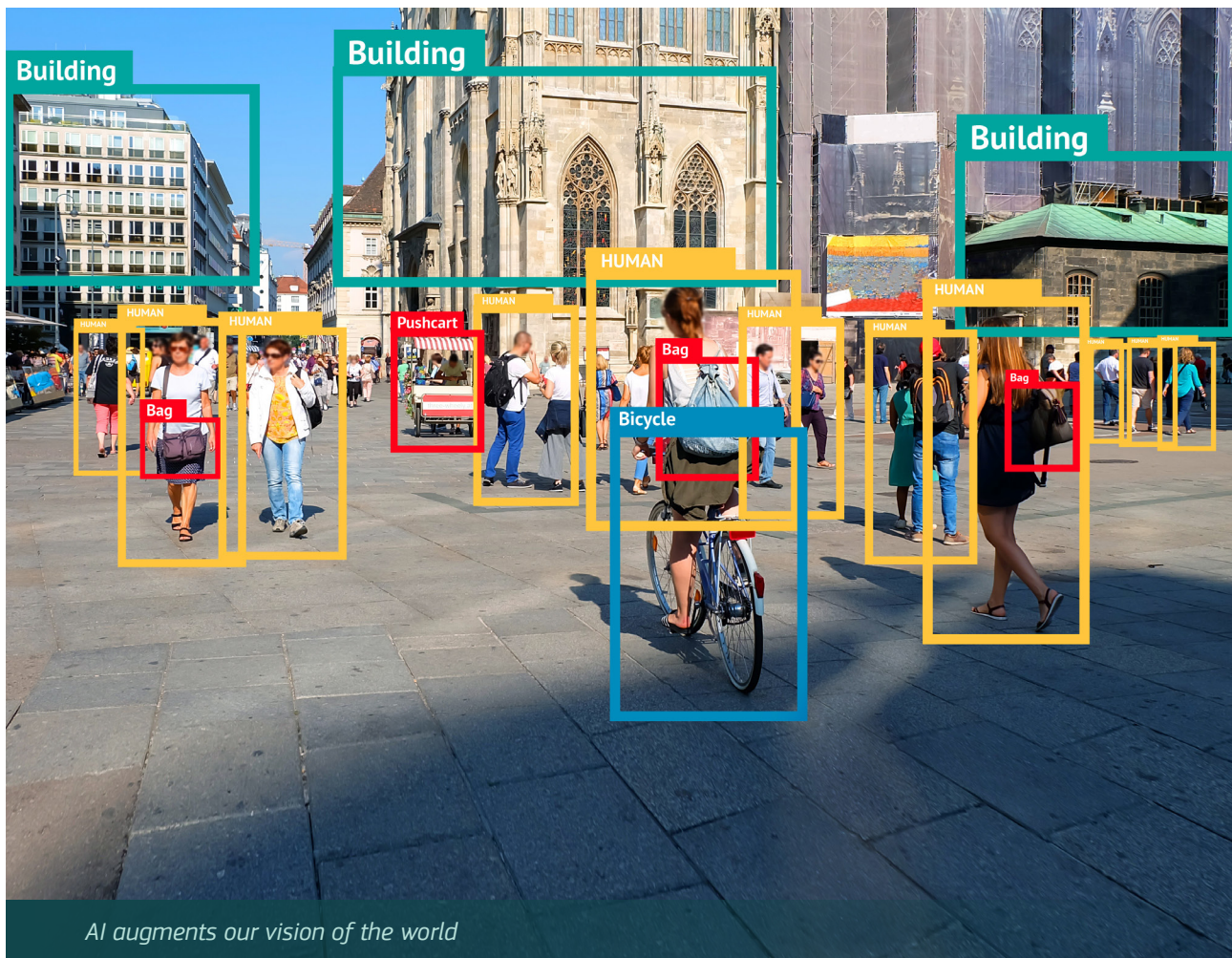
The report does not include any analysis of AI applications in different sectors of the economy. This will be the subject of a separate JRC report in preparation, as well as the focus of an AI-Watch facility to monitor and analyse AI developments launched in December 2018. Other JRC reports on resilience, fairness, mobility, cybersecurity, hybrid threats, and China will be published soon, complementing and adding details to some of the arguments addressed in this report.

What you should take away

Currently, the global competition on AI is largely between the USA and China at the present time. For the EU, it is not so much a question of winning or losing a race but of finding the way of embracing the opportunities offered by AI in a way that is human-centred, ethical, secure, and true to our core values.

We have many areas of strength, including excellent research, extremely rich cultural diversity, and leadership in some industrial sectors, such as automotive and robotics.

We should therefore not feel intimidated or fearful about the future. However, we need to act and shape this future because EU countries are stronger together in this global competition than acting alone.



AI augments our vision of the world

SUMMARY

AI is a generic term that refers to any machine or algorithm that is capable of observing its environment, learning, and based on the knowledge and experience gained, taking intelligent action or proposing decisions. There are many different technologies that fall under this broad AI definition. At the moment, ML techniques are the most widely used.

In this chapter, we introduce the basic concepts of AI and highlight recent developments. The opportunities opened up by AI are many, and in some cases not yet foreseen, but so are the challenges. These will be explored in particular in Part 2 of the report, but in this chapter, we highlight that the algorithms powering AI display all the characteristics of a black box: we can access the inputs and outputs, but do not understand fully what happens in-between, and how certain outputs, including decisions and actions, are derived. Given this major drawback, it is imperative to **increase our scientific understanding, develop transparent mechanisms to assess the quality and performance of AI so that as a society we are able to build and retain trust in this disruptive technology.**

ABOUT AI

Traditionally, Artificial Intelligence (AI) refers to machines or agents that are **capable of observing their environment, learning, and based on the knowledge and experience gained, taking intelligent action or proposing decisions.**

To date, there is no agreed definition of what constitutes AI. When reflecting of AI and machines, the general public often thinks about humanoid robots which have always been part of the public's perception of intelligent machines, as popularised in science fiction novels and films. In fact, robotics is only one subfield of AI and makes use of many ML techniques, including neural networks.

The origins of AI date back to the 1940s and 50s, but it is only recently that it has entered into our everyday life, from filtering out the spam in our mailbox (see Box 1), to automatically tagging our loved ones in family photos, to real-time machine translation, and many more applications in business, industry, defence, and social media.

BOX 1. Early success case of ML in cybersecurity: Spam filtering

We are all familiar with the sense of annoyance when our inbox gets filled with junk mail or spam. In the early 2000s, the application of a ML technique called *Bayesian Networks* brought a new generation of *more effective e-mail filters* that were able to adapt themselves to their environment, learning to detect spam e-mail by analysing large quantities of samples. Unsolicited e-mail has become not just a nuisance but a *potential security threat* as mail messages are used to distribute malware, conduct phishing campaigns or carry out targeted attacks. Modern e-mail filters continue to fight these threats employing a *new generation of algorithms and AI methods* to create more effective and resilient countermeasures.

Although AI has a long history of development, recent major breakthroughs are due to the convergence of increased computing capacity, availability of data, and new algorithms.

2.1 A brief history of AI

There have been ups and downs in the history of AI with *logic-based* approaches in the 1950s and early 60s, *knowledge-based expert systems* in the 1970s and 80s, and *data-driven* approaches (from 2000 onwards) with periods of disillusionment and reduced funding in-between. We are now at the beginning of a new phase of high expectations, fuelled by vastly increased computing processing capabilities and data. Their combination supports new developments such as Machine Learning (ML)

that has emerged as a subfield of AI to predict a behaviour from many examples given as input.

ML represents a paradigmatic shift in computing. Traditionally, a programmer would write computer code setting the rules needed to process data inputs to get an answer as output. In ML, the **computer receives input data as well as the answers expected from the data, and the ML agent needs to produce the rules** (see Fig. 1). These rules can then be applied to new data to produce original answers. **An ML system is trained rather than explicitly programmed.**

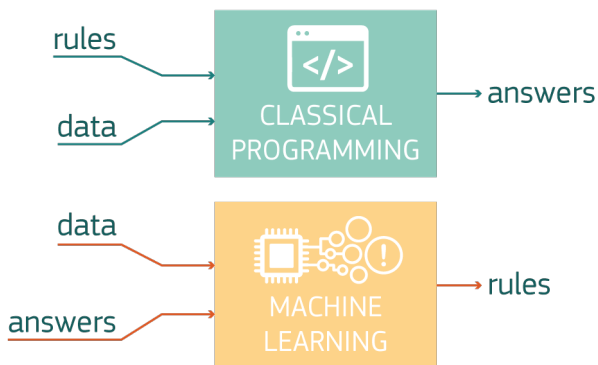


Figure 1. Paradigm shift in ML²

Source: after Chollet, 2017

2.2 Recent developments in ML

The recent interest in AI, and ML in particular, results from three parallel developments. First, increasingly realistic computer games have required specialised graphic processors. When the PC graphics card manufacturer Nvidia published the CUDA programming interface to its graphics accelerator cards in 2007, it became possible to take advantage of fast parallel computing at low cost.

Second, very large amounts of data have become available as computers and their users have been networked. The digitalisation of images, videos, voice and text has created an environment in which ML thrives. This has allowed AI researchers to re-

visit old artificial neural network models, training them with very large datasets to address some hard problems such as object recognition from digital images and machine translation. Whereas it was previously believed that computers needed to understand language and its structures before they could translate text and speech from one language to another, for many practical uses it is enough to process millions of sentences to find out the contexts where words appear and make translations without linguistic knowledge. This has considerably speeded up the process, leading to real-time translation applications now available on most mobile phone platforms.

Third, there have been new advances in ML algorithms. In addition, specialised open source ML software libraries have become available that make the creation and testing of ML algorithms easier. These trends started to come together around 2012 (see Box 2)

Whilst ML is the generic class of algorithms that learn from the data, their accuracy depends very much on the quality of the training dataset, and how well they have been structured, semantically labelled, and cleaned by humans to make them representative of the problem to tackle, and reduce the number of parameters in the data.

BOX 2. Turning point in image recognition

In 2012, a multilayer network trained using Nvidia's graphics processor cards showed outstanding performance in a well-known image recognition competition. The competition was based on the ImageNet database that contains about 14 million human-annotated digital images. The ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is now one of the main benchmarks for progress in AI. Its object detection and classification challenge uses 1.2 million images for training, with 1 000 different types of objects. In 2017, the best neural networks were able to guess the correct object category with 97.7 per cent 'top-5' accuracy, meaning that the correct object class was among the five most probable classes as estimated by the network.

- *Deep Learning* (DL) is a subset of ML that can cope with noisier data by increasing significantly the number of neural layers and neurons³, and the amount of data used for the training. DL has facilitated object recognition in images, video labelling, and activity recognition, and is making significant inroads into other areas of perception, such as audio, speech, and natural language processing.

- *Reinforcement Learning* (RL) is another set of algorithms that focus on experience-driven sequential decision-making, i.e. they make software agents take action to maximise some notion of cumulative reward. RL combined with DL is the basis of many recent successes in complex games, such as Go, Poker and Dota, where computers have been able to beat leading human experts in the field. Practical applications of these algorithms are starting to be applied in many domains, such as autonomous driving, unmanned aerial vehicles, stock markets, and defence.

In addition to image and text recognition and gaming, the field of human-computer interaction has seen many impressive improvements recently as a result of the AI applications. A long-standing goal has been to enable people to have a natural conversation with computers, as they would with each other. In recent years, the application of DL and the combination of multiple AI components, including natural language processing, voice synthesis and voice translation, have revolutionised the interaction we have with computers, as discussed in Section 2.3 below and exemplified by the virtual assistants now available on most mobile phone platforms, such as Apple's Siri, Microsoft's Cortana and Amazon's Alexa.

The extent of progress in computer-generated speech is such that even in singing it is increasingly difficult to distinguish a human from a machine (see Box 3).

BOX 3. DL for singing processing

The singing voice has traditionally been a challenging instrument to analyse and synthesise given its expressive character and the variety of timbre and expressive resources that characterise styles and singers. However, recent data-driven ML techniques, especially DL, have substantially boosted the quality and accuracy of singing processing techniques.

Singing synthesisers might soon make them indistinguishable from human singing. As ML algorithms can also compose new music and imitate the style of classic composers of the past, we may soon be confronted with a major shift in our understanding of what music is, how we experience it, as well as new models of music production, distribution and consumption (Gomez et al., 2018).

Real-time translation of both text and voice have been developed by Google and Baidu, while Microsoft researchers have created algorithms that read a document and answer questions about it almost as well as a human. This is a major milestone in the push to have search engines such as Bing and intelligent assistants such as Cortana interact with people and provide information in more natural ways, much as people communicate with each other.

2.3 Recent developments in social robots

Advances in robotics have allowed AI to be embedded in physical entities that take into consideration the characteristics of their 'body' as well as those of the physical and social environment in which they operate. This raises the potential for socially intelligent robotic agents to co-exist with humans, as in the case of co-workers, personal companions or self-driving vehicles.

Advances in perception systems make it possible to translate low-level signals into high-level abstract concepts. For instance, natural language processing (NLP) techniques connect sounds from speech with language semantics. This contributes to creating a perception-cognition-action loop for autonomous

robots that makes them able to understand and interact with the surrounding environment.

These developments in robotics bring additional perspectives to the integration of AI in our everyday life: (i) the agents are becoming increasingly capable of perceiving the context within which they act by developing an understanding of the human physical environment, including complex human social behaviour; (ii) the embodied form of AI can physically act upon and shape human environment and potentially affect human behaviour as well as human to human social interaction not only at the individual level but also in small groups; and (iii) the interaction between humans and intelligent robotic systems increasingly involves human emotions which might stimulate empathic behaviour towards artificial systems and have an impact on human development over time.

In this context, a relatively new strand of research has emerged, which investigates the ways in which humans interact with robots and how various robot behaviours and designs affect human cognitive and socio-emotional development. A seamless human-robot interaction has the potential to contribute to transparent and explainable AI, and for vulnerable populations too (e.g. children and the elderly), because of its intuitive nature even with minimalist design characteristics.

For this reason, while the field of robotics is advancing, researchers in the field of human-robot interaction create scenarios in which they test various robot characteristics in real-life settings. In this way, they investigate the impact of socially intelligent robots on human cognitive and socio-emotional development by involving multiple stakeholders. This type of research allows society to actively shape the ways in which robots might be integrated into our everyday life and contribute to a more inclusive and sustainable use of AI.

Whilst these developments are important because they bring AI into our everyday experience, they represent what is often defined as the *Narrow AI*.

General AI, which would involve the development of consciousness, is many years away. The current consensus of the private-sector expert community is that general AI will not be achieved for several decades (NSTC 2016a).

2.4 Current challenges

Each specialised sub-area of AI has its own specific challenges, but here we focus on those related to algorithm design and evaluation in ML as this is one of the main areas of development, as discussed above. Other challenges, including ethical, legal and cybersecurity, are discussed in Chapters 6, 7 and 10, and by different scientists (see also Gomez et al., 2018).

“We need to advance the explainability, accountability and transparency of algorithms in general and DL systems in particular.”

Nowadays, many ML techniques display some characteristics of a ‘black box’ model, i.e. we know what goes in and what comes out of the algorithm, but we do not have a full understanding of its inner workings. This limits the scientific understanding of algorithms, the capability to recover from adversarial examples, and complicates human supervision in practical applications (see Chapter 10 on cybersecurity).

We need to define evaluation frameworks that are meaningful and in realistic settings to match practical application contexts. In this respect, we should consider engineering best practices, impact assessment methods, user satisfaction and business metrics, in order to develop smart and transparent benchmarking strategies. We should train the next generation of ML developers to apply and communicate these strategies and follow best practices for AI evaluation.

We need to advance the explainability, accountability and transparency of algorithms in general, and DL systems in particular, both from the perspective of ML research and from that of users in different applications. We need to extend data and algorithmic literacy across society to increase the ability of critical thinking with respect to machine intelligence and challenge it when necessary.

Further research is also needed to understand fairness in the context of automated decision-making. An algorithm or decision is fair when it does not discriminate against people because of their membership to a specific group (such as gender, race, sexual orientation). It is unbiased if it is not affected by the membership of a person to a protected group. Attempts to formalise algorithmic discrimination leads to a long list of fairness criteria that are always harmful to one group of stakeholders and beneficial to the other, making it difficult to arrive at an agreed set. Moreover, these criteria are helpful to identify if there has been discrimination in decision-making, but they do not explain the source of unfair treatment. This is best done by looking at biases in the training data, its labelling,

and the algorithm development process which may encode biases by the development team or context.

■ 2.5 Summary and conclusions

AI is a generic term that refers to any machine or algorithm that is capable of observing its environment, learning and, based on the knowledge and experience gained, taking intelligent action or proposing decisions. There are many different technologies that fall under this broad AI definition. At the moment, Machine Learning techniques are the most widely used.

The methodological developments in AI date back more than 50 years but the recent breakthrough has come from advances in computing power, availability of data, and complexity of the algorithms. Now the many applications of AI are entering into our everyday lives, from filtering spam, to machine translations, image recognition, and music generation, and are increasingly used in industry, government and commerce. We are probably only at the beginning of this process because the development of ubiquitous sensor networks, and the IoT, will increase exponentially the sensing capabilities of AI, the volumes of data available to train the algorithms, and the reach of AI through decisions and actions.

The opportunities are many, and in some cases not yet foreseen, but so are the challenges. These will be explored in particular in Part 2 of the report, but in this chapter we have highlighted that the algorithms powering AI display all the characteristics of a black box: we can access the inputs and outputs, but do not understand fully what happens in-between, and how certain outputs, including decisions and actions, are derived. Given this major drawback, it is imperative to increase our scientific understanding develop transparent mechanisms to assess the quality and performance of AI so that we are able as a society to build and retain trust in this disruptive technology.



Europe has excellent research in AI and leading industrial sectors on which to build

SUMMARY

The overview of the global and European AI landscape we report in this chapter is the fruit of a comprehensive analysis of over 35 000 key players worldwide. The analysis confirms the intense competition in AI taking place worldwide with three main leaders: the USA, Europe, and China. Each region has about one quarter of all key players in the AI field, including both research and industry, but a distinctive different mix of players: while Europe is well balanced in the number of research and non-research players, the USA has approximately three times as many industrial/corporate players as research ones, and China has about six times as many research players as industrial ones. The strength of the corporate world in the USA is also indicated by its dominance in the number of start-ups (almost half of the total worldwide) and venture capital (more than one third of the total). China, on the other hand, is making a concerted effort to turn research into patents, and accounts for almost 60% of the world total.

Europe is currently in a good position regarding the quality of its research, with more than 30 % of top AI publications, just behind the USA (33 %). This result is the fruit of the high level of research funding and the intense inter-country collaboration provided by the framework programmes. Key areas of strength in Europe on which to build upon are automated and connected vehicles, and robotics. Continued policy support seems key maintaining and improving Europe's position in an ecosystem which is rather polarised around a few big powers.

EU IN THE AI COMPETITIVE GLOBAL LANDSCAPE

■ 3.1 The international policy context

Dutton (2018)⁴ has identified at least 21 countries that have launched or are preparing national AI strategies and programme in 2017-2018, of which 8 are in Europe, 3 in North America, and the rest in Asia and Oceania. Here we summarise the developments in the USA, Canada, India, Japan, and South Korea, leaving Europe to Chapter 4, and China to Chapter 5.

■ 3.1.1 USA

In October 2016, the Obama administration published three reports that concluded that AI technology will be the driving force behind transformations across both the economy and national security. One of the reports (NSTC, 2016c) proposed a national AI R&D strategy linked to some other initiatives with six main directions:

1. Make long-term investments in AI research.
2. Develop effective methods for human-AI collaboration.
3. Understand and address the ethical, legal, and societal implications of AI.

In the context of strong global competition on AI, the EU is well placed if it builds on its strengths, like research and robotics, and works together.

4. Ensure the safety and security of AI systems.
5. Develop shared public datasets and environments for AI training and testing.
6. Measure and evaluate AI technologies through standards and benchmarks.

The current administration seems to be supporting the unrestricted development of AI, and in the May 2018 Summit on AI at the White House with leading industrialists Michael Kratsios, Deputy Assistant

to the President for Technology Policy, announced the government's goals: (1) maintain American leadership in AI, (2) support the American worker, (3) promote public R&D; and (4) remove barriers to innovation. To achieve these objectives, Kratsios announced a new Select Committee on Artificial Intelligence to advise the White House on interagency AI R&D priorities and to consider the creation of Federal partnerships with industry and academia.⁵

In September 2018, DARPA, the government defence research agency, announced an investment plan of over US\$ 2 billion to overcome some of the perceived limitations of current AI technologies, i.e. their dependence on large amounts of high-quality training data, poor ability to adapt to changing conditions, limited performance guarantees, and inability to provide users with explanations of their results. The 'AI Next' programme seeks to explore new theories and applications that could make it possible for machines to adapt to changing situations developing contextual reasoning capabilities.⁶

These are signals of continued support from the Trump administration towards the development of AI, which will also be boosted by an undisclosed part of the large USA defence budget. Compared to other regions of the world where government plays a major part in both policy and investment, in the USA, private companies play a central role in the development of technology in general as well as AI.

The USA stock market was worth some US\$ 34 trillion in April 2018, 43 % of the total world markets, compared to Europe's 15 %, and Asia and Japan 15 %.⁷ Tech companies account for one quarter of this value⁸, and the top five (Apple, Amazon, Google/Alphabet, Microsoft and Facebook) account for some US\$ 4 trillion⁹, or 11 % of the total USA market, and 33 % of the total valuation of the European market. Not surprisingly, the R&D investment of companies like Amazon and Alphabet is very significant, at US\$ 16.1 billion and US\$ 13.9 billion respectively in 2017 (Knight, 2018), much more than the civilian investment of the government. In comparison, the total budget for the National

Science Foundation, the main agency funding R&D on AI together with DARPA and the Department for Transportation (in the specific area of autonomous and unmanned systems), amounts to US\$ 5.3 billion in the 2019 budget.¹⁰

The power of the USA tech giants is of course not just financial, but more crucially about the vast quantities of data they are able to gather and analyse from all over the world. Google accounts for 86 % of all internet searches in the world,¹¹ processing some 40 000 queries per second, or 3.5 billion per day.¹² Facebook has some 2.2 billion active users per month.¹³ It is thus the largest content manager in the world, although it does not produce any of its own data. All the data, and more importantly for AI training, the context for the data, is provided by its customers in exchange for the service provided by the platform. Its Like button has turned the company into the largest recommendation platform in the world, able to profile and target small groups and individuals for both marketing, and as seen recently with the Cambridge Analytica case, also for political purposes. As indicated in Chapter 2, the development of AI rests largely on the ability to have large volumes of well-structured and labelled data, and in this respect the USA giants have a clear lead in the West.

3.1.2 Other countries

Canada: Canada announced its AI strategy in the 2017 budget, which allocates CAN\$ 125 million over five years to achieve four main objectives:

- To increase the number of outstanding AI researchers and skilled graduates in Canada.
- To establish interconnected nodes of scientific excellence in Canada's three major centres for AI in Edmonton, Montreal and Toronto.
- To develop global thought leadership on the economic, ethical, policy and legal implications of advances in AI.

BOX 4. Some examples of AI developments in the USA

Google is an AI world leader. It researches and develops AI products and services, has developed its own AI chip, the Tensor Processing Unit (TPU), and created TensorFlow, one of the most widely used open source AI/ML libraries. DL algorithms are already a considerable part of several Google products including recommendation systems, Android, Gmail, Maps, and translation services. The number of DL directories Google uses has gone from zero in 2012 to 4 000 in 2016.¹⁴ In May 2018, Google announced **Google duplex, an AI system for accomplishing real-world tasks over the phone.** Google, and other large operators, also use AI extensively to manage their data centres and save energy.¹⁵

Facebook is the biggest social media company in the world. It has established a Facebook AI research (FAIR) team which is one of the most advanced, with several labs around the world. Facebook supports the open source AI libraries **PyTorch** and **Caffe2** that compete with **Google TensorFlow.** The company has also released plans to develop its own AI chips.

Amazon uses AI extensively in its recommendation systems and logistics. It also offers both consumer and business-oriented AI products and services. **Amazon Echo** brings AI into the home through the intelligent voice server, Alexa. For business,

Amazon AI empowers fake reviews detection, chatbots, product recommendations, big data management, etc. **Amazon Web Services** are probably the biggest cloud system in the world providing infrastructure as a service to over **1 million users.**¹⁶

Microsoft has also been investing heavily in AI and infrastructure as a service. **Microsoft Azure** is one of the three biggest cloud (and AI) providers together with **Google Cloud** and **Amazon web** services. Cortana is the Microsoft virtual assistant that competes with Amazon Alexa, Apple Siri, Google Duplex and others.

Apple is the most valuable company in the world worth over **US\$ 1 trillion.** Apple's main AI divisions are **Siri** team, and the **Core ML** team. Siri's team focus is on NLP and computer vision, both of which are necessary to power voice assistant features and new, more cutting-edge technology such as augmented reality apps that rely on object recognition. Core ML is the **ML Application Programming Interface (API)** that Apple launched last year to help AI tasks and AI-focused apps and services from third-party developers run more efficiently on Apple devices.

- To support a national research community on AI.

The strategy is led by the Canadian Institute for Advanced Research¹⁷ in close partnership with the Canadian government and the three new AI Institutes: the Alberta Machine Intelligence Institute in Edmonton, the Vector Institute in Toronto, and MILA in Montreal. Compared to other national strategies that have a strong emphasis also on sectoral developments, data, and supporting industry, the Canadian strategy is primarily focused on research.

India: A discussion paper for an AI strategy was published in June 2018 by NITI Aayog (a government

think-tank) focusing very much on social inclusiveness: #AIforAll.

The strategy follows three distinct, yet inter-related components:

- a) Opportunity: the economic impact of AI for India.
- b) AI for Greater Good: social development and inclusive growth.
- c) AI Garage for 40 % of the world: solution provider of choice for the emerging and developing economies across the globe.

In relation to c) the strategy argues that India provides a perfect ‘playground’ for enterprises and institutions globally to develop scalable solutions which can be easily implemented in the rest of the developing and emerging economies. NITI Aayog focuses on five sectors that are envisioned to benefit the most from AI in solving societal needs, and where only private-sector-led initiatives may not lead to the desired societal outcomes:

- a) **Healthcare:** increased access and affordability of quality healthcare,
- b) **Agriculture:** enhanced farmers’ income, increased farm productivity and reduction of wastage,
- c) **Education:** improved access and quality of education,
- d) **Smart Cities and Infrastructure:** improved connectivity for the rapidly increasing urban population, and
- e) **Smart Mobility and Transportation:** smarter and safer modes of transportation and better traffic and congestion management.

The strategy makes some 30 recommendations. With respect to research, it proposes a two-tier system: new Centres of Research Excellence in AI will focus on fundamental research and act as technology feeders for the International Centres for Transformational AI, which will focus on creating AI-based applications in domains of societal importance.

To facilitate adoption of AI, the strategy recommends in particular:

- 1) Establishing a market place for AI, and
- 2) Facilitating the creation of large foundational annotated data sets.

Although the discussion paper acknowledges that significant public investment would be needed to implement the strategy, it does not put a figure on it, leaving the matter to the political discussion.

Japan: In 2015, the Japanese government announced a five-year plan, placing AI and robotics as the stepping stone of its renewed strategy for science, technology and innovation. Japan’s approach to artificial intelligence is a combination of the USA and Chinese models. It is based on the 5th Science and Technology Basic Plan (2016–2020) of the Japanese Council for Science, Technology and Innovation. The plan’s objective is to establish Japan as a super smart society (i.e. Society 5.0). The programme is headed by the prime minister, and its field-specific programmes are headed by selected corporate heads. In April 2016, the government established a Strategic Council for AI Technology promoting research and development of AI technology, coordinating with industries related to the industries that utilise AI (so-called ‘exit industries’), and moving forward with the social implementation of AI technology. The Council coordinates the activities of three leading research centres:

- 1) The Centre for Information and Neural Networks and Universal Communication Research Institute of the National Institute of Information and Communications Technology.
- 2) The RIKEN Centre for Advanced Intelligence Project of the Institute of Physical and Chemical Research.
- 3) The Artificial Intelligence Research Centre of the National Institute of Advanced Industrial Science and Technology.

The AI Technology Strategy was published in March 2017. It identifies four priority areas for the development and integration of AI: increased productivity through user-driven hyper customisation of services, medical healthcare and welfare to support an increasingly ageing population, mobility to support

safe and environmentally friendly travel for all, and information security. What is particularly interesting in this strategy is that each of these four priorities is driven by an image of the society Japan is aiming for. A clear example of socially driven AI. The strategy is articulated in three phases:

- 1) Utilisation and application of data-driven AI developed in various domains (until 2020),
- 2) Public use of AI and data developed across various domains (2020-2025), and
- 3) creation of an ecosystem built by connecting the multiplying domains (2025-2030).

The three centres identified above will act as lead for each of the priority areas identified, and act as hubs for R&D in partnership with industry and academia.

The strategy pays particular attention to the collection, management and sharing of data across government, academia and the private sector. With respect to the latter, it recognises that there are still many issues to be solved, including the handling of personal data, but has set up a Data Distribution Acceleration Working Group to help share best practice.

South Korea: In 2016, the government introduced a mid-to-long-term Master Plan for the Intelligent Information Society. One of the plan's goals is to foster the 'establishment of the basis for AI technology'. It announced it would spend 1 trillion won (US\$ 840 million) by 2020 to boost the artificial intelligence industry. The government plans to support the establishment of a high-profile research centre that will serve as a 'pivot' of the nation's research and development in the AI field. Samsung and LG Electronics, SKT, KT, Naver, and Hyundai Motor have decided to join the initiative. Each company plans to invest about 3 billion won (Marquart, 2016). In May 2018, the government added another 2.2 trillion won to its AI strategy focusing on three priorities: training an additional 5 000 AI scientists, developing applications in the areas of national defence, medicine, and public safety and funding infrastructure and incubators for start-ups.

■ 3.2 Analysing the key features of the AI landscape

The field of AI is experiencing a period of intense progress, due to the consolidation of several key technological enablers: faster processing, vastly increased amounts of data, and better algorithms. Because AI is expanding so quickly across multiple sectors, and there has yet to be a single and agreed definition in which technologies and applications are included as 'AI', it is not possible to track its evolution through official statistics at this stage, as would be done for more-established fields.

“Agriculture, healthcare, mobility and security are among the most promising applications of AI for many countries.”

BOX 5. The Techno-Economic Segment (TES) approach in a nutshell

The first step in the methodology is to define the boundaries of the TES (AI in this case), by detecting players that focus on AI as their primary or secondary activity. Players are defined as **research centres, academic institutions, and companies** which have participated in one or more of the following economic activities: **R&D processes, industrial production and marketing, specific AI-related services**. The players' activities may be explicitly stated, e.g. in the description of the companies' activities in business registers or segment specific firms' repositories, or derived from the analysis of their R&D activities, e.g. text from patents, conference proceedings and research projects. To carry out this analysis, we developed a comprehensive dictionary of domain specific keywords covering the technological aspect of the TES (in this case AI). A selection of the

keywords that helped to identify the highest number of AI-related players and R&D activities is shown in Figure 2. Once we defined the boundaries of the TES and mapped the players, we examined their interconnections to describe the ecosystem. We considered different types of relations such as: co-participation in R&D activities, co-location, similarity in technology, etc. We used multilayer network analysis to understand the relative importance of the players in the ecosystem and their positional advantages/disadvantages. We also modelled the information flows throughout the network, and studied the network's resilience over time. By clustering players by type of technology, geographical location, and linkages to other players we are also able to identify country or regional profiles of technological specialisation and study their evolution over time.

The JRC has developed a specific methodology to assess emerging techno-economic segments (TES) like AI, providing a synthetic overview of their ecosystem and dynamics. The main aim is to understand the entire ecosystem as it emerges: who are the key players, in which technologies they specialise, how they are distributed, how they are connected to each other,

and how innovation spreads through the network. The methodology (De Prato et al., 2018) makes use of different types of factual data such as patents, business registers, scientific and trade publications, and industrial associations records.¹⁸ To make sense of this heterogeneous data it applies both statistical and ML techniques to detect early

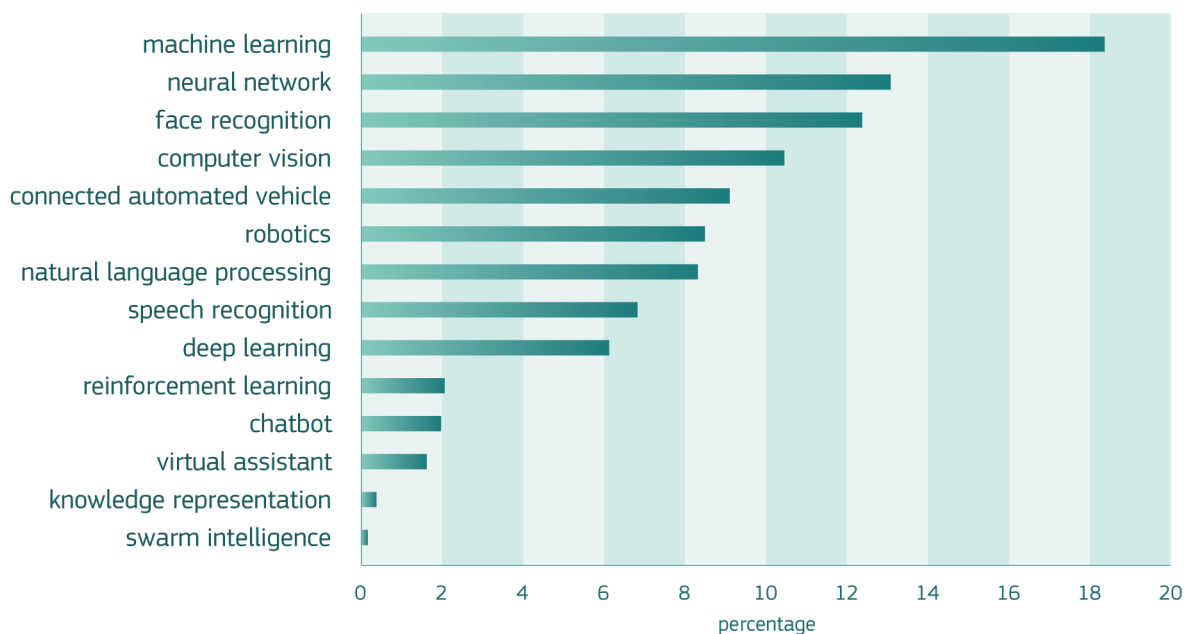


Figure 2. Percentage of AI players detected by the most representative keywords (2000-18)

Note: The conceptual proximity of some of the keywords, and the overlap between techniques and applications, may result in the same player being detected more than once.

developments, map them, monitor their evolution, and benchmark the ecosystem with the necessary scientific evidence.

Given the volume and variety of sources analysed, and the amount of painstaking work needed to match different data structures and identify connections through the data sources, we believe this is the most accurate and comprehensive database of the AI ecosystem to date. In the following section, we provide a global overview, and highlight some characteristics of the EU with respect to this global landscape.

3.3 Overview

Considering the AI landscape worldwide, the TES analysis has detected around 35 000 players in the period 2009-2018, of which 16 000 were involved in at least one research or innovation activity, and 19 000 are players engaged solely in industrial activities.

The TES analysis shows (Fig. 3) that **the EU is among the geographical areas with the highest number of players active in AI, just behind the United States and just ahead of China**. The USA had about 28 % of the players, with the EU (25 %), and China (23 %) following closely behind. All together these three regions account therefore for more than three quarters of all AI players worldwide. Within the EU (Fig. 5) the largest economies have the most AI players, with the UK representing 25 % of the EU AI players and

just over 6 % worldwide, Germany about 15 % of the EU's AI players and 3.7 % worldwide, and France about 11 % of the EU's and 2.7 % worldwide.

Consideration of the number of players relative to their country's GDP provides additional valuable information about countries with a smaller economy but a flourishing AI industry. Figure 4 shows that Israel and Singapore are making a huge effort to keep up in the race for AI; while they already have a non-negligible 2.1 % and 1.3 % of total world AI players, the indicator relative to GDP scales up to 2.45 and 1.63 players per billion euro respectively and positions them as first and second in the ranking. The EU presents 0.59 AI players per billion euro, slightly higher than the USA (0.57) and China (0.43).

In the EU we can see a similar picture (Fig. 5). The United Kingdom, Germany, France, Spain, Italy, and the Netherlands have the largest number of players but, when weighted by GDP, Bulgaria, Estonia, Cyprus and Malta stand out, although when numbers are small, ratios have been treated with some caution.

Figure 6 illustrates clearly the different mix of R&D and industrial players in the different regions and countries. On the right side of the graph, the bars indicate the number of players engaged in at least one R&D activity in the period 2009-2018 (firms, research institutions and academia). On the left, we have the non-R&D, i.e. industrial players only.

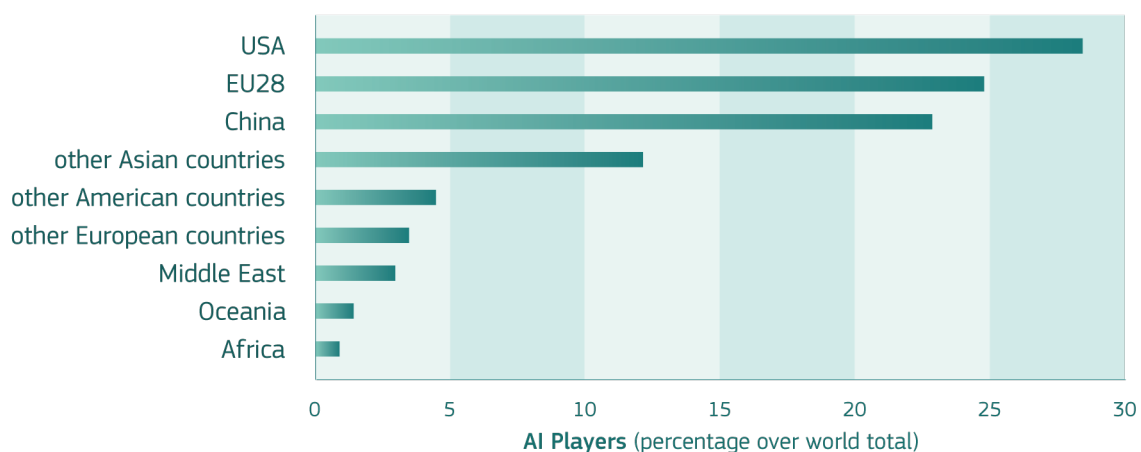


Figure 3. AI players in the world by geographical zones (% over world total), 2009-2018

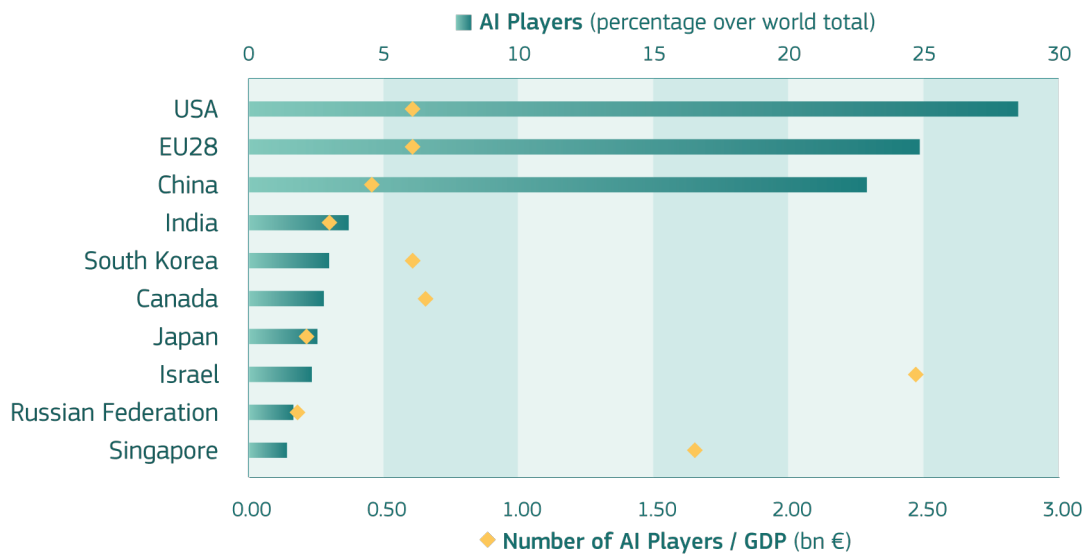


Figure 4. Top 10 AI players in the world and relation with GDP by geographical zones, 2009-2018

As shown, while Europe is well balanced in the number of research and non-research players (approximately 25 % each of the global number), the USA has approximately three times as many industrial/corporate players as research ones (41 % of all industrial players in USA vs. 13 % R&D), and China has about six times as many research players as industrial ones (7% industrial players and 42% R&D of the world total in each category). While we cannot forget that the TES analysis relies, so far, only on R&D materials published in English and thus is possibly penalising China, and other coun-

tries publishing strongly in their own language, Figure 6 shows clearly the Chinese effort to enter the R&D worldwide network and possibly come to dominate in the future. This overwhelming ratio of R&D versus non-R&D players is also observed in South Korea and, to a lesser extent, in Russia. The high proportion of non-R&D players in the USA, India, Israel and Canada reflects their vibrant industrial ecosystems, while the balance in the EU reflects a very strong research and academic environment with potential to nurture a flourishing AI industry. Figure 7 analyses the players that deserve special

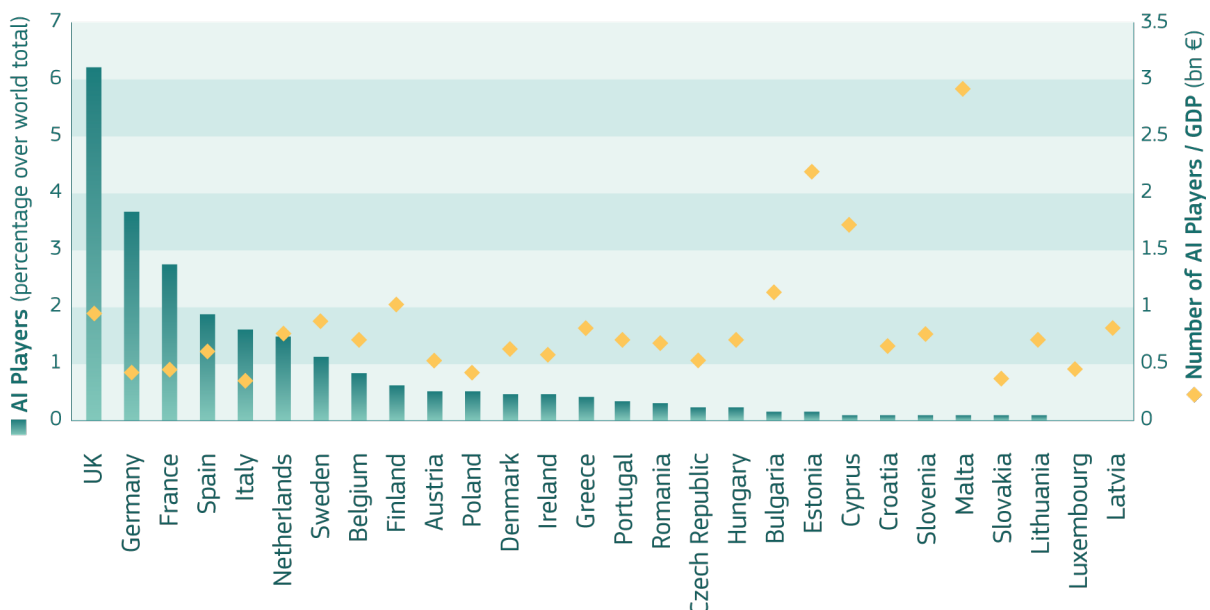


Figure 5. AI players in the EU and relation with GDP by country, 2009-2018

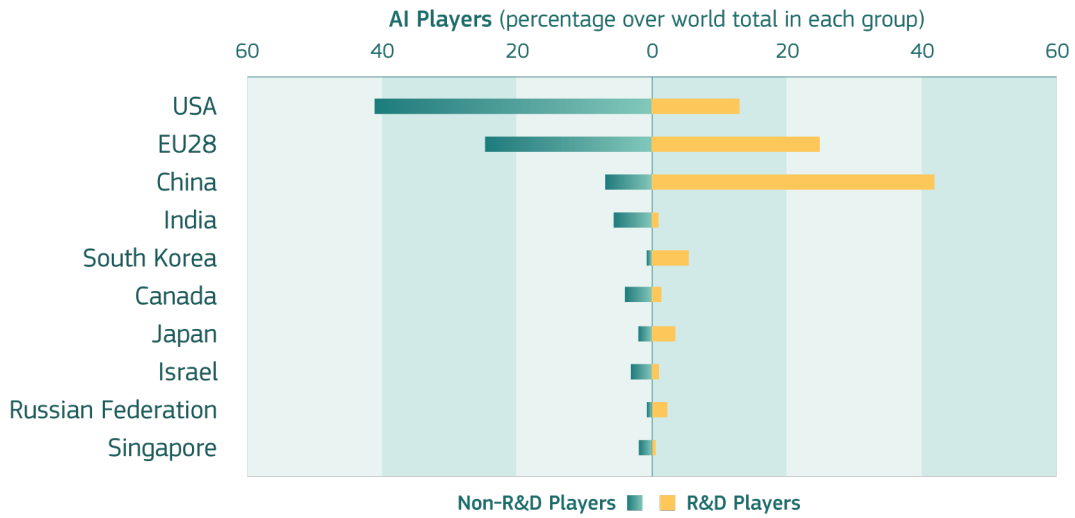


Figure 6. R&D versus non-R&D players in top 10 countries by number of AI players, 2009-2018

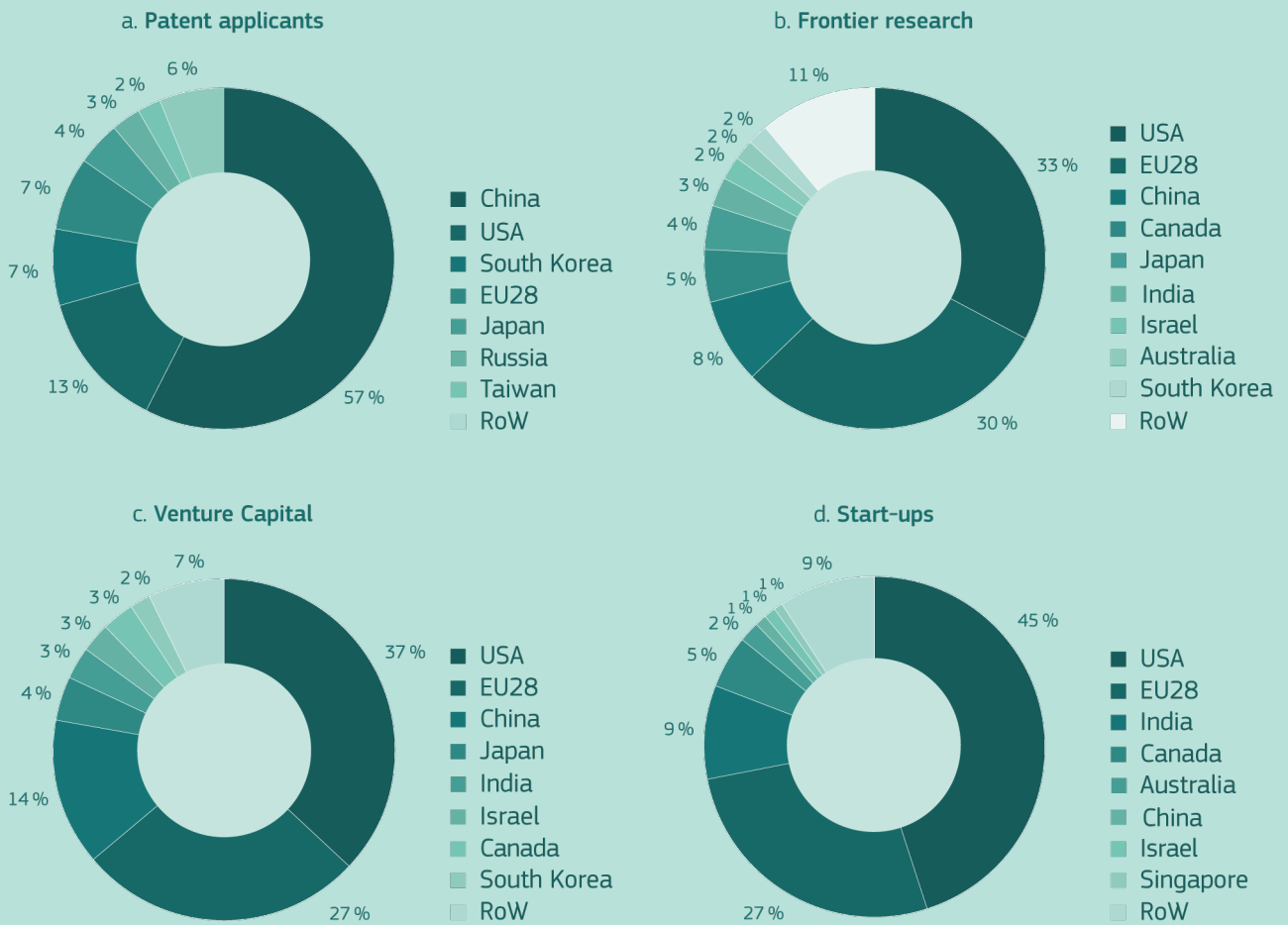


Figure 7. Number of AI players in selected types of activity (% over world total), 2009-2018

attention, like innovative players and firms considered to be start-ups or receiving venture capital. The highly developed industrial ecosystem in the USA puts this country at the forefront of AI-related activities with 37 % of all firms receiving venture capital, and 45 % of start-ups active in the field. **Both venture capital and start-ups are concentrated in a few countries, with the EU featuring strongly with 27 % of start-ups and 27 % of venture capital.**

It is worth noting the number of start-ups in India (9 %) and Canada (5 %), while the relatively low number in China reflects a different funding mechanism for new companies in that country.

China has a relatively strong position in venture capital (14 %), behind the USA and the EU. Moreover, in the past few years, China has conquered the top position in patenting, with 57 % of patenting applicants in AI worldwide coming from that country.

Behind the USA, **the EU has a prominent position in the research scene (frontier research), with 30 % of papers submitted to the top AI international conferences** are coming from EU companies

or research institutions. The importance of the EU in this field stems from lively AI scientific research activity supported by the EU framework programmes for R&D, which have fostered an intense inter-country collaboration and high participation from most EU countries. Figure 8 shows the distribution of players by country which participated in AI-related research projects under the FP7 and H2020 European research programmes in the period 2009-2018 with a good spread among the larger and more research-active countries, Germany, UK, Italy, Spain, France and the Netherlands.

3.4 Technological capacity

Research, both basic and applied, is the backbone of the technological capacity needed to bring the European industry to a leading position in the international landscape. In this section, we look at a few technological sub-domains of the AI landscape to analyse the EU's key strong areas with respect to other world regions. The largest number of AI players address AI domains related to ML (e.g. neural networks), and applications such as face recognition, speech recognition, computer vision or those encompassed by connected vehicles.

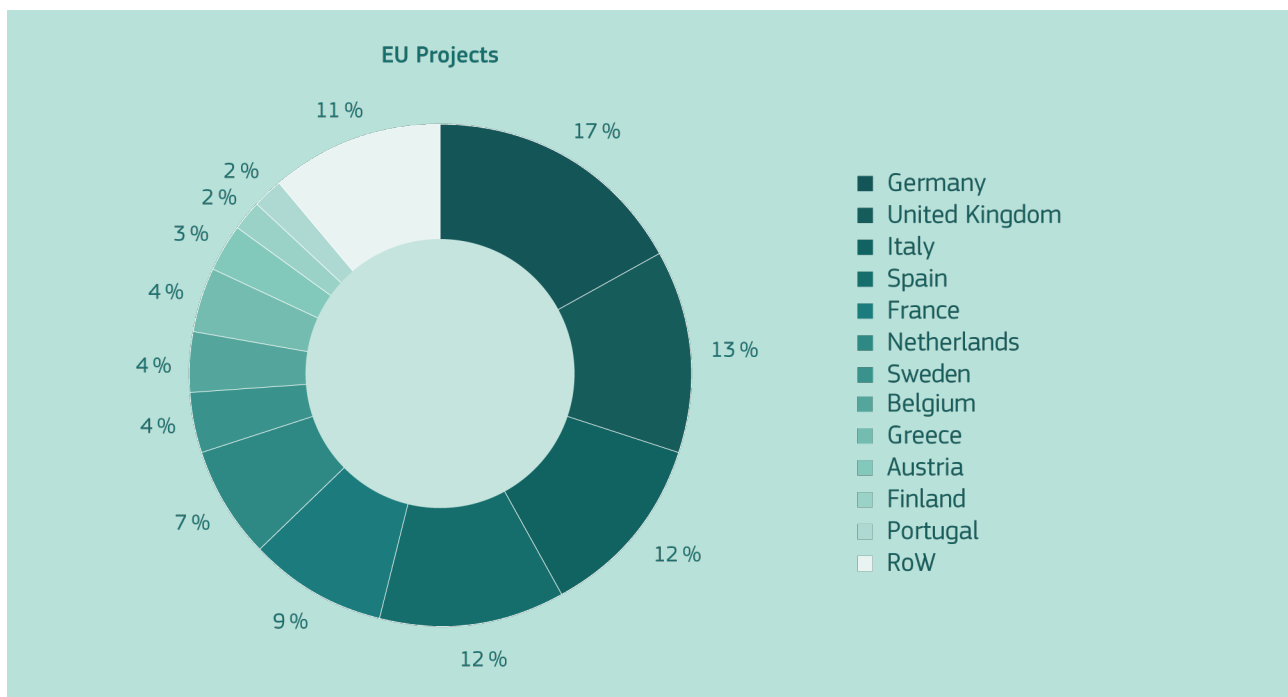


Figure 8. Distribution of EU players (%) in EU-funded AI research projects, 2009-2018

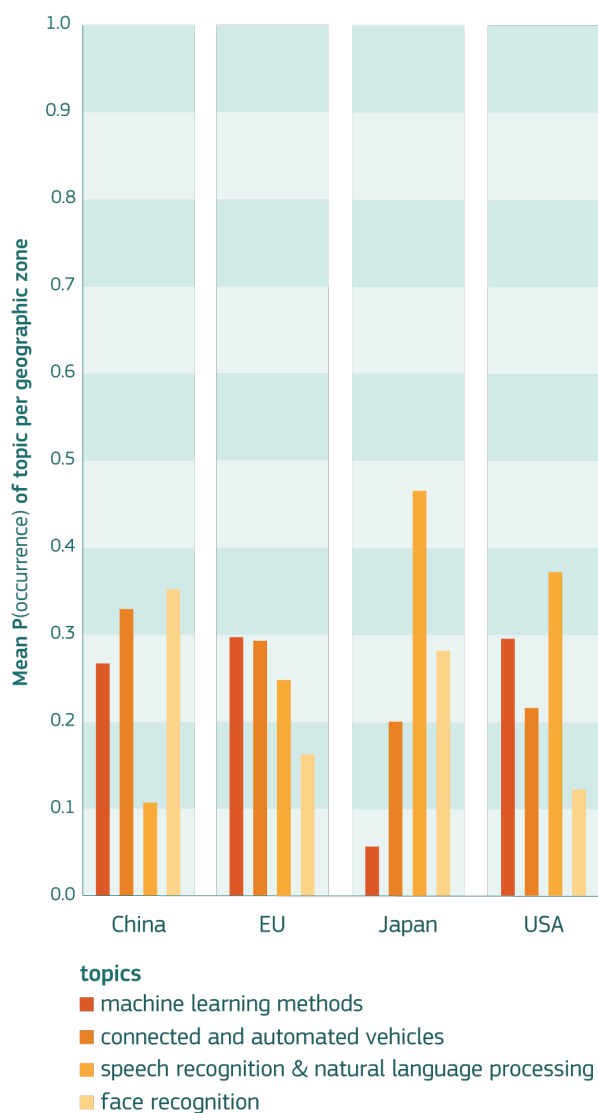


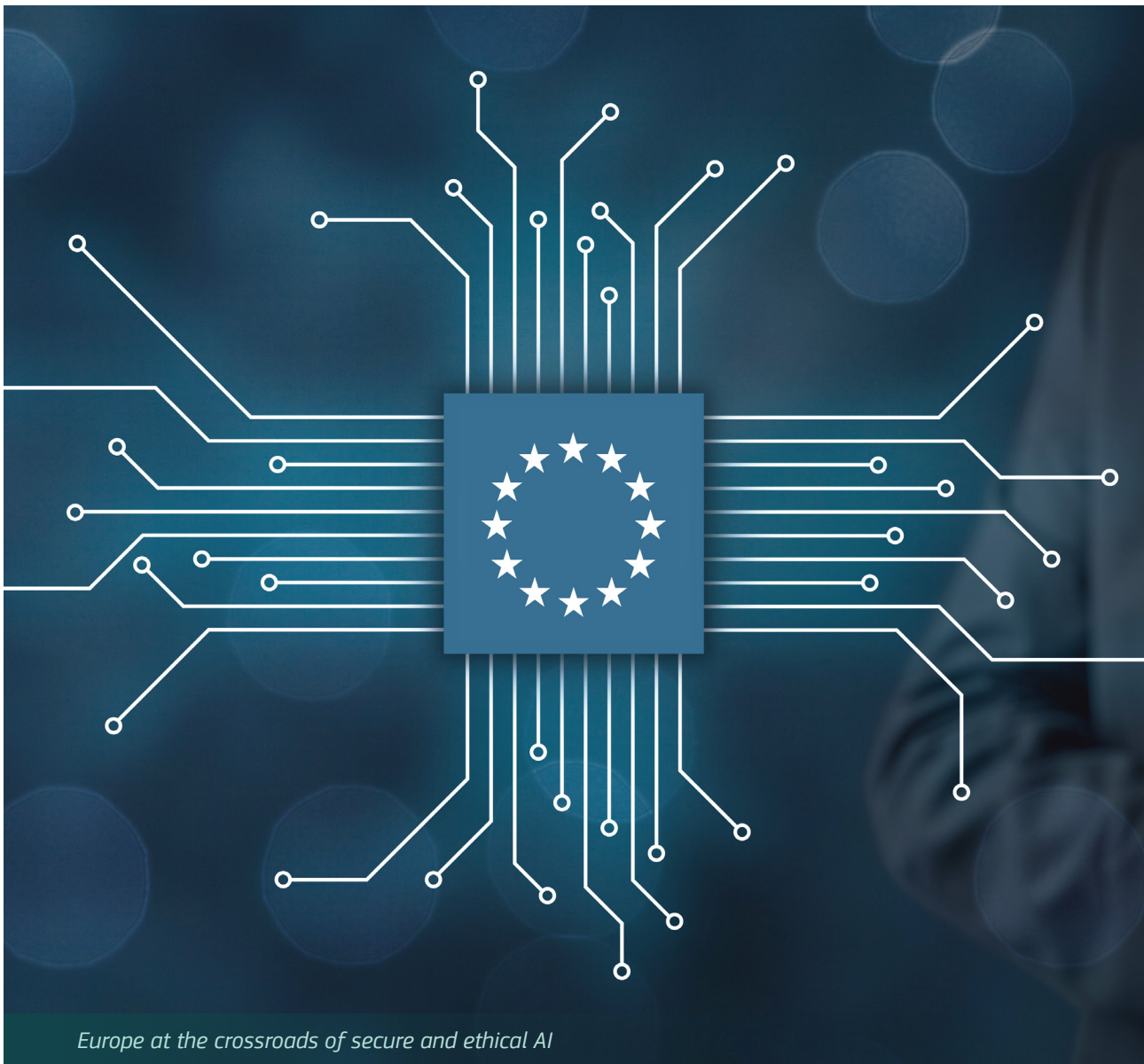
Figure 9. Occurrence of AI topics (%) by geographical zone, 2000-2018

From the TES analysis, we identified four key topics in the AI technological landscape for 2000-2018: ML methods; connected and automated vehicles; speech recognition and natural language processing; and face recognition. Figure 9 shows the key areas of attention/specialisation of each geographical area. As indicated, the USA and Japan have focused on the area of speech recognition and natural language; ML has captured great interest in the EU, the USA, and to a lesser extent in China, while face recognition is in the spotlight in China and Japan. The EU is the region presenting a more balanced coverage of all four topics.

3.5 Summary and conclusions

The overview of the global and European AI landscape we report in this chapter is the fruit of a comprehensive analysis of over 35 000 key players worldwide. The analysis confirms the intense competition in AI taking place worldwide with three main leaders: the USA, Europe, and China. Each region has about one quarter of all key players in the AI field, including both research and industry, but a distinctive different mix of players: while Europe is well balanced in the number of research and non-research players, the USA has approximately three times as many industrial/corporate players as research ones, and China has about six times as many research players as industrial ones. The strength of the corporate world in the USA is also indicated by its dominance in the number of start-ups (almost half of the total worldwide) and venture capital (more than one third of the total). China, on the other hand, is making a strong effort to turn research into patents, and accounts for almost 60 % of the world total. China has also put in place a strongly coordinated approach to AI, including government policy, industrial applications and research with the objective of becoming the world leader in AI by 2030. This is an ambitious but achievable target (Chapter 5).

Europe is currently in a good position in the quality of its research, with more than 30 % of top AI publications, just behind the USA (33 %). This result is the fruit of the good level of research funding and the intense inter-country collaboration provided by the framework programmes. Key areas of strength in Europe on which to build upon are automated and connected vehicles, and robotics. Continued policy support seems key to maintain and improve the European position in an ecosystem rather polarised around a few big powers.



Europe at the crossroads of secure and ethical AI

SUMMARY

This chapter shows that **many European countries as well as the EC are developing strategies and programmes to guide the development of AI**, with shared concerns over the need for an agreed ethical framework and applications that clearly benefit European society and uphold the European values enshrined in the Treaties. The High-Level Expert Group established in 2018 by the European Commission is developing ethical and investment guidelines that will provide a framework for subsequent developments.

Most national strategies analysed share similar attention to strengthening their research base, including the setting up of one or more national centres for AI, **support for their industry and SMEs** and awareness of the need to share data better between all the stakeholders: the public sector, industry, and the public. They also focus on applications aimed at **modernising public administrations**, as well as specific sectors such as health.

AI IN THE EU

■ 4.1 Strategies and plans

■ 4.1.1 The European Union

Many activities related to AI are currently taking place in the EU. They include national strategies and programmes, and coordination activities among the Member States and the European Commission.

Many of the enablers for the development of AI have been addressed during the last five years under the Digital Single Market strategy. This is aimed at ensuring access to online activities for individuals and businesses under conditions of fair competition, consumer and data protection, removing geo-blocking and copyright issues.

In April 2018, the EU Member States signed a *Declaration of cooperation on Artificial Intelligence*¹⁹ in which they agreed to work together on the most important issues raised by AI, from ensuring Europe's competitiveness in the research and deployment of AI, to dealing with social, economic, ethical and legal questions.

During the same month, the EC issued a Communication on AI for Europe (EC, 2018a) with three main objectives:

- Boosting the EU's technological and industrial capacity and AI uptake across the economy, both by the private and public sectors. This includes investments in research and innovation and better access to data.
- Preparing for socio-economic changes brought about by AI by encouraging the modernisation of education and training systems, nurturing talent, anticipating changes in the labour market, supporting labour

Many European countries as well as the EC are developing strategies and programmes to guide the development of AI so that it benefits all Europeans.

market transitions and adaptation of social protection systems.

- Ensuring an appropriate ethical and legal framework, based on the Union's values and in line with the Charter of Fundamental Rights of the EU. This includes forthcoming guidance on existing product liability rules, a detailed analysis of emerging challenges, and cooperation with stakeholders through a European AI Alliance to develop AI ethics guidelines.

The Communication earmarked EUR 1.5 billion to support AI research for 2018-2020, having already invested some EUR 2.6 billion in AI-related research in the Horizon 2020 Framework Programme (EC, 2018b). However, the aim is to join forces between European, national, and private-sector investments to reach some EUR 20 billion per year for the next decade. As part of this partnership effort, the EC established the European AI Alliance as a multi-stakeholder forum for engaging in a broad

and open discussion of all aspects of AI development and its impact on the economy and society. It is steered by the High-Level Expert Group on AI (AI HLEG), which consists of 52 experts who have been selected by the Commission for this task. The AI HLEG is now focusing on two main reports with input from the members of the European AI Alliance:

- (1) Draft AI ethics guidelines, which will offer guidance on how to implement ethical principles when developing and deploying AI, building on the work of the European Group on Ethics in Science and New Technologies and the European Union Agency for Fundamental Rights; and
- (2) Mid- and long-term policy recommendations on AI-related challenges and opportunities, which will feed into the policy development process, the legislative evaluation process and the development of a next-generation digital strategy.

The EC is currently working with Member States to have a coordinated plan on AI by the end of 2018. The main aim is to maximise the impact of investment at the EU and national levels, encourage cooperation across the EU, exchange best practices, and support the EU's global competitiveness in this sector.

The plan should last until 2027, i.e. the remaining time under the current financing framework and the entire period of the next Multiannual Financial Framework 2021-2027 to support better European and national strategies.

The areas currently listed as priorities for collaboration among the Member States include:

- Mapping research excellence and joining efforts to increase partnerships in Europe.
- Interoperable dataset repositories (or 'data spaces') making quality data available for a broad range of users, including trusted

data and validation/certification systems for algorithms.

- Access to know-how support and shared testing facilities such as smart hospitals or precision-farming solutions.
- Networks of Digital Innovation Hubs to facilitate the uptake of AI.
- Support to start-ups.
- Skilling and upskilling of the workforce and action to attract and retain talent in Europe.
- Improved public services through numerous AI-enabled applications, such as citizen-government interaction in the context of smart cities and tax-fraud detection.

The plan for AI will also link to the ecosystem of supporting technologies: AI chips, 5G, High-Performance Computing (HPC) that are already in the pipeline or are planned. In this respect, it is worth noting that in June 2018 the EC proposed a Digital Europe programme²⁰ with a budget of EUR 9.2 billion for the period 2021-2027. There are five areas proposed as the focus: high-performance computing, artificial intelligence, cybersecurity, advanced digital skills, and ensuring their wide use across the economy and society.

Within the Digital Europe Programme, the EC proposes to develop common 'European libraries' of algorithms that would be accessible to all to help the public and private sectors identify and acquire whichever solution would work best for their needs. Open platforms and access to industrial data spaces for artificial intelligence will be made available across the EU in Digital Innovation Hubs, providing testing facilities and knowledge to small businesses and local communities.

In addition to the above regulatory package, other aspects are currently being addressed such as: Liability (including the revision of the Product



“The Coordinated Plan for the Development of AI in Europe aims to maximise the impact of investment at EU and national levels, and promote cooperation.”

Machinery Directive), and an Action Plan on Disinformation.

European action is not limited to the development of an appropriate legislation (and standards) but also includes a set of relevant initiatives to speed up technology development and its uptake. Some relevant examples for AI include, for example 5G, Galileo, very high-speed and ubiquitous connectivity, European Processor Initiative, Joint Undertaking for High Performance Computing, and the Big Data Public-Private Partnership.

Moreover, sectoral policies are being discussed to regulate the use of the ICT technologies (including AI) in specific segments, for example:

- The Third Mobility Package aims to allow citizens to benefit from safer traffic, less-polluting vehicles and more advanced technological solutions, while supporting the competitiveness of the EU industry. Particular focus is given to autonomous mobility that has the potential to make transport safer, more accessible, inclusive and sustainable. The Commission’s proposal encompasses a strategy aiming to make Europe a world leader for fully automated and connected mobility systems.

- The Clean Energy for all Europeans package was presented in November 2016 and aims to provide the stable legislative framework needed to facilitate the clean energy transition.
- The Commission Directive 2018/844 aims to improve energy efficiency through digital solutions such as building automation and electronic monitoring of technical building systems.

By nature, AI technology will penetrate all industrial sectors, digital services and social life. It is therefore extremely difficult to reconcile the need for a clear common framework and the rapid adoption in some domains when the technology is already mature to enter the market. To monitor the development and uptake of AI in different sectors and its impacts on work and society, in December 2018, the EC is launching an AI-Watch facility with the following objectives:

- 1) Develop an overview and analysis of the European AI ecosystem,
- 2) Monitor the uptake of AI applications across the economy,
- 3) Monitor the progress of AI technology,
- 4) Assess the evolution of service robotics over the last 10 years,
- 5) Gather information on all EU Member States' national initiatives on AI,
- 6) Provide an overview on the use of AI in public services,
- 7) Develop an AI index including the dimensions relevant for policymaking,
- 8) Make information available to the public using an online tool.

In the following sections, we provide a summary of selected national strategies on AI.

4.1.2 France

France is one of the more active countries in developing a national AI strategy. The government has recently published three main reports which are reviewed below:

The French AI plan (FIA, 2017) offers a strategy for the coming years focusing on research and education, innovation, and the social and economic impacts of AI.

The report by the Office of the Parliament for Science and Technology (OPECST, 2017) 'Toward a Controlled, Useful and Demystified Artificial Intelligence' focuses on social and regulatory aspects and raises important issues to be addressed in developing a socially useful AI, whilst being realistic about expectations on potential impacts.

The Mission Villani report (2018) addresses six main areas:

1. An innovative and ambitious industrial and economic policy: the mission opted to concentrate on some key sectors: health, transportation, environment, and defence-security. The policy will require public support but along with some structuring and precise challenges.
2. To devise a data policy fitted for the stakes. Access to a sufficient amount of data is a major issue in France and Europe. The mission is trying to favour wide access, the creation of data ecosystems while providing new protection to users.
3. To anticipate and monitor the impact of AI on work and employment.
4. AI as a tool for a sustainable and ecological economy, providing a vision for a 'greener'

AI enabling an ecological transition. For instance, the EU could take a leading role in the development of specific chips in the semi-conductor industry, preparing the EU industry for the post-silicon era.²¹

5. Setting a framework for ethics and trust to enable the growth of AI.
6. Reducing the brain drain and finding ways to offer attractive working conditions to both French and international researchers.

The final report pleads for an inclusive and diverse AI stressing, for instance, a role to improve gender balance. To develop an inclusive policy for AI, the report suggests incorporating a dual objective: *'First, to ensure that the development of AI technology does not cause an increase in social and economic inequality. Second to call on AI in order to reduce this'* (Mission Villani, 2018: 133).

The mission advocates building a data-focused economic policy grounded in a European data ecosystem. The report holds that the state has to be a key driver in these various areas of transformation and suggests setting up an inter-ministerial coordinator to implement this strategy, with support from a shared specialist centre. One of the novel aspects is to set up four to six Interdisciplinary Institutes for Artificial Intelligence (3IA institutes) nationwide, organised into a network: the National Network of Interdisciplinary Institutes for Artificial Intelligence.

These new entities are meant to improve the relationships between public basic research and private firms. It is also a means to avoid the brain drain of researchers. This network should be integrated in a future European AI research area network that could be modelled on the EMBL (European Molecular Biology Laboratory), which has been operating successfully since 1974. Other European partnerships could be built with the instruments of the H2020 programme, like the current public-private partnerships in robotics and Big Data. Enhancing the appeal of careers in public research is another concern,

with the suggestion to at least double the starting salary for researches in the public sector in view of the high degree of competition with industry.

4.1.3 United Kingdom

In 2016, the Government Office for Science (UK Gov., 2016a) released a note on 'Robotics, automation and artificial intelligence', which explored the opportunities and recommended three actions concerning challenge areas, facilities and skills. The same entity released a report that same year 'Artificial intelligence: opportunities and implications for the future of decision making' (2016b) that included a presentation of what AI is, a review of the use of AI by governments, and a discussion on the effects on labour markets. In 2015, the Alan Turing Institute was created as the national institute for data science, headquartered at the British Library (Hall & Pesenti, 2017).

In 2016, the House of Commons Science and Technology Committee released a report on 'Robotics and artificial intelligence' (House of Commons, 2016). The report deals mostly with the broad issues (economic, social, ethical, and regulatory) triggered by this disrupting technology, but the last chapter examines *'the research, funding and innovation landscape for robotics and AI'*. Its conclusions were rather critical: *'Government leadership has been noticeably lacking. There is no Government strategy for developing the skills, and securing the critical investment, that is needed to create future growth in robotics and AI'* (UK House of Commons, 2016: 33). In its answer to the report, the government acknowledged that more could be done to take a leading role, and in March 2017 it announced an industry-led review on the conditions necessary for the AI industry to continue to thrive and grow in the UK. In April 2018, the government published its AI Sector Deal (UK Gov. 2018) as part of a large industrial strategy to place the UK at the forefront of AI development and use. The Deal invests £ 950 million, of which two thirds is new money and the rest is budget already allocated.



“ Most national AI strategies focus on strengthening research, supporting industry, sharing data, and modernising public administration. ”

The Deal has five main pillars:

Ideas: Support research in data science AI with £300 million, including £49 million for the Alan Turing Institute, stimulate the uptake of AI, including within the public sector, and invest up to £100 million to support AI innovation to raise productivity. The Institute’s public policy programme, in particular, is where the UK government comes for independent advice on data science, AI, and ethics.

People: Work with schools, universities and industry to ensure a highly skilled workforce, including £ 400 million to support STEM teaching in schools, and up to 200 additional PhD bursaries in AI-related fields with the target of reaching 1 000 government-supported PhD places per year in 2025. Also enable access to highly skilled global talent and promote diversity.

Infrastructure: Enhance the UK’s data infrastructure making more public-sector data available, including geospatial data under the guidance of a newly created Geospatial Commission, and promote Data

Trusts as a way to share public- and private-sector data in a secure way. Promote the upgrading and maintenance of the physical digital infrastructure, including £1 billion to develop 5G and extend broadband.

Business environment: The government has set up an AI council with industry and academia to oversee the strategy and a new Office for AI to support its delivery. Support for business and innovation will include a £ 2.5 billion investment fund, reaching £ 7.5 billion in partnership with the private sector.

Places: Support the continued development of London as Europe’s capital for AI but also expand support to regional clusters with the appropriate infrastructure and skills. Further support for the Alan Turing Institute to become the national academic institute for AI and data science.

4.1.4 Finland

The government appointed a steering committee in May 2017 due to report in April 2019. A first report was published in October 2017 (Finnish Gov. 2017) setting out a vision of a country embracing AI in every aspect of life in an open and ethical way. It identified key areas for action including:

- Enhancing competitiveness by developing industrial ecosystems to help in the application of AI, and measures to facilitate the access and use of AI and data for companies.
- Increasing the quantity and quality of data shared in Finland, through legal frameworks that are ‘based on the importance of the data to business operations (not on data protection first)’, regulatory sandboxes for experimentation, increased access to the public of the government’s Mydata network, and support to companies to turn their data into products.
- Launching an AI accelerator pilot in 2018 based on open data, open code, open interfaces with companies providing anonymised data for research, partnership with universities to research on the data, and access to small companies to experiment with the use of AI-based solutions. In addition, a number of open piloting and testing environments will be set up to facilitate experimentation.
- Establishing a new Centre of Excellence for AI and basic research, including a virtual university and the creation of massive open online courses (MMOC) to support the upskilling of the population and workforce.
- Supporting research and innovation, and extending the application of AI in public service, for example with the development of personalised AI assistants, and new partnership models between the public and private sector to facilitate the adoption and use of AI.

4.1.5 Other EU countries

Several other Member States already have AI strategies in place or are in the process of adopting them. From those illustrated above, most strategies have similar aims and objectives: embracing AI and become a leading country in the field, supporting innovation and business, promoting the use of AI in the public sector, boosting R&D and skills, establishing one or more national centres of excellence, and developing strategies to promote the sharing and use of data more widely between the public and private sectors. A European strategy needs to build on these national efforts and investments to make the whole greater than the sum of its parts.

4.2 Summary and conclusions

This chapter shows that **many European countries as well as the EC are developing strategies and programmes to guide the development of AI**, with shared concerns over the need for an agreed ethical framework and applications that clearly benefit European society and uphold the European values enshrined in the Treaties. The High-Level Expert Group established in 2018 by the European Commission is elaborating ethical and investment guidelines that will provide a framework for subsequent developments.

Most national strategies analysed give similar attention to strengthening their research base, including the setting up of one or more national centres for AI, **support for their industry**, and SMEs, and awareness of the **need to share data better** between all the stakeholders: the public sector, industry, and the public. They also focus on applications aimed at **modernising public administrations**, as well as specific sectors such as health.





China is implementing its ambitious strategy on AI.

SUMMARY

This chapter highlights some of the main features of the AI developments in China, which has identified AI and related applications as strategic to its future economic and social development, and to its role as a global superpower. As a result, it has put in place a strongly coordinated approach to AI, including government policy, industrial applications and research with the objective of becoming the world leader in AI by 2030. This is an ambitious but achievable target as the JRC analysis of the AI innovation ecosystem in China confirms. This opinion is shared by main market analysts. In fact, AI deployments with a high impact on daily activities have already started, including authentication processes, medical diagnosis, premium insurance, transportation and retail facilities, and security. Potential obstacles to the full development of the strategy relate to the difficulty of having AI adopted in traditional industries, and prepare enough specialists in the country able to understand the domain of application, its needs and expectations, identify the best algorithm and approach, and then customise it and develop it further.

THE AI ECOSYSTEM IN CHINA

■ 5.1 China's economic and policy context

The development of AI is a global phenomenon, but the case of China is particularly interesting and not as well known as that of the USA. For this reason, we dedicate a slightly longer section to this country, drawing on a comprehensive study carried out for the JRC by Feijoo (2018).

China is already a leading global force in the digital economy. For example, it has 42 % of the global share of e-commerce, it processes 11 times more mobile payments than the USA, it is home to one-third of the world's unicorns²², and leads the implementation of 5G mobile communications (Woetzel et al., 2017). This digital success is fuelled by three main factors: (i) a large and young Chinese market enabling rapid commercialisation of digital business models; (ii) a rich digital ecosystem expanding quickly beyond a few large companies; and (iii) strong governmental support for companies through favourable economic and regulatory conditions, and a multiple role as strategic investor, consumer of digital technologies, and provider of access to key data.

At the macroeconomic level, China can no longer depend solely on increases in capital and labour to achieve the desired levels of sustainable economic growth, nor on producing cheap goods to sell elsewhere as has been the predominant model for the last three decades. Against this background, there has been a recent shift into innovation as a strategic

China has put in place a strongly coordinated approach to AI, including government policy, industrial applications and research with the objective of becoming the world leader in AI by 2030.

driver to increase the quality of products and productivity, and expand internal consumption, thus decreasing the dependency on external markets. Digital technologies in general, and AI in particular, are expected to lead to higher efficiency in products and services, achieving large productivity gains and a more dynamic economy, with businesses able to compete globally and even export 'Made in China' digital business models and solutions (Woetzel et al., 2017).

But there are other interests beyond mere economics. China, as a global power, is using initiatives such as ‘One belt, one road’ or institutions such as the Asian Infrastructures Investment Bank to shape a geopolitical environment more favourable to its interests (Delage, 2017). Digital technologies, and AI in particular, are essential constituents of this strategy, showcasing a state-of-the-art new industry fully developed in China and a different approach from the West to economy and policy through technology. The strict application of national laws to cyberspace and the use of AI-related technologies to maintain social harmony and control are the main examples of this differentiated approach.

The potential of AI is clear to China’s policymakers. As stated by President Xi, China ‘... should unswervingly follow an independent innovation path featuring Chinese characteristics, stick to the guiding principles of independent innovation, leap-frogging development in key sectors, with development supported by science and technology and oriented towards the future’ (Xi, 2014, p. 134). In fact, since 2014, the government has launched a series of key national economic initiatives that are relevant to AI with the goal of creating a EUR 13 billion AI market in China by 2018, and helping China to lead the world in AI by 2030.

In the case of AI, and in contrast to what happened in the recent past, development of the sector relies on a plan that encompasses all relevant players within an ecosystem, from universities and research centres to existing companies and new firms in the entrepreneurial / innovation milieu, which is also under development in China. It will be, therefore, a first real assessment of China’s chances to create an independent home-grown, fully-fledged innovation and industrial ecosystem.

However, AI is not an isolated ecosystem as it interacts and benefits from not only the development of successful innovations such as ecommerce or mobile payments but also from cloud computing, industry 4.0, robotics, IoT, blockchain, and IT security ecosystems. The particular case of cloud

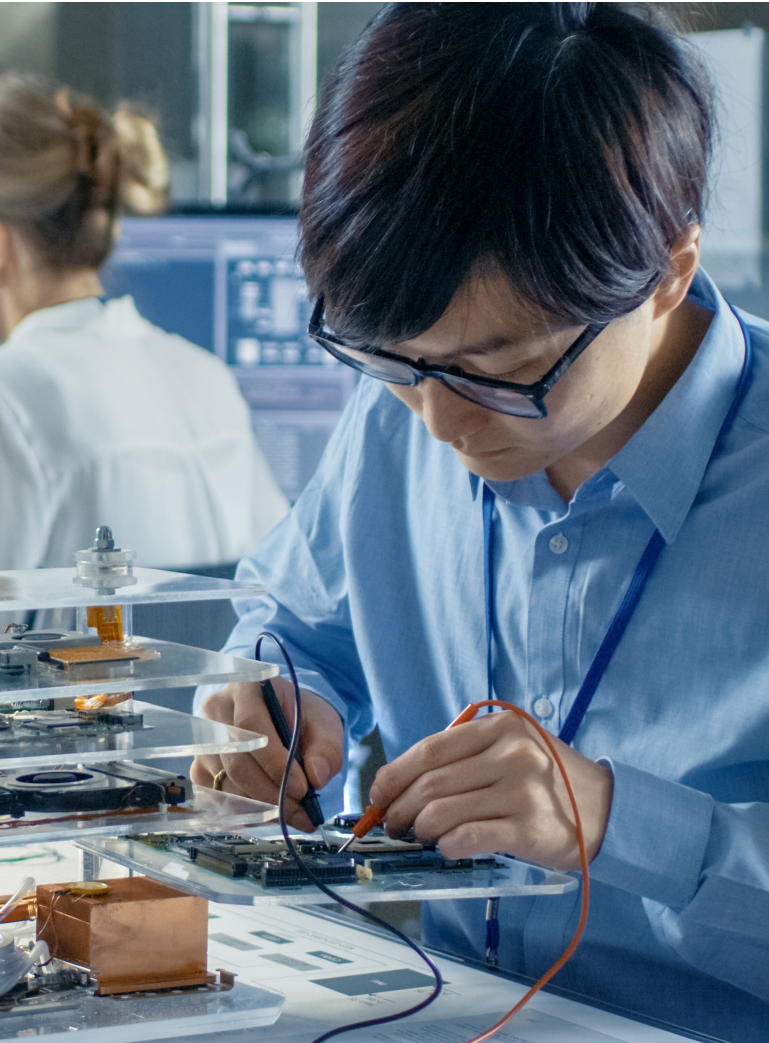
computing is relevant since it is expected that a considerable part of AI services will be provided from the cloud. Therefore, market developments in both AI and cloud computing will be considerably aligned.

“ Since 2014, the government has launched a series of key national economic initiatives that are relevant to AI with the goal of creating a EUR 13 billion AI market in China by 2018, and make China lead the world in AI by 2030.”

■ 5.2 Government policies and initiatives

In this section, we summarise the main government initiatives related to AI following a timeline up to 2018. The timeline is interesting because it displays how **policies and initiatives started at industry level slowly go up the value chain to include applications and consumer products**. It is particularly relevant to follow the developments that allowed companies in different sectors to share and access data to train newer and better algorithms. Crucially then, the government issued

specific AI guidelines covering the whole value network in a holistic perspective.



Made in China 2025, State Council, May 20, 2015

The goal of this strategy is to make China an advanced and prestigious manufacturing power by 2025. In fact, intelligent manufacturing is positioned as the main development direction of 'Made in China'. The main industries targeted are machinery, aviation, shipping, automobiles, light industry, textiles, food and electronics, and the general aim is to improve the precision manufacturing and agile manufacturing capabilities.

The strategy includes the promotion of an open industrial ecosystem based on the dynamic perception of consumer demand and new industrial internet applications, such as intelligent monitoring, remote diagnosis and management, and whole industry chain tracing.

Internet +, Action Guidance, State Council, July 4, 2015

The internet+ strategy focuses on intelligent manufacturing and innovation, promoting the development of cloud computing, IoT, intelligent industrial robots, industrial control systems, IntelliSense components, industrial operating systems and software, and additive manufacturing technologies. It also includes an upgrade of production equipment, processes and a focus on (big) data sharing. The strategy objectives specific to AI are:

- To increase public support for the development of AI and core technological breakthroughs;
- To promote the popularisation and application of AI in fields such as smart home, smart terminal, smart car, and robots;
- To create a new computing cluster that supports ultra-large-scale DL;
- To build a massive training resource base including voice, images, videos, maps and other data;
- To strengthen the construction of innovation platforms of basic AI resources and public services;
- To research, develop and industrialise key AI technologies such as computer vision, intelligent speech processing, biometrics, natural language understanding, intelligent decision control and new human-computer interaction;
- To support cooperation among security enterprises and internet companies, in order to develop and promote big data analysis

technologies such as image recognition and enhance the intelligent service level of security products;

- To promote the application of internet technologies, intelligent sensing, pattern recognition, data analytics and robotics;
- To nurture a number of key enterprises and innovation teams to lead the global development of AI.

Robot Industry Development Plan (2016 - 2020), Ministry of Industry and Information Technology, National Development and Reform Commission, Ministry of Finance, April 27, 2016

The main goal of this plan is the development of intelligent industrial and service robots. Six areas are defined for industry: welding robots, vacuum robots, fully autonomous programming intelligent robots, human-computer collaboration robots, dual-arm robots and heavy-duty autonomous vehicles, plus four areas for the service domain: fire rescue robots, surgical robots, intelligent public service robots, and intelligent nursing robots. The plan also proposes that the government establishes robotic innovation centres and encourages banks and private investors to support the development of robotics and robots.

New Generation AI Development Plan, State Council, July 8, 2017

This is the key strategy for AI where guidelines, strategic goals, key tasks and supporting measures for a new-generation AI development by 2030 are set up. The plan defines very clearly the strategic goals in AI development:

- By 2020, most AI technologies and applications will be on a par with the most advanced levels worldwide, and the AI industry will become a new important economic growth net contributor. AI will have an impact on our daily lives.

- By 2025, major breakthroughs will be made in the basic AI theory. Some of the technologies and applications will reach the world leading level. AI will become the major driving force for industrial upgrading and economic restructuring in China. It will have an impact on society.
- By 2030, AI theory, technologies and applications will be world leaders. China will become the world's major AI innovation centre.

The plan mentions six key tasks:

- To build an open, collaborative AI technology innovation system, cutting edge in terms of basic theory, key common technologies, innovation platforms, and high-end talent.
- To foster a high-end and efficient smart economy, develop the emerging AI industry, promote industrial upgrade and create AI innovation areas.
- To build a safe and supportive smart society, develop efficient and intelligent services, including for social governance, enhance public safety and security based on AI and promote sharing and mutual trust in social interactions.
- To strengthen military and civilian integration in the field of AI, promote the two-way conversion of AI technology between military and civilian sectors and jointly build up and share military and civilian innovation resources.
- To construct a ubiquitous, safe and efficient intelligent infrastructure system and strengthen the upgrading of infrastructures such as networks, big data and high-performance computing.
- To plan significant science and technology projects, and strengthen overall coordination.

■ 5.3 Regional / local initiatives

There are many initiatives taking place in China at the regional and local level and supported by the respective governments. This seems to create a particular type of competition between different locations that is a main feature of the AI innovation ecosystem. Below, we give a few illustrative cases.

Beijing: In 2017, the local government of Beijing issued a guideline to support technologies, applications and the AI industry. It sets the goal for the city to become an AI innovation hub with global influence by 2020 - anticipating national objectives. According to the Beijing government, the region has more than half of the key AI research institutes in China, nearly 400 AI enterprises, and during the first nine months of 2017 there were investments in 51 AI projects totalling US\$ 1.6 billion.

Zhongguancun: often referred to as China's Silicon Valley - is the main high-tech business district of Beijing. It includes some of the most prestigious tech universities and about 500 colleges, national scientific research institutions and R&D centres for multinational companies. There is a plan to locate the largest industrial cluster for integrated circuits in China there. The high-tech companies located in the district generated about US\$ 600 billion in revenues in 2017 and the area is home to 67 unicorns. It is estimated that 30 000 overseas returnee entrepreneurs have established over 8 000 companies in Zhongguancun. The district also published a plan at the end of 2017 for cultivating AI talent with the expectations of achieving an output for the AI industry there of more than EUR 7 billion in 2020.

Shanghai: Shanghai has its adaptation of the central government plan called '13th Five-Year Plan for Shanghai Manufacturing Transformation and Upgrading', issued by the Shanghai Municipal People's Government in 2016. It aims to capture the opportunities from emerging technologies such as AI, quantum communication, VR and precision medical treatment, as well as new trends in

industrial organisation, innovation and entrepreneurship. In 2017, Shanghai had about 500 innovative parks hosting 16 000 start-ups.

Guangzhou: Guangzhou has launched a full set of industrial parks supported by the local government: the Guangdong Intelligent Manufacturing Demonstration Base, the Intelligent Industrial Park in Huangpu Machinery Valley, the Zengcheng Intelligent Equipment Manufacturing Industrial Park, the Huadu New Energy Vehicles and Intelligent Manufacturing Equipment Industrial Park and the Liwan 3D Printing Industrial Park. All of them include direct support to innovative companies, particularly those linked to the transformation of the manufacturing industry.

Hangzhou: A radically distinct approach is taking place in Hangzhou. Alibaba's success has shaped Hangzhou's tech scene. The city is home to many incubators, funded in part by government subsidies, that are filled with entrepreneurs who previously worked at Alibaba.

■ 5.4 The specificities of the AI innovation ecosystem in China

The AI innovation ecosystem in China is most developed in the areas of image, face and voice recognition, together with the techniques that support such fields (basically machine and DL and the algorithms that accompany both). There is also recent and particular emphasis on the development of AI chips. As a consequence, the more developed areas of application in China are security, healthcare, transportation, urban traffic management, commerce and driverless cars. Adding all of them together configures a scenario of deep economic and social changes that have already started to take place in China. Applications of AI are now widely available and impact the daily lives of China's citizens, combining convenience, efficiency, personalisation and surveillance (See Box 6).

China's large companies work closely with the government, and take advantage of the positive

perception of consumers to experiment with applications unheard of outside China. This fact permeates the AI innovation ecosystem, but arguably, the foremost instance of alignment between consumers, companies and government is the implementation of the so-called ‘social credit system’, tentatively scheduled for 2020. This AI-based system, which would be controversial in other countries, will be able to collect user behavioural data in both physical space and cyberspace and provide a rating of citizen reputation. This rating could then be used to give or deny access to a range of services provided by the state.

Military involvement is another salient feature of the AI ecosystem in China: civilian AI breakthroughs are applied for military use (under the national strategy of ‘military-civil fusion’) and, vice versa, military-based developments are transferred to civilian industry. According to the Central Military Commission for Science and Technology, AI could accelerate military transformation by means of intelligent and autonomous unmanned systems, AI-enabled intelligence analysis, war-gaming simulation and training, defence, offence, and command in information warfare; and intelligent support to command decision-making (Kania, 2017). In the other direction, China wants to significantly increase the spillover of military developments into the civilian industries.

Another feature of the AI innovation ecosystem in China is the competition among cities and regions for talent, in particular from the new policies launched by second-tier cities. As a result, there are considerable differences in the accumulation of talent and opportunities in new tech industries between East and West China.

A final feature of the ecosystem refers to the limited role awarded to critical views of AI in China. Such views could be perceived as a lack of loyalty to a policy that is highly supportive of AI. However, this absence of criticism could lead to a lack of verifiable algorithms, trustworthy systems, or a situation in which moral responsibilities are outsourced to algorithms. The calls for policy support of algorithmic

BOX 6. Some examples of AI applications in China

Alibaba’s core business is selling goods and providing a platform for B2B trade. Just to give a sense of scale, during ‘Singles’ Day’ a creation of the company that takes place on 11 November, Alibaba processed at peak time more than **325 000 orders per second** through pop-up stores selling products fitted with VR mirrors for virtual try-outs. An AI fashion consultant helped match items with the one selected. In one day it sold **more than US\$ 25 billion worth of goods**. By contrast, on 2017 Cyber Monday (November 27), the biggest online shopping day in the US, all retailers combined brought in **US\$ 6.59 billion** (Knight, 2018).

Tencent’s main source of data is WeChat which, in 2018, reached more than **1 billion monthly active users**.²³ Its flagship project is its healthcare platform to book appointments and make payments which, in 2017, was available in **38 000 medical facilities serving 110 million customers**. In 2018, it launched its AI-powered diagnostic medical imaging service in 100 hospitals. The system is trained using **over 1 billion images from Tencent social network** complemented by anonymised patient data.

iFlytech, a former start-up, is now a consolidated company specialised in user interfaces and voice recognition. Its AI-enabled general practitioner medical robot passed the written test of China’s national medical licensing examination in November 2017. **The robot can automatically capture and analyse patient information and make an initial diagnosis.**

There are many start-ups in the area of face recognition. Among them Megvii claims that their cameras are able to identify **120 people per second** in a crowded train station, while LLVision is developing AI-enabled glasses storing **up to 10 000 images of suspects** on the device to achieve faster face recognition. They are part of an ecosystem of companies supporting **the government’s objective to set up a national facial recognition database that will store information about the country’s 1.3 billion citizens**, and be able to identify them within three seconds.²⁴

transparency and accountability present in the EU and the USA (e.g. Garfinkel, Matthews, Shapiro, & Smith, 2017) but notoriously absent in China.

The developments in China are very impressive. However, the JRC analysis identified some gaps in the current ecosystem and the roadmap to 2030. These structural gaps are not specific to AI but are also valid for other technically-based innovation ecosystems in China, a view supported by the main market analysts (Barton et al., 2017; Mubayi et al., 2017).

The first gap refers to the slow adoption of AI within traditional industries, many of which are supported by local governments and some have a high level of debt, where the impacts on jobs could be particularly difficult. A related issue refers to the gap between existing talent and its application within existing industries, including the distance between research and practical implementations in (traditional) industry.

Some experts also include the lack of fundamental research into AI. This gap implies the need to develop a sufficient quantity of AI talent and ensure that the local education system is able to supply it. The difficulties of increasing the pool of talent are also related to the set of knowledge and skills needed to succeed in the application of AI to a particular domain and the shortcomings in the innovation-resistant education system in general. The issue at stake is that it will not suffice to create an AI application - a programmer could do it almost mechanically with the appropriate directions - but to become a 'functional specialist' able to understand the domain of application, its needs and expectations, identify the best algorithm and approach, and then customise it and develop it further (Bacala et al., 2018).

■ 5.5 Summary and conclusions

This chapter highlights some of the main features of the AI developments in China, and in particular the synergy between government policy, private-sector investment, and research. China has identified AI and related applications as strategic to its future economic and social development, and to its role as a global superpower. As a result, it has put in place a strongly coordinated approach to AI, including government policy, industrial applications and research, with the objective of becoming the world leader in AI by 2030. This is an ambitious but achievable target as the the JRC analysis of the AI innovation ecosystem in China confirms. This opinion is shared by main market analysts (Barton et al., 2017; Mark Purdy et al., 2017). In fact, AI deployments with a high impact on daily activities have already started, including authentication processes, medical diagnosis, premium insurance, transportation and retail facilities, and security. Potential obstacles to the full development of the strategy relate to the difficulty of having AI adopted in traditional industries, and preparing enough specialists in the country able not only to advance technological innovation but also to adapt it to the needs of potential users.

“A final feature of the ecosystem refers to *the limited role attributed to critical views of AI in China.*”



PART 2:
MULTI-DIMENSIONAL
PERSPECTIVES



AI is shaping our ways of thinking, acting, socialising and working.

SUMMARY

In this chapter, we highlight some of the key dimensions of the ethical and societal debates on AI. As shown, **the debate about ethical and social implications of AI for individuals and societies needs to move forward and fast**. Political and regulatory frameworks seem to be lagging behind AI developments, which will soon pervade almost every aspect of our daily lives.

To counter this situation, 'regulatory foresight' and 'ethics foresight analysis' at the same level of 'technological foresight', are needed to keep up with emerging implications of AI. This, however, cannot happen without a strong commitment to the following issues: *clear ethical frameworks and guidelines*, including the right to choose between being cared for by a human or a robot, and the right to refuse being profiled, tracked, or measured, and *responsible AI design*, which is necessary to build trust in the applications.

We conclude that, to build and retain trust in AI, we need a multi-layered approach that includes the critical engagement of civil society to discuss the values guiding and being embedded into AI, public debates in different fora to translate these values into strategies and guidelines, and responsible design practices that encode these values and guidelines into AI systems, so that they are ethical by design.

ETHICAL AND SOCIETAL PERSPECTIVES

■ 6.1 Introduction

When addressing ethical and social impacts of technology, it is not just about avoiding detrimental or unintended consequences, or guaranteeing that humans have control over their technologies. Such externalist and instrumentalist perspectives tend to separate technologies from humans. That is, technologies as mere neutral tools devoid of values, and humans as sole masters defining the terms of engagement.

In reality, the relationship between technologies and humans is much more complex. On the one hand, technologies are built by specific people and organisations (Bijker et al., 1987), so they embody and replicate social norms, values, and other economic, ecological, political, or cultural forces at any given time. On the other hand, technologies shape how we work, move, communicate and ultimately live. Therefore, **ethical and social implications of technology need to consider this fundamental intertwining of the human and the technological domains.** ‘Humans are technological beings, just as technologies are social entities’ (Verbeek, 2011, p.4).

Calls for the development and use of AI to be placed ‘within limits of our fundamental norms, values, freedoms and human rights’ (EESC, 2017) are coming from a variety of viewpoints, including academia, civil society organisations, practitioners, think-tanks and businesses. There are encouraging signs of a commitment, or at least an interest in

To build and retain trust in AI we need critical engagement of civil society to discuss the values guiding AI and responsible design practices that encode these values and guidelines into AI systems.

collaborative work, between governments, industry and ethicists to design technologies for common good and identify specific principles for accountability, transparency, safety and social good (Pauwels, 2018). In this respect, we should also mention Floridi’s (2018) plea for the EU to use ‘soft’ digital ethics²⁵ as a good approach for AI, because it can support regulation and legislation and balance different values and interests.

In a concise format, we are presenting an overview of key ethical and societal issues in AI/ML, as currently discussed in policy, academia, civil society and business. The aim is to inform a comprehensive framing of ethical and societal challenges, based not only on fundamental values and rights, but also on wider discussions on who benefits from AI developments, to what end, and under what conditions.

Drawing on Stahl, Timmermans and Flick (2017), we make a distinction between issues that could have an impact at individual level (such as autonomy; identity; dignity; privacy and data protection) and those that have impact at societal level (such as fairness and equity; collective identity and good life; responsibility, accountability and transparency; privacy in connection to surveillance and datafication; democracy and trust).

■ 6.2 Overview of individual and collective implications of AI

■ 6.2.1 Challenges at individual level

It is important to think about how artificial intelligence might bring new challenges in relation to individual human beings. In this context, it is crucial to consider how the concepts of autonomy and identity of individuals as well as security, safety and privacy issues might change under the influence of AI.

a. Autonomy

Autonomy is one of the central values in Western ethics (Juth, 2009). It represents the idea that everyone has the capacity to know what is good or bad for them, to live according to their standards and that they have power over their thoughts and actions, thus promoting individual choice, rights and freedoms. AI might contribute to human development by extending human capabilities, or could stimulate productivity and prosperity and lead to active work until a later age. It could also have the opposite effects, as discussed in Chapter 9.



“It is crucial to think how the concepts of autonomy and identity of individuals as well as security, safety and privacy issues might change under the influence of AI.”

When it comes to interaction, there is also a need for people to know if they are interacting with another person or a machine. It is often not clear for us today if we are communicating via digital media with a bot or a human being. Also, the content presented to us and selected by algorithms interferes with our freedom to receive and impart information (Rathenau Institute, 2017).

b. Identity

According to Fearon (1999), identity is seen either as a social category with its membership rules or ‘a set of attributes, beliefs, desires or principles of action that a person thinks distinguish herself in socially relevant ways and that the person takes a special pride in’. In this case, the meaning of identity becomes very close to dignity or pride. As we interact and even develop close relationships with AI systems or robots, we need to understand better possible transformations of how we perceive them and how they shape our conception of the world. For instance, profiling, targeted advertising or other AI-powered techniques can deeply affect our identities (Floridi, 2015). The way in which human intelligence and cognitive capacities are affected by the interaction with machines and AI is an important area of research, as indicated in Section 2.3.

c. Dignity

Human dignity is the right of a person to be respected and valued and treated in an ethical manner, as mentioned in the Article 1 of the Charter of Fundamental Rights of the EU. It should be one of the basic notions used for the protection from harm caused to humans, and vulnerable groups in particular, and it should be respected while developing new technologies (Rathenau Institute, 2017).

Individuals’ rights and responsibilities could start eroding as a result of the increasing interaction of humans and machines (EDPS, 2018). At the moment, smart devices have no moral responsibility and that is why it could be potentially harmful to let them manage human beings (EGE, 2018). However, the European Parliament called for the EC to consider a specific legal status for robots (EP,

2017), which is still a controversial proposal when considering, for instance, that at the present time accountability is ultimately related to human responsibility (EECS, 2016).

d. Privacy and data protection

In the world of hyper-connected devices, crucial questions that are posed relate to privacy, security and data protection. Smart meters, toys, fridges and phones, to mention just a few, all have built-in AI systems that are sending our data to the manufacturers of these devices, often without our knowledge. AI applications in healthcare are even more sensitive, for instance when suggesting diagnosis or treatments (AI Now, 2017) or when data is (mis)used by insurance companies or technology companies (Powles and Hodson, 2017). The right to protection of personal information and privacy is therefore increasingly recognised as crucial in European societies (EGE, 2018).

The new GDPR that came into force in May 2018 is raising expectations in relation to the protection of personal data used by algorithms. The full impact of the Regulation will only emerge in the years to come, but it is clearly a major issue that needs monitoring.

6.2.2 Challenges at societal level

AI and big data have broader consequences for society and the well-being of communities, and nations worldwide. These changes have direct impact on reshaping of power relations and the transformation of the social contract. It is important to examine these changes in order to mitigate risks and dangers.

a. Fairness and equity

Will AI decrease or increase social inequalities? To address this question, we can distinguish two levels: social inequalities generated by the application of AI in the workplace, and individual level inequalities as a result of unfair AI-driven decision-making. With respect to the former, the evidence is as yet unclear, as discussed in Chapter 9. With respect to



the latter, research has shown that the use of algorithms could discriminate against particular groups or individuals, for instance in the criminal justice system (Chouldechova, 2017) and in recruitment processes (O'Neill, 2016). Bias could come from the training data or the algorithms, or from the biases of the people who build them (The Future Society, 2017; CNIL, 2017; see also Section 2.4). Besides technical research (such as de-biasing or fairness evaluation tools), and education of developers and society as a whole to be alert about potential unfairness, there is a need for the active involvement of government in fostering the culture of equity, free from stereotypes, and equal access 'across and within societies' (EGE, 2018) to benefit from the opportunities provided by AI.

b. Collective human identity and the 'good life'

AI impacts what we can consider the 'good life', that is, it can deeply transform how we live today and in the future. Similar to other technologies, AI can bring about social, economic, political and

ecological changes to our experience, critical capacity and perceived reality (Borgmann, 1987; Feenberg, 1991, 1999; Mitcham, 1980; Coeckelbergh, 2018; Higgs et al., 2000). In this sense, it confronts us with who we are and how we want to live in our society pursuing goals of well-being (Montréal Declaration 2017; IEEE, 2018).

c. Responsibility, accountability and transparency

Responsibility and accountability

We should think of responsibility in congruence with human rights and values (CoE, 2017). Accountability is connected to responsibility and it is extremely important because it deals with biases and discriminations caused by data mining and profiling. Accountability is seen as a necessary condition for the social acceptability of AI (Mission Villani, 2018). Besides the complexity of the online environment, locating responsibility and accountability becomes more difficult with the blurred relationship between

people and data, action and consequences, which could be considered as a wicked problem without a clear answer (Markham et al., 2018). In the context of AI, liability often follows responsibility. Current laws and regulations struggle with the emerging issues as discussed in Chapter 7.

Transparency

Transparency means not only transparency of algorithms (often referred to as ‘black boxes’), but also of data and Automated Decision Making (ADM). The importance of the latter is particularly salient when ADM is used by public agencies, such as predictive policing or risk assessment in the criminal justice system (AI Now, 2018). The decisions made by algorithms may not be understood or explained and it is not clear who is responsible for them. Therefore, transparency is seen as key for accountability, together with a right to meaningful information in automated decisions as foreseen by the GDPR, or even a ‘right to explanation’ (Selbst and Powles, 2017; Edwards and Veale, 2017; see Chapter 7). Although it is not likely that the code will ever be completely transparent because of IP rights, other information such as variables used, with their values and deviations, and the amount and type of training data used could be made explicit and contribute to an ‘effective transparency’ (CoE, 2017b).

Besides transparency, there is a need to audit algorithms, i.e. checking that they conform to certain properties (Kroll et al., 2016). Auditing should be done in collaboration with government and civil society. However, lack of access to the training data is the main problem for public auditing of algorithms (Villani, 2018).

d. Surveillance/datafication

Digital politics and media researchers argue that we live in a mass-surveillance culture, characterised by datafication of individuals’ lives that is difficult to escape (Van Dijck, 2014; Rathenau Institute, 2017; Korff et al., 2017). The world’s biggest digital companies control a large part of the online public sphere and hold most personal data. Through the analysis of big data, they obtain ‘truths’ about hu-

man beings, their behaviour, needs and aspirations (Markham et al., 2018, p.4), as well as control over the users’ social behaviour and consumer preferences (Bakardjieva and Gehl, 2017).

Connected to mass surveillance is the appearance of a phenomenon of voluntary self-surveillance that gained popularity through the so-called Quantified Self movement. Self-surveillance is seen through the use of tracking devices and corresponding platforms, with the goal of controlling one’s indicators for the allegedly better management of life and health (Vesnic-Alujevic et al., 2015). While thinking about controlling their lives, users are often not aware of the fact that there are more parties involved in the technology and that the data they leave are not only available to themselves. A right ‘not to be measured, analysed or coached’ by AI is needed (Rathenau Institute, 2017).

f. Democracy and trust

While dealing with ethical and societal implications of using algorithms, it is crucial to tackle their impacts on democracy and the trust of citizens. The effects of new pervasive technologies on power relationships between governments, citizens and businesses are not straightforward. As we have discussed above, the lack of autonomy, privacy, responsibility, accountability, transparency and possibility for mass surveillance can impact deeply our democratic systems. In addition to issues discussed above, we should consider:

Political profiling

Big data may be used not only for commercial advertising but also for political profiling, targeting and manipulation. A recent example is Cambridge Analytica and the way it used algorithms to target USA voters in the electoral campaign for the presidential election in 2016. The accumulation of large amounts of data in the hands of one company or political party may have serious consequences for the erosion of democracy (Brundage et al., 2018). Social scoring systems based on AI optimisations of social processes might be dangerous for privacy, equality and freedom (Botsman, 2017).

Bots, disinformation and deep fakes

AI has been used as a political advertising tool to influence voter behaviour but also to manipulate the public. Influencing public opinion through algorithm-enabled persuasion and manipulation, the use of political bots to spread false information has become easier than ever before. Bots have decreased trust in online environments and have created an advantage for populist politics that reduces trust (Brundage et al., 2018). The so-called ‘deep fakes’, false audio-visual content generated by AI, are a novel challenge that we are likely to face in the very near future (EPSC, 2018).

Freedom of expression

Free flow of information and the right to receive and impart information without interference are guaranteed by Article 10 of the European Court of Human Rights. These rights are being questioned because of the impact of algorithms and profiling in limiting users’ choice. In this context, platforms are becoming new information gatekeepers (Colombo et al., 2017) and their responsibility in maintaining freedom of expression needs to be clarified (Rathenau Institute, 2017).

Algorithms in governance

The use of algorithms in governance has become very popular among researchers, and policymakers. Researchers are mostly interested in the pervasiveness of algorithmic governance and its effects on power, freedom and human rights. According to Danaher (2017), ‘algocracy’ means using data mining and analytics to predict and control human behaviour. This is done through hyper-nudges: dynamic and personalised algorithmic nudging that can change and update based on the behaviour of their intended target (Yeung, 2016). This could threaten democratic societies if citizens lost agency and the possibility to participate freely in public life.

At the same time, there may be opportunities to leverage some of the characteristics of the current digital transformation, including algorithms in governance to increase democratic accountability and participation, and drastically change policy design

and implementation (see Chapter 12). A societal debate about the opportunities and challenges in the use of algorithms in government is needed to direct future developments.

g. Cumulative knowledge as a common good?

In the longer term, there is also a need to start to reflect at all levels in society on the extent to which the digital data that constitutes the cumulative knowledge about society should also be considered a national asset, in a similar way to cultural assets and landscapes. The Mission Villani alluded to this issue in arguing for an AI strategy structured around the goals of sovereignty and strategic autonomy (Mission Villani, 2018). In this respect, we need to reflect on the need to widen the view of digital data as a public good to also include a dimension of preserving data about society for future generations, and consider whether the state should exercise some degree of sovereignty over national data.

AI is of the utmost importance to help build coherence among the actions and policies to address the Sustainable Development Goals (SDGs) and achieve the EU Agenda 2030. The use of big data and AI is essential to identify synergies and trade-offs among the various strategies for achieving the individual goals, and develop a systemic approach to SDGs. The level of democracy of the digital world and the way we address AI affects the sustainability of our future.

■ 6.3 Summary and conclusions

In this chapter, we highlight some of the key dimensions of the ethical and societal debates on AI. As shown, **the debate about ethical and social implications of AI for individuals and societies needs to move forward and fast**. Political and regulatory frameworks seem to be lagging behind AI developments, which will soon pervade almost every aspect of our daily lives. To counter this situation, ‘regulatory foresight’ (EGE, 2018) and ‘ethics foresight analysis’ (Floridi (2018) at the same level of ‘technological foresight’ (EDPS, 2018a), are

needed to keep up with the emerging implications of AI. This however cannot happen without a strong commitment to the following issues:

Clear ethical frameworks and guidelines

There is a recognition in Europe at both EC, and national level that a clear ethical framework is needed. This is a rapidly evolving area with guidelines from the High-Level Expert Group expected by the end of 2018, as well as panels on ethics set up by national governments. We are also expecting work from the OECD, the UN, and professional organisations such as the IEEE to contribute to the debate. Then we will need to see how these proposed frameworks align and what gaps or inconsistencies, if any, they leave. At least in Europe, we can expect that **the ethical framework and guidelines should be compatible with the EU principles and regulatory frameworks.**

In reviewing the issues, two new rights are suggested:

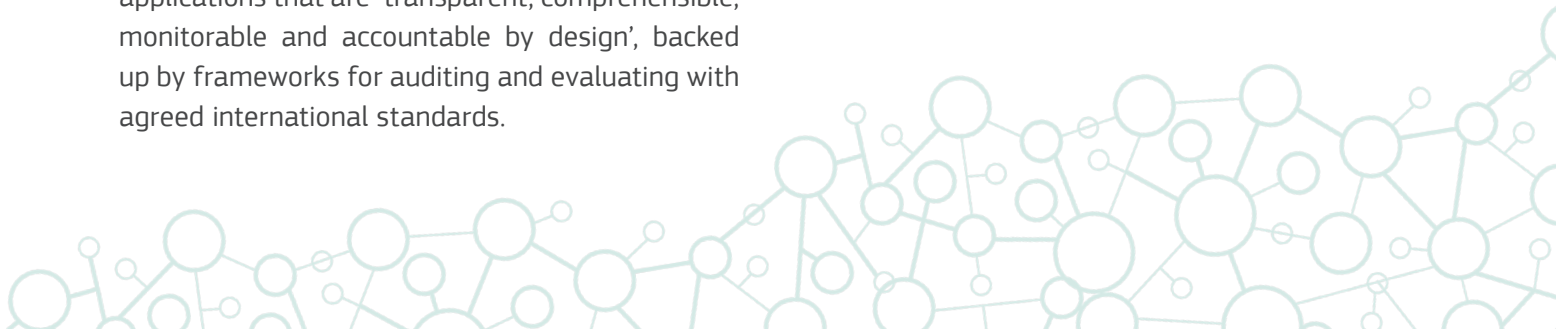
- 1) The right to a meaningful human contact. Every person should have the right to choose human contact and refuse to be cared for by a robot if he or she does not feel comfortable with it (EP, 2016). At the same time, robots should respect the autonomy of humans in decision-making.
- 2) The right to refuse being profiled, tracked, measured, analysed, coached or manipulated (CoE, 2017; Rathenau Institute, 2017).

Responsible AI design

The translation of values and norms into the design and operation of AI systems should be part of regulatory frameworks and guidelines. For instance, ‘privacy by design’ or ‘data protection by design and by default’ are now part of the GDPR. Similarly, we should think of designing AI applications that are ‘transparent, comprehensible, monitorable and accountable by design’, backed up by frameworks for auditing and evaluating with agreed international standards.

This **approach to responsible design is necessary to build trust in the applications**, but it is not by itself enough. As introduced in Section 6.1, technology is a social product which embodies and replicates social norms, values, and other economic, ecological, political, or cultural forces at any given time. Therefore, it is crucial to develop a multi-layered approach to AI design, including:

- **The critical engagement of civil society** in the development and use of algorithms in different application areas with discussion about who benefits, and what values should be encoded into the systems (Jasanoff, 2013; Boucher et al., 2014);
- **The establishment of multi-stakeholder fora** to promote such public debate and translate their outcomes into socio-political strategies for AI enforcing ethical and social values (Cath et al., 2018; Stilgoe, 2017);
- **The design practice of considering ethical and social values** (Aldewereld et al., 2015; Van den Hoven et al., 2017; Nascimento et al., 2016), and potential sources of bias in the entire development process of the AI system: from the selection of multidisciplinary and diverse design teams, to the selection and labelling of training data, evaluation of outputs, and assessment of outcomes.





SUMMARY

In this chapter, we highlight the main legal issues raised by AI that concern, among others, fundamental rights, including personal data, the transparency and explainability of algorithms, and the suitability of existing liability rules. We also review the European legal framework with respect to data ownership, access and sharing regimes and intellectual property rights, and consider the different approaches to regulate disruptive technologies. We note the tensions between protecting rights of individuals and firms and encouraging innovation with trying to maximise openness and transparency. The task of policymaking is to manage change and regulate the adoption of new technologies to ensure that they are accepted to society and respect our fundamental values. We conclude that Europe is well placed to establish a distinctive form of AI that is ethically robust and protects the rights of individuals, firms, and society at large. For example, the General Data Protection Regulation, opposed by many when in preparation, is now perceived as a European asset and is inspiring similar approaches outside Europe. Extending this notion, we should consider the high standards of the European legal and regulatory landscape in a similar way as those in environmental quality which are an asset for Europeans and their future generations to build upon, not a barrier.

LEGAL PERSPECTIVE

■ 7.1 Main legal challenges identified in European AI strategies

As indicated in Chapter 4, several European Member States, as well as the European Commission, are considering the challenges and opportunities afforded by AI. This section summarises the main legal challenges raised by widespread use of AI as identified in AI strategies and policy documents²⁶. Section 7.2 and 7.3 develop more in detail data and Intellectual Property (IP) related issues.

One of the most critical issues in the strategies examined is how to ensure that AI algorithms operate in full respect of the European values and fundamental rights²⁷ as well as of its ethical principles (see Chapter 6). Key elements of concern are potential forms of discrimination by algorithms, and the need to develop not only ethical guidelines but also tools and design principles such as ‘ethics by design’. Other lines of action include mandating the assessment of impact on human rights or discrimination when algorithms affect individuals or are used in public administration, or require algorithm auditing mechanisms. Within this context, there is a need to reflect on if and when automated decision-making should be allowed. Naturally, some users and sectors (e.g. justice, health) are more sensitive than others in this regard.

In terms of *personal data protection*, certain provisions in the General Data Protection Regulation (GDPR) (EC, 2016a) have addressed some of the concerns mentioned above. These include: i) data protection by design and by default, ii) privacy impact assessments, iii) (a relative) right to explanation as concerns the logic involved in automated decision-making, and iv) the right not to be subject to a decision based solely on automated processing²⁸. These provisions could serve as inspiration for other areas. As the GDPR has only recently come

The task of policymaking is to manage change and regulate the adoption of new technology so that they are acceptable to society and respect our values.

into force, there is a need to monitor its application in the AI context. Some authors question the suitability of the GDPR to deal with AI (see Pouillet 2018a and b; Delforge et Gerard 2017; Watcher 2017) on the basis of the difficulty to comply with the principles of purpose limitation (since AI systems are by nature general-purpose systems) and proportionality, as well as the requirement to obtain legitimate consent. Criticism also concerns the perceived limitations as regards the so-called right to explanation.

Within this context, *transparency* and *explainability* of algorithms are regarded as essential to understanding how decisions are made and give citizens, users and operators the possibility to challenge them. However, it is not yet clear how these requirements should be addressed. Some references may be found in public documents concerning: i) the processes to be put in place, ii) the interference with IP protection, iii) the responsible actors to scrutinise algorithm decision making, and

iv) the development of guidance to assist in explaining AI. The EC is analysing the issue and has commissioned a study, *AlgoAware*²⁹, to shed light on these (and related) aspects.

Liability is another issue frequently raised in the legal literature and in policy debates. This is an essential aspect to ensure that possible victims have efficient redress systems, but also that industry has legal certainty. The complexity of AI systems and the variety of actors intervening in the value chain, coupled with its self-learning and autonomous abilities, can make it very difficult to allocate responsibilities. Moreover, since the liability regime has been only partially harmonised, this subject is also a matter of national law.

The main question is therefore: does the current liability and safety framework provide adequate mechanisms to deal with AI products and services liability? If not, what adjustments are necessary? This issue is being explored by the Expert Group on Liability and New Technologies, recently created by the EC (see Chapter 4) which is expected to report in mid-2019. The group is split into two assemblies. The first assembly will focus on the Product Liability Directive³⁰ and will examine to what extent the provisions of the Directive are adequate to solve the questions of liability in relation to traditional products, new technologies and new societal challenges³¹. The second assembly, dealing with new technologies, 'will assess whether and to what extent existing liability schemes are adapted to the emerging market realities following the development of the new technologies such as Artificial Intelligence [...] and assist the Commission in developing EU-wide principles which can serve as guidelines for possible adaptations of applicable laws at EU and national level as regards the new technologies [...]'.³²

By mid-2019, the group is expected to issue a guidance document on the interpretation of the Product Liability Directive in light of technological developments and to publish a report on the broader implications for potential gaps in and orientations for

the liability and safety frameworks for AI, the IoT and robotics.³³

AI products or services, like everything else, cannot be 100% guaranteed to be safe. A possible approach to manage safety is to define acceptable levels of safety for identified risks to be characterised by the probability of occurrence and severity criteria. This requires assigning quality levels for software and processes applicable from design to validation, according to the criticality of the software modules and corresponding datasets contributing to the risks identified. Additional mitigation measures of a technical (e.g. diversified parallel developments) or operational nature (e.g. a certain level of human control) may need to be implemented when residual risks are considered unacceptable.

In addition to the areas mentioned above, national AI strategies also refer to the challenges posed by cybersecurity (see Chapter 10), and the need to facilitate testing platforms, including regulatory sandboxes, to support innovation as well as other legal areas such as consumer protection and competition law. Both legal commentators and the general press have recently warned of possible anti-competitive behaviour that could take place in this context. Such concerns relate to collusion practices (including, for example, tacit collusion in price fixing), price discrimination, data concentration and to the role of privacy in competition assessment. National public authorities and stakeholders have recalled the need to remain vigilant in this regard (UK House of Lords 2018). The increasing power of certain platforms also raises concerns as regards their accountability as quasi-public actors in the digital society (EPSC 2018).

The regulatory regime for non-personal machine-generated data (or sensor data) has been a recurring question in the policy debate concerning AI or, more generally, the industry 4.0 revolution. This specific issue is explored further below.

7.2 Ownership, access and data sharing

The discussion on the ownership of data has been extensively debated in the history of information technology law. The advent of the big data era brought a new impetus to this discussion both for both personal and non-personal data (Osborne Clarke, 2016). This section will focus mainly on the current discussion on the legal regime for non-personal machine-generated data although in certain cases the distinction may be artificial (see also Section 12.1).

The issue of the lack of a comprehensive legal framework for machine-generated data was raised in the Communication on Building a European Data Economy (EC 2017a). The EC, worried about the under-exploitation and lack of trade in machine-generated data, put forward a series of legislative and non-legislative options for discussion, among them the creation of a new data producer's right. However, in the later Communication Towards a common European data space (EC 2018c), data ownership is hardly mentioned. Instead, the EC focuses on access to data and on fundamental measures such as a proposal to review the PSI Directive (EC, 2018d), a recommendation on access to and preservation of scientific information (EC, 2018i) and a guidance document on the business-to-business and business-to-government exchange (EC, 2018f). Recently, the EC has set up an expert group to assist in assessing issues around business-to-government data sharing.

There is no legal or statutory title providing for ownership of data either by traditional property rights or by intellectual property since data per se is not a copyright or patentable subject matter. At most, under certain circumstances data could become protected by the database's *sui generis* right or by the Trade Secrets Directive (EC, 2016c), provided that certain conditions are met. In practice, provisions on data access, use and sharing are regulated at the contractual level on the basis of a de facto ownership. Focusing on competition law, some authors point out that the strict requirements of the

essential facilities doctrine as well as the criteria developed to assess possible abuse of a dominant position in refusals to licence make traditional competition law tests of limited use to promote access to data (Drexel et al., Drexel 2017, Pouillet 2018b). Recent studies predict that conflicts are likely to arise in the near future, but recognise that legal intervention may be premature at this stage (Osborne Clarke, 2016 and in a more nuanced sense also Deloitte, 2018).

The need to create a new intellectual property right to incentivise data production and collection, commercialisation, and access is something that has been explored by scholars with diverging views. For example, Zech (2016) and Van Asbroeck et al. (2017) favour the introduction of new ownership rights for industrial data, while Drexel (2017a), Hugenholtz (2018) and Ramalho (2017a) are quite critical of such an idea. In short, recent economic research falls short of making a clear case for creating new IP rights for data (Kerber, 2016; Kerber & Frank, 2017; Duch-Brown et al., 2017).

While discussions on ownership are still inconclusive, the debate has shifted to the issue of data access. In this regard, different options are being considered, including specific regulation, creating a new *qualified* right to access data, or a system of compulsory licences to grant access to data, either on a horizontal or sectoral basis³⁴ (Drexel 2017b, JIIP 2018; Mezzanotte, 2017; Weber, 2017). Some scholars have even suggested a specific copyright exception, a fair use or an open norm that cover AI and other related uses without the authorisation of the copyright holders (among others, Geiger et al., 2018, and Schafer et al., 2015). Here we should recall that in 2016 the EC tabled a legislative proposal that, among other things, proposed a mandatory exception for text and data mining (TDM) for research purposes subject to a series of conditions (Article 3 in EC 2016b).³⁵ Although generally welcomed by the research community and by copyright scholars, the proposal has not been exempted from criticism in relation to the conditions under which it should operate, being more restrictive than

the regime applicable in other countries (Geiger et al., 2018; Samuelson, 2018). All in all, copyright exceptions would only authorise the re-use but would not, strictly speaking, facilitate access to datasets.

As indicated at the beginning of this section, some scholars have noted that the distinction between personal and non-personal data is becoming more and more blurred. It is not only that non-personal data or anonymised data can be subject to techniques linking them back to individuals, but also that recent reforms have made clear that the notion of personal data includes data that allow individualisation, even if they do not reveal the individual's civil or legal identity³⁶ (Poullet, 2018b). This must be taken into account when considering specific regimes for sensor or industrial data that may also include data susceptible to being considered as personal data.

“ The EU can provide guidance to the European countries on regulatory issues of AI while fostering the development of a single market for new technologies. ”

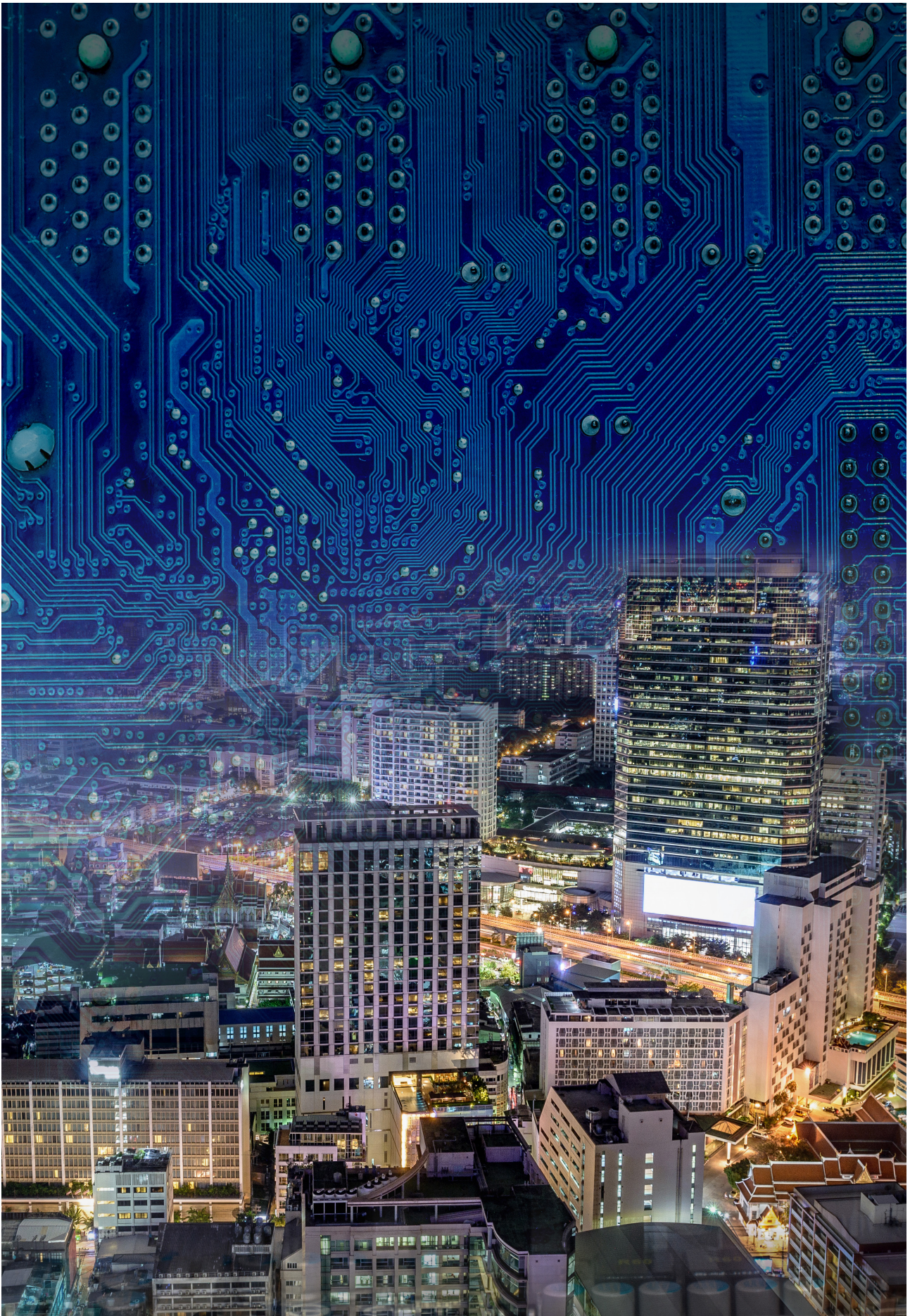
7.3 The protection of AI inventions/ creations by intellectual property rights

National AI strategies have so far paid little attention to the challenges AI poses to the IPR legal framework. However, the increasing importance of AI technologies and the gaps identified by scholars in both the copyright and patent system (e.g. Guadamuz 2017; Michaux 2018, Ramalho 2018; 2017b; Schönberger 2018 or WEF 2018) reveal the need for further investigation.

As regards the patentability of AI, the announcement of the European Patent Office (EPO) of specific guidance on patentability criteria for AI is particularly welcomed. The new section on Artificial Intelligence and Machine Learning provides some clarification on when AI-related subject matter has a technical character.³⁷ The protection of AI-generated works or inventions seem to be more problematic. In light of the humanist approach of copyright law, it is questionable that AI-generated works deserve copyright protection. As regards patent laws, although a priori nothing prevents AI-generated inventions from being patented, the assessment of the non-inventive step or the allocation of ownership may also raise issues that are not entirely solved.

While some copyright scholars clearly advocate for AI-generated works to be placed in the public domain, others have put forward a series of proposals aimed at ensuring a certain level of protection. With notable exceptions, these proposals are still too vague. They do not always sufficiently detail the possible elements underpinning such protection.

There is no doubt that certain AI-generated creations/inventions may share the characteristics of information goods – non-excludable and non-rivalrous nature – that justify the creation of quasi monopolistic rights to foster innovation and commercialisation. However, there are concerns whether incentives are needed, especially in cases where the investment cost is low, and what consequences such rights might have on the market, including on



creations or inventions made by humans. Would more property rights encourage or rather deter innovation? We clearly need to investigate these issues further from a law and economics approach before favouring one solution or another.

■ 7.4 Regulatory approach

Every disruptive technology induces changes with an impact on society. Regulation is often considered necessary to address some of these impacts. The first difficulty when deciding about the most appropriate regulatory framework for AI is to anticipate the impact it may have on society. The second is to cope with the pace of AI advances. Other challenges include the level of complexity of AI, its autonomous and self-learning features and its pervasiveness across sectors. These features require both qualified interdisciplinary capacity (to better understand technologies and predict impact) and flexibility (to envisage frameworks that, while enforceable, can adapt with time) in policymaking. The implementation of regulatory sandboxes may be particularly useful in this respects.

While a hard-law approach may be needed in certain instances, co-regulation and soft-law schemes are often preferred in order not to stifle innovation – even if there may be cases where regulation also acts as an enabler for innovation. The role of certification and standardisation bodies, as well as that of other organisations representing stakeholder interests (notably citizens and consumers), are considered important. There is also a discussion on the approach of ‘regulation by design’, the need to act in an anticipatory vs. reactive way, innovation or precautionary principles, or to have either sectoral or horizontal regulation. While some institutions have proposed the adoption of regulation addressing part of the challenges highlighted in this section (EP 2017), most of them remain rather cautious.

In a global context, governance models should be explored in order to ensure that regulatory efforts are not isolated and to avoid regulatory patchworks leading to forum shopping (EGE 2018) or to a

“Every disruptive technology induces changes *with an impact on society.*”

regulatory race to the bottom (Goldfarb and Trefler, 2018a). Several actors have called for a global dialogue on the establishment of norms and ethical rules or principles (e.g. UK House of Lords, 2017, or Mission Villani, 2018).

Considering the various issues at stake, and specifically their ethical dimension, it becomes crucial, whichever the approach adopted, to ensure stakeholder involvement and, in particular, the engagement of civil society in the law-making process. Inclusive and/or innovative platforms are being created in an attempt to respond to these challenges. Thus, as already noted, the EC has recently established the European AI Alliance to engage in a broad and open discussion of all aspects of AI development and its impact on the economy and society. The UK is ‘establishing a Ministerial Working Group on Future Regulation to scan the horizon and identify the areas where regulation needs to adapt to support emerging technologies such as AI’, and a ‘Regulators’ Pioneer Fund to support regulators to develop new approaches which enable emerging technologies’.³⁸ Also in the UK, a recent report has proposed the creation of a new body, the Office for Responsible Technology, with three main functions: empower

regulators, inform the public and policymakers and support people to find redress (Miller et al., 2018).

7.5 Summary and conclusion

AI is most likely going to have significant impacts on our lives, even more than past waves of digitalisation. As with any major technological change, this will bring significant societal benefits, but will also pose new challenges across many economic segments, including transport, energy, healthcare, defence, education, leisure, farming and finance. The task of policymaking is to manage change and regulate the adoption of new technology in order to ensure that they are acceptable to society and respect our fundamental values. The EU can provide guidance to the European countries on regulatory issues of AI while fostering the development of a single market for new technologies.

To foster the development and uptake of AI, citizens and industry operators need to be reassured that it complies with ethical and regulatory frameworks. Consumers and users of AI devices should benefit from mechanisms allowing redress in case of damages and from tools that allow them to supervise decisions taken by AI systems. Economic actors also need sufficient legal certainty to invest in AI technology. The initiatives adopted both at the European and national levels provide some hints on how these aspects may be addressed.

The future EC guidelines on the ethical development of AI and the findings of ongoing studies on transparency and explainability, along with parallel initiatives undertaken by other stakeholders, may lay the foundations for a fairer AI. The work of the Expert Group on Liability and New Technologies is also expected to bring clarity on the most suitable rules to cope with unanticipated damages caused by AI systems.

Compared with the other AI champions, such as the USA and particularly China, Europe is noted for its high regulatory standards for personal data protection. The GDPR is inspiring regulation across the

world and informing good practices for the use of personal data even beyond EU data markets. The European policymaker can have a role in the setting of standards for human-centred AI, as has already been done in relation to privacy.

Access and use of quality data are fundamental for many AI applications. However, empirical evidence shows that an optimal framework for trading and sharing of data has still to be realised (see Section 12.1). Within this context, scholars have put forward a series of proposals to adjust the regulatory framework for non-personal machine-generated data. These go from enacting new ownership rights on data, to establishing access regimes or incentivising data sharing through soft-law instruments. Sectoral approaches and solutions focused on regulating access (rather than on privileging the creation of new rights) are gaining prominence. In a context where the European data landscape is decentralised and fragmented (in comparison with that of other regions), the discussion on data governance becomes even more relevant and opens opportunities for different models to emerge. (See 12)

In addition, the potential capacity of AI systems to generate inventions or creations has attracted the attention of scholars, some of them sceptical about the capacity of the existing IP framework to accommodate this phenomenon. Further economic and legal research is needed to assess to what extent adjustments to the legal framework or the creation of new rights are needed at all.

Last but not least, we have briefly referred to the needs and tensions in the general debate on how to better regulate AI. In the discussion around regulating AI, **the question is not only if there is a need for regulatory intervention but mainly which approach should be adopted to find a balance to support innovation and at the same time protect wider societal interests.** The question is also how to ensure that policymakers are well equipped to make adequate and timely decisions in a rapidly changing, technology- and market-driven environment with an intrinsic global dimension.



AI will change what and how we learn

SUMMARY

From an educational perspective, we observe in this chapter that AI has potential positive impacts on shortages of skills, learning, and teaching. Three crucial points stand out from the review: firstly, the need to understand better how the interaction with AI impacts human intelligence in cognitive capacities in both adults, and even more importantly, children. Secondly, we need to think beyond current needs and practices, and consider how AI is likely to change the relationship between education and work, and human development. Thirdly, we highlight possible risks related to AI in education, particularly privacy and ethical ones. In this chapter, we also provide an initial overview and geographic distribution of the academic offer of study topics related to AI. This is only the beginning of a monitoring and analysis process of academic supply but it already highlights potential priority areas where greater effort is needed to prepare society to fully harness the benefits of AI.

EDUCATIONAL PERSPECTIVE

8.1 Introduction

Since the beginning of the 1980s, and until recently, educational applications of AI have mainly focused on the knowledge-based approach (see Sleeman and Brown, 1982). The most prominent line of research has been concerned with Intelligent Tutoring Systems, or ITS (Woolf 2009). These systems have been successful mainly in relatively limited and unambiguous domains, such as mathematics and physics (Ritter et al., 2007; Graesser et al., 2005).

As student testing plays an important role in many educational systems, numerous projects try to use AI for automatic test generation and assessment. AI is also used to diagnose student attention, emotion, and conversation dynamics in computer-supported learning environments, for example for course development and management, in an attempt to generate optimal groups for collaborative learning tasks, and to recognise patterns that predict student drop-out (Nkambou et al., 2018; Rosé et al., 2018). To do this effectively, large datasets are needed for training the systems, and the student behaviour needs to be actively monitored to provide feedback for personalised learning. This creates technical needs to monitor students unobtrusively, for example using video processing and remote eye-tracking, with associated ethical and regulatory challenges.

In special needs education, AI-based approaches have shown potential, for instance, in the early detection of dyslexia (Drigas and Ioannidou, 2013). A well-published example is the Swedish company Lexplore that has developed a system that quickly

Education and training are crucial to harness AI, but AI can also help us rethink what competences and skills will be needed in the future to live a fulfilling life, and master the digital transformation of our society.

scans for students at risk, and detects dyslexia by tracking reader eye movements (Jakobbson, 2017). AI-based systems have also been successfully developed for the diagnosis of autism spectrum disorder and attention-deficit disorders.

This section will take a closer look at some of these challenges related the impact of AI on skills demand, and on learning and teaching.

8.2 AI impact on skills demand, learning and teaching

8.2.1 Direct AI impact on advanced skills demand

The development of new AI and ML models requires very high levels of competence in several areas. This is one of the reasons why AI experts are now being paid very high salaries. The number of AI neural experts may be doubling annually, but the basic knowledge needed for state-of-the-art work in this area requires advanced levels of scientific, mathematical and technical skills that are not easy to acquire. Development of new AI methods requires a good understanding of statistics, linear algebra, differential equations, as well as computer architectures and esoteric³⁹ programming approaches and tools. The required skill set is scarce, and estimates of the actual number of people with AI skills vary significantly, ranging from tens of thousands to a hundred thousand. It is estimated that there are around 22 000 PhD researchers in AI and some 5 000 people who have written academic articles or presented at AI conferences in recent years.⁴⁰

It is expected that the high visibility of AI and the current demand will relatively rapidly direct talent to this area. As an example, since its launch in May 2018, about 90 000 students from over 80 countries have enrolled in the six-week Elements of AI –course organised as part of the AI Education programme at the Finnish Centre of AI.⁴¹ Due to the high wage differentials, current students in statistics, mathematics, mathematical physics, computer and chip design, and perhaps neurophysiology may reconsider their career paths and find new opportunities as experts in AI. Moreover, high-level AI competences may also emerge from unexpected places, for example, through open software and open hardware communities.

One rather immediate result of this situation is that high-level AI talent could probably be provided as a service, similarly to the Infrastructure of software as a service. This may mean that there is not going

to be massive needs for high-level AI competences. Another issue is the likely impact of AI on changing skills demands for the existing workforce and the population at large. Some of these are mentioned in the next section on the impact of AI on learning. In Section 8.3, we report on a first scan of the type and distribution of AI-related study topics in European countries.

8.2.2 Impact of AI on learning

The way in which human intelligence and cognitive capacities are affected by the interaction with machines and AI is an important area of research. Whilst recent scientific literature has focused on the interactions between AI systems and adults, there are important differences in the way children deal with artificial systems that need to be researched further.

In general terms, AI can be used in three different ways that may have different implications for the development of human cognitive capabilities both in children and adults.

First, AI can support existing capabilities. When competences are understood as combinations of domain-specific expertise and behavioural repertoires⁴², AI can now reduce the need for domain specific knowledge and therefore make transversal and domain-independent generic skills more important.

Second, AI can speed up cognitive development and create cognitive capabilities that would not be possible without technology. The mechanisation of human work has made possible things that would be impossible without technology; similarly, the mechanisation of cognitive work makes possible new activities that have not been possible before. This has already happened. It would be impossible to design a modern microprocessor or a neural chip without computer-aided design tools that use extensive bodies of design knowledge.

Third, AI may reduce the importance of some human cognitive capabilities, or make them obsolete.

For example, as AI can convert speech to text and vice versa, and do mathematical calculations, dyslexia or dyscalculia may become socially less important than it has been in the past. This has clear benefits for individuals, but the overall impact of making skills redundant through AI is not easy to predict. From a pedagogic point of view, it may be more beneficial to use AI to help people to develop competences that allow them to overcome difficulties in reading and counting, instead of using AI to make redundant skills that underpin important cognitive capabilities.

It is also often assumed that AI systems enable new levels of personalisation and diversity for information systems. Much of this personalisation, however, results from fine-grained categorisation that puts users into pre-defined classes. Although these systems may be able to efficiently simulate personalisation, they do not necessarily support deeper levels of diversity. AI systems can be excellent predictive machines, but this strength may be an important weakness in domains where learning and development are important.⁴³

8.2.3 Impact of AI on teaching

There are some clear opportunities for AI in teaching, such as student assessment in its various forms, and personalised tutoring systems. These systems use a knowledge-based architecture with a *domain model* that describes the area to be learned and a *student model* that describes the current state of the student's knowledge and learning. An expert system or pedagogical model manages the introduction of learning materials to the student through an adaptive and interactive user interface. As student behaviour and learning can also be monitored in ITS environments in great detail, intelligent tutoring environments have also been an important source of data for research into learning (Porayska-Pomsta 2015).

As these supervised AI learning algorithms are based on historical data, they can only see the world as a repetition of the past. This has deep

ethical implications. When, for example, students and their achievements are assessed using such AI systems, the assessment is necessarily based on criteria that reflect cultural biases and historically salient measures of success. Supervised learning algorithms create unavoidable biases, and these are currently being debated extensively.

The rapid advances in NLP and AI-based human-machine interfaces will also generate new pedagogical possibilities. For example, learning by teaching machines shows clear potential, while real-time machine translation also opens up new possibilities in language learning.

Whilst it is possible to imagine many exciting possibilities for AI in teaching, **without clear policies** that put emerging technical possibilities in the broader context of the transformation of **education and the future of learning, educational AI will probably mainly be provided as solutions to existing problems**. Instead of renewing the system and orienting it towards the needs of a post-industrial economy and knowledge society, AI may therefore mechanise and reinvent outdated teaching

“Educational AI will probably mainly be provided as solutions to existing problems.”

practices and make them increasingly difficult to change. It is therefore crucial to look beyond the current practices and ask fundamental questions about what competences and skills are needed in a digitally transformed society, and how we should teach them.

8.3 AI skills and academic supply

To have a first understanding of the current academic supply on AI, we analysed all the 780 universities in the EU with a website containing both study programmes and keywords related to AI. A first screening showed that a higher number of bachelor studies have pages in local languages compared to English, while master studies are mostly advertised in English and doctoral studies are almost only in English. This implies that our analysis, based on English only, has been able to capture almost all doctoral programmes in the EU countries, and most masters, but clearly underestimates the offering

numbers from undergraduate courses in national languages. Notwithstanding this limitation, to be addressed in future work, we identified 693 study topics related to AI.

Figure 10 shows the country profiles for academic offer (at all levels) by thematic field of the programmes, which are the result of the combination of techniques and application fields taught in the academic programmes in the AI domain: computer vision and natural language processing, ML and DL, and robotics. The first one is a combination of two subfields: methods and developments in image processing, object detection and computer vision, and the field that treats natural language, including speech recognition, machine translation and computational linguistics. The figures show that in most countries a high proportion of programmes focus on robotics (81% in all EU), followed by computer vision and NLP (12% in all EU) and computer vision and NLP (7%).

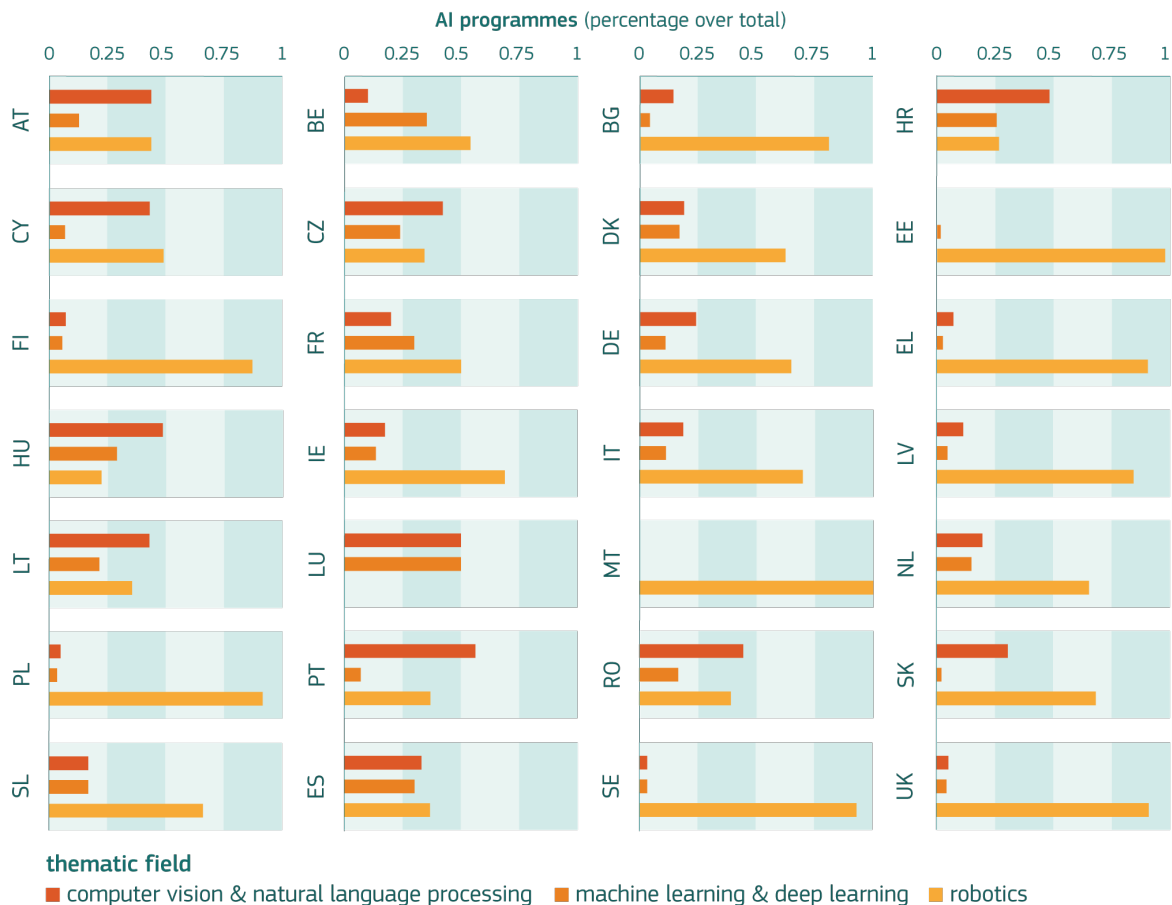


Figure 10. AI thematic profiles in academic programmes (as % of total programmes)

In Figure 11, we show a geographical distribution of universities offering study topics related to AI. The locations are superimposed on the distribution of science and technology skills at NUTS2 level, to show where there are significant gaps and opportunities to upskill the local workforce (see also Chapter 13). It is worth underlying that this is only a preliminary analysis and that the academic offer is underrepresented where study topics are offered in the national language, as at this stage we only analysed material available in English.

8.4 Summary and conclusions

In this chapter, we have identified that AI already contributes to education particularly through intelligent tutoring systems, student assessment, and helping to identify potential learning challenges.

We have explored potential impacts on shortages of skills, on learning and on teaching. Three crucial points stand out from the review: firstly, the need to understand better how the interaction with AI

impacts human intelligence in cognitive capacities in both adults, and even more importantly, children. Secondly, we need to think beyond current needs and practices, and consider how AI is likely to change the relationship between education and work, and human development. Thirdly, we have highlighted possible risks related to AI in education, particularly privacy and ethical ones (for further reading see Tuomi, 2018).

We have also provided an initial overview and geographic distribution of the academic offer of study topics related to AI. The academic supply of study topics related to AI shows a strong focus on robotics, reflecting one of the current areas of strength in Europe. Further research is needed to improve on this analysis, which nevertheless provides some material for reflection on areas of the current low level of science and technology skills where reinforcing the provision of courses may be considered important. This is relevant to the discussion in Chapters 12 and 13 on possible strategies to start preparing society, and the most vulnerable regions, to the challenges that AI will bring.

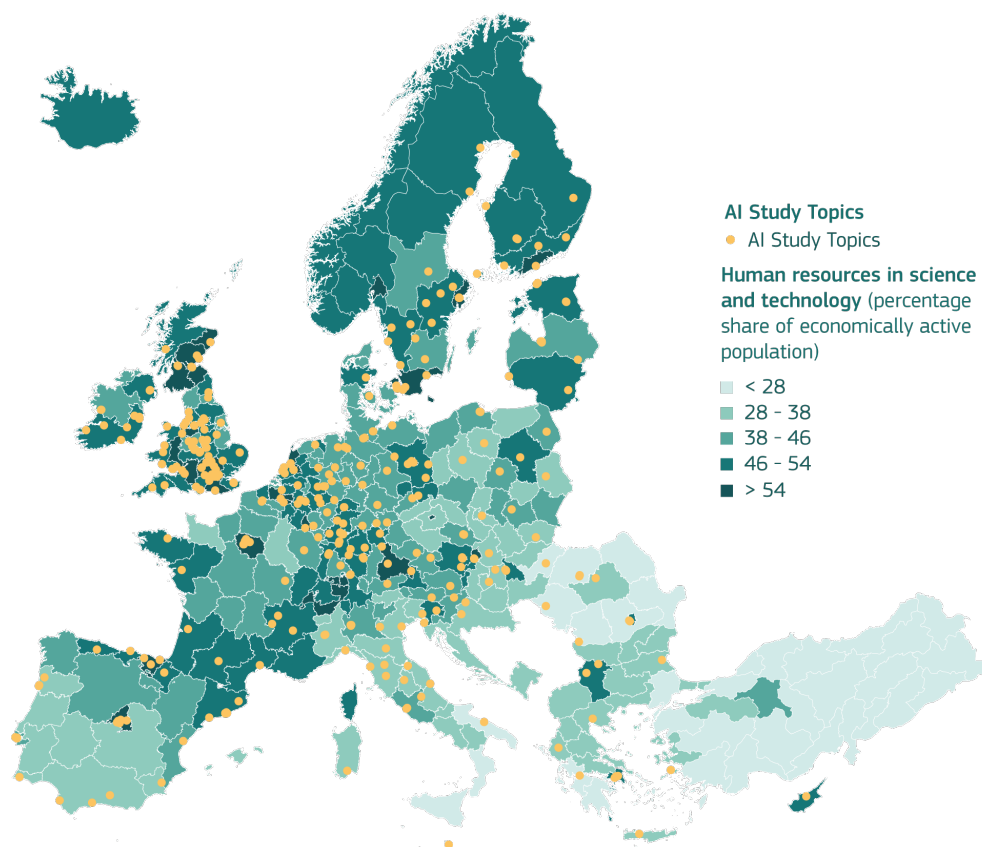


Figure 11. Distribution of universities offering AI study topics over human resources in science and technology 2017.

Source: Eurostat Regional Yearbook 2018, and JRC



AI has a wide range of potential economic implications for employment and income distribution.

SUMMARY

In this chapter, we analyse the literature and evidence available to date on the potential impacts, both positive and negative, with respect to work, growth and inequality. In relation to work, we find that neither theory nor evidence are very conclusive at the present time. AI could complement and enhance human activity, replace an increasing number of routine tasks, or both. Another area of uncertainty is the extent to which AI has the potential to spur economic growth. When considered as a general-purpose technology, AI could spread across many jobs and industrial sectors, boosting productivity, and yielding strong positive growth. To the extent that ML generates new insights from data, it may also contribute to the production of new ideas and innovation itself. Economic growth models are starting to explore various scenarios but there is no empirical evidence yet that favours one or the other.

When it comes to inequality, we find that AI can affect unfavourably the distribution of income through many channels. The most discussed concern job polarisation, reduction in job quality at the lower end, and also greater difficulty for lower-skilled workers to adjust to change and find new jobs, with longer periods of unemployment than those with higher qualifications and skills. Some regions of Europe are more vulnerable to these challenges if no action is taken to prepare society. Therefore, a very proactive strategy needs to be put in place to build the resilience of regions across Europe.

ECONOMIC PERSPECTIVE

In this chapter, we discuss the potential impact of AI on jobs, growth and inequalities, based on the scientific literature and evidence available to date. There are concerns over the potential impact of AI on the labour market and income inequalities. Some alarmist studies predict massive job losses as AI replaces human labour. On the other hand, historic evidence shows that previous waves of innovation replacing workers with machines ultimately generated more jobs and higher incomes. We examine if there are reasons why this time things could be different in the case of AI. AI is likely to improve productivity and economic growth over a wide range of sectors in the economy. This could give a welcome boost to the current productivity slow-down. At the same time, the non-rival and general-purpose characteristics of AI could speed up change across multiple sectors and lead to greater impacts than past waves of innovation, creating more friction in labour markets. We also focus attention on the distributional impact of AI and the risk that it may exacerbate among existing regional and income disparities.

■ 9.1 Potential impact of AI on jobs

Debates on the impact of technological change on employment go back at least to the beginning of the Industrial Revolution. Back then, the Luddites attacked spinning machines because they displaced human labour in the textile industry. However, the long-run historical evidence eases our fears about being made redundant by machines. Ever since the Industrial Revolution, and despite many subsequent waves of innovation and mechanisation, human employment and incomes have strongly increased, not decreased. Still, with every major new invention, these fears resurge – as with AI.

History suggests that technological innovation may be a substitute for human work but ultimately increases employment. The nature of AI may accelerate and boost economic growth and create new jobs. We need to ensure that this benefits us all.

Do we have any reasons to assume that this time will be different, or can we continue to draw comfort from the historical evidence to date?

Traditional economic theory revolves around the magnitude of the substitution effect between workers and machines. It suggests that machines (including AI algorithms in computers) and human labour are not perfect substitutes. Machines may replace workers but cannot do entirely without them; they are to some extent complementary.

When the demand for machines increases because they are more productive and cheaper than workers, this will also increase demand for keeping workers who are necessary to operate the machines as the goods produced with the machines become cheaper. Moreover, machines will increase workers' productivity and wages for the remaining human tasks. In this view, technological improvement boosts the demand for labour. The long-run historical evidence seems to support this.

More recently, however, economists have started to look at this from a new perspective. First, recent evidence shows that while employment may be growing in advanced industrial economies, a rapid increase in the capital share is causing a relative decline in the labour share in value added (or in total income generated by the economy) (Karabarbounis and Neiman, 2013). **Evidence of rising income inequality lends further support to the sceptical view on the impact of recent technological advancement on labour markets.** That triggered the search for new economic models to explain these observations.

A recent task-based model of employment (Acemoglu and Restrepo, 2016, 2017, 2018a, 2018b) provides a better explanation for these phenomena. This model starts from the observation that the primary effect of automation is to replace workers with machines when the latter are more productive at a specific task. That displacement effect inevitably reduces the wage share of value added for that task and increases the share of capital or profits in value added. However, technological innovation may also lead to the creation of new tasks for workers that did not exist previously and cannot be done (yet) by machines. This reinstatement effect may increase human employment. The combination of displacement and reinstatement effects reallocates tasks between workers and machines. As such, displacement and reinstatement are in addition to the traditional effects in the substitution model.

The crucial difference with the traditional substitution effect is that the latter changes the demand

for workers and machines but without a reallocation of these production factors to specific tasks. Other effects need to be added to complete the picture. Machines may increase the productivity of workers for their remaining tasks and thereby push up wages and/or reduce product prices. Changes in relative prices across products will affect demand for products from different sectors and change the composition of the product basket in the economy.

Recent empirical evidence (Acemoglu and Restrepo, 2018c) based on USA data offers some support for this task-based model of labour markets effects of automation. The share of wages in GDP has been declining strongly in manufacturing, but not in services. The authors show that the traditional price-driven substitution and productivity effects are too weak to explain this decline. The main drivers of the decline have been the change in task content and, to a lesser extent, the sector composition effect. They decompose the task content effect in both a displacement and reinstatement effect. In the USA at least, displacement due to automation of existing tasks has been stronger than growth in new tasks.

Evidence for newly created jobs is based on new occupational titles in job statistics. It is not clear, however, whether new titles represent new tasks or only a re-bundling of existing ones. **There is no consensus on the definition of tasks that may vary across different levels of granularity.** It is not clear how distinct tasks are from other related tasks, how skills that can be domain-specific or domain-general affect the ability to perform a task, and how the automation/replacement of particular tasks affects the structure of individual occupations.

Gregory et al. (2018) apply a comparable task-based framework to EU data for the period 1999–2010. They confirm the existence of a strong employment-reducing substitution effect but also find that complementary demand and spillover effects more than compensate for this, so that the net employment effect of technology is positive. However, this finding depends on capital income gains from

technological progress feeding back into product demand. If only wage income gains feedback into demand, the total labour effect is only half as large. This underlines the importance of income redistribution policies (see Section 9.3).

Note that the historical period covered by the USA and EU studies dates back long before AI emerged as an automation technology. As such, it offers only a general view on the impact of technological innovation on employment, not a specific AI perspective.

Economists have tried to find more recent empirical evidence that is more closely related to the introduction of AI technology, in two ways. A first avenue was to try to be forward-looking and estimate the risk of automation of occupations. Some alarmist studies produced estimates of the percentage of human occupations at risk of being automated in the foreseeable future, varying from almost 50 % (Frey and Osborne, 2017) down to 13 or 9 % (Arntz et al., 2016; Nedelkoska and Quintini, 2018).⁴⁴ A major problem with these studies is that they focus only on the displacement effect and ignore all the positive employment effects that AI could induce – as explained in the task-based model by Acemoglu and Restrepo (2018c). Another problem is that the level of granularity of the definition of tasks creates high variability in the results. Occupational or even sub-occupational level approaches to skills are too coarse-grained to be reliable. Both problems lead to an overestimation of job displacement effects and excessively alarmist results.

The degree of routine in tasks is an important criterion to assess whether they can be replaced by AI (Brynjolfsson et al., 2018), although creative and social interaction tasks are not necessarily beyond the scope of today's ML algorithms provided that large quantities of data are available to train the algorithm. The organisation of work in a particular social context, rather than the task itself, may also be an obstacle to automation (Fernandez-Macias and Bisello, 2017).

A second research direction was to focus on the impact of industrial robots on employment and wages. Robots constitute a mechanically embodied form of AI that is more easily observable and measurable than algorithms. However, they are an imperfect proxy measure for AI because robots are rival products that can only be used for one task at the time while AI algorithms are non-rival and can be used for many tasks at the same time. In fact, neither the task-based model nor the pure labour displacement studies consider the implications of the non-rival nature of AI. Also, robots are mainly employed in manufacturing, especially car manufacturing, not in services. This results in selection bias in the evidence. Some evidence from the USA suggests a negative impact of robots on employment and wages. There is conflicting evidence for the EU where one study suggests a negative impact while a second study for Germany comes to the opposite conclusion (Chiacchio et al., 2018; Dauth et al., 2018).

Although the literature is not conclusive on the net employment effects of robotisation, there is ample evidence pointing to major reallocations and broad structural changes. The speed and efficiency of these changes can also influence the net employment effect: Korinek and Stiglitz (2017), for example, discuss the possibility of technological unemployment as a transition phenomenon. Such a case gets strong support from the recently emerged view among scholars of globalisation that the reallocation of workers among sectors is not frictionless, it can take many years, and the implied costs may be equivalent to multiple years of lost income for workers (see Hornok and Koren, 2017, for an overview). It is reasonable to expect that AI-induced reallocations would face similar challenges.

The speed of AI uptake is thus important not only for the expected productivity gains, but for the implied reallocation process, too. However, attempts to predict this speed are subject to wide uncertainty (McKinsey Global Institute, 2018).

A final, more philosophical point is to recognise that employment is not just paid market work. It also includes work at home or in communities, and may have an intrinsic value. As stated by Korinek and Stiglitz (2017), “jobs provide not only income but also other mental services such as meaning, dignity and fulfilment to humans”. To counteract the potential loss of paid jobs, policies may need to promote other fulfilling ways to spend time (Stevenson, 2017).

■ 9.2 Potential impact of AI on growth

We tend to look at the employment effects of AI from a worrisome angle – the risk of workers being displaced by machines – and forget about the promising effects of AI: AI machines can boost productivity compared to human labour and therefore generate more economic benefits for all. This would be very welcome in the current period where economies are suffering from slowing productivity growth.

Economists see human labour, (AI) machines, and innovative ideas as production factors in the growth equation (Aghion et al., 2017). These factors produce an output that is sold, the value of which is shared between labour and machines (or capital). Growth economists look at the rate of growth of that production as well as the distribution of the value of production between capital and labour. They use existing models of economic growth to simulate two types of future AI-driven scenarios.

The first type of scenario focuses on the production of goods and services, and on the substitution or complementarity between humans and machines in that production process. Even an extreme scenario whereby nearly all human tasks would be carried out by AI yields a reassuring outcome: the labour share in the value of output will remain high because remaining human tasks become very productive and highly remunerated. The ‘last’ human task will fetch an extremely high wage. That wage will have to be redistributed among unemployed workers in order to ensure some income for the latter.⁴⁵ Economic growth will continue in that scenario

at a steady pace, although an initial increase in the share of AI machines (capital) in the value of output may slow down growth.

The second type of scenario focuses on the impact of AI on the production of innovative ideas which, so far, has been the unique property of human labour. The invention of new ideas can be seen as a production process with a series of tasks, some of which can be automated and taken over by AI. Again, we can simulate an extreme scenario whereby all tasks for the production of new ideas are taken over by AI. This would lead to a ‘singularity’⁴⁶ scenario whereby economic growth would become infinitely high in finite time. However, even that scenario will run into constraints, including physical constraints to carry out an infinite number of machine tasks. Less extreme scenarios lead to a step-wise increase in the level of the economic growth rate but without becoming explosive.

These growth scenarios remain rather theoretical. In terms of empirical evidence, **they run into the same problems as employment forecasts: there are no data yet to gauge the impact.** In addition, available evidence for previous episodes of massive mechanisation might not fit the peculiar features of the AI revolution. Still, growth economists are trying to push forward the ‘time frontier’ in the empirical evidence, either by looking at past waves of innovation in other technologies or by examining the economic nature of AI as a new technology.

Cockburn et al. (2018) find preliminary evidence that **AI is a general-purpose technology⁴⁷ that fits a wide range of applications and spreads rapidly across many sectors and industries.** This could accelerate the productivity and growth effects. They also find that AI can, unlikely any previous technology, automate and accelerate the invention process itself, and thus the speed of innovation. This would ultimately translate into accelerated productivity growth. As an example, Agarwal et al. (2018) show how the use of AI in the pharmaceutical and chemical industries speeds up the process of discovery of new molecules for useful applications. These

'recombinant growth' scenarios show how AI can be better than humans at detecting promising new combinations of existing technologies or knowledge.

Other authors caution about the time it takes to roll out these optimistic forecasts. Brynjolfsson et al. (2017) argue that the Solow 'productivity paradox' - rapid growth in AI investments combined with slow increases in productivity benefits - applies to AI as well. **The realisation of the full effects of AI on productivity may take many years** because it requires the development of complementary innovations, organisational changes and new human labour skills. Brynjolfsson et al. (2018) find some evidence that the uptake of a general-purpose technology like AI follows a J-curve effect. The initial impact may be negative because heavy intangible investments in reorganisation and re-training drag down growth that only rebounds in later years. Historical evidence suggests that earlier general-purpose technologies followed a similar pattern. The fact that AI is a general-purpose technology would not be a reason to expect things to be different this time.

Besides its direct effects on the production of goods or ideas, **AI can also affect growth by changing firm behaviour and market competition.** If existing leading products or technologies are easily imitated by competitors in AI, it can either discourage potential innovators, or lead firms at the technology frontier to innovate more in order to escape fierce competition by the imitators. At any rate, easier imitation implies that AI-intensive production sectors might become congested in relatively little time. AI-driven platforms that handle massive amounts of data may adversely affect competition and lead to the emergence of a dominant platform - particularly if there are economies of scale, local externalities or rents. Algorithmic collusion may distort competition in some markets. This situation of monopoly, or at best oligopoly, may restrict access to the data collected by the platforms with non-trivial implications for trade and government regulations (Goldfarb and Trefler, 2017), and makes a case for antitrust considerations (Giles, 2018).

“The realisation of the full effects of AI on productivity may take many years.”

AI might also affect firm organisation along several dimensions. First, it might change endogenously the degree of complementarity/substitutability between workers with different skills - something that is considered fixed in the standard growth models. By facilitating the outsourcing of some low-skilled tasks, AI would at the same time strengthen the complementarity of those who are retained within the firm with the high-skill tasks that cannot be outsourced. By the same token, if AI reduces information asymmetries and facilitates monitoring, it can promote downstream delegation of power and more decentralisation of authority. The end result for AI intensive sectors would be larger and more horizontally integrated firms, with more profit centres, which outsource an increasing number of tasks to independent self-employed agents (Aghion et al., 2017).

Finally, similarly to what has happened with the IT revolution, **the effect of the diffusion of knowledge from AI might disproportionately benefit high-tech sectors** (Baslandze, 2016).

9.3 Potential impact of AI on inequality

The majority of the literature predicts an unfavourable impact of AI on the distribution of incomes. The most widely discussed and debated aspect is the increase in labour income inequality. Evidence suggests that labour market polarisation plays an important role in this. We observe polarising labour markets because tasks that are not easily performed by AI tend to be found at opposite ends of the skills spectrum while AI tends to replace humans in tasks that correspond to the 'mid-skill' category (Autor et al., 2003). Acemoglu and Autor (2011) and Autor and Salomons (2017) show suggestive evidence from the USA on how job polarisation translates into wage polarisation or even a polarisation in working conditions.

BOX 7. The impact of AI on working conditions

Because its use in actual work processes is still very limited, there is very little hard evidence yet on the impact of AI on working conditions. A recent Eurofound report based on the qualitative analysis of five emerging technologies in European manufacturing (including advanced robotics and the Internet of Things, which make intensive use of AI) concludes that there are some concerns in this sense, in particular with respect to the degree of autonomy, privacy and control of workers (Eurofound 2018). Digital factories where all objects are equipped with connected sensors and where workers collaborate with advanced robots can easily become digital panopticons, where human operators feel permanently monitored and controlled. **It seems difficult to maintain any sense of privacy at work if every object is a monitoring device.** And while management through algorithms and big data analytics can significantly improve the efficiency of the labour process, it can also lead to routinisation of tasks, work intensification and the asphyxiation of any sense of autonomy. On the other hand, **the introduction of advanced robots, connected devices and big data in the workplace can also contribute to a decline of repetitive and routine work**, as well as a reduction in the number of hazardous tasks. AI-enabled workplaces are likely to be safer, with more skilled workers carrying out less repetitive work.

There is evidence for Europe, too, that labour market polarisation leads to a widening of wage gaps. Goos et al. (2014) find that improved technology has led to increased demand for both well-paid, high-skilled as well as low-paid low-skilled jobs while the demand for middle-income jobs decreased, thus supporting the hypothesis that technological progress can lead to income inequality. Yet, there are substantially differing trends in labour market polarisation between European countries (Goos et al., 2014; Darvas and Wolff, 2016), which can be explained by country-specific institutions and policies (Fernández-Macías, 2012; Fernández-Macías and Hurley, 2016).

A report by the OECD (2018) shows a different type of polarisation that is regional and occurs within countries. For instance, the share of jobs at high risk of automation varies by 12 % between regions in Spain but only by 1 % between regions in Canada. Technological progress tends to perpetuate the developmental divide within countries as regions that are expected to be more negatively affected by technological progress also exhibit low productivity growth and high unemployment rates.

Despite similar positive trends in labour demand for both high- and low-paying jobs, we observe diverging trends in respective job quality. On the one hand, technological progress leads to increasing wages in high-paying jobs that require skills which complement AI (Deming, 2017). In contrast, technological progress causes even further reductions in wages at the lower end of the wage distribution down to a level that does not support a reasonable standard of living (Autor and Salomons, 2017). The accompanying digitalisation of the economy and the emergence of platforms causes an increase in precarious forms of self-employment that are characterised by a limited duration, such as seasonal or on-call work, as well as the absence of social security coverage (OECD, 2018). **This, too, contributes to increasing inequality even further.**

As discussed in a previous subsection, the speed of adoption of AI in the labour market could cause

transitional unemployment if AI makes workers redundant at a faster pace than they can move on to new jobs (Korinek and Stiglitz, 2017). **Differences in the pace of AI adoption across different regions and sectors, as well as differences between workers in their ability and speed to adjust to occupational changes may further increase inequality.** For instance, historical data suggests that low-skill workers are slower than high-skill ones in adjusting to sudden structural labour market changes and transitioning to new sectors and occupations. This results in longer period of transitional unemployment for low-skill workers (Goolsbee, 2018). The same has been established in the context of globalisation: Dix-Carneiro (2014) finds that female, less-educated and older workers face substantially higher costs of reallocation.

New evidence from the impact of robots on the manufacturing sector in Germany suggests that transitional unemployment effects may not be that strong as a large part of the workers manage the transition within their firms and across occupations. Yet, this job security comes at the cost of reduced wage growth for adjusting workers (Dauth et al., 2018).

Such significant reallocations can easily generate a sense of unfairness. Rodrik (2018) argues that so far this has been mainly associated with globalisation. Nevertheless, such sentiments may be evoked in the future with respect to automation and digital technologies, leading to a political backlash against technological progress, and AI in particular.

Job polarisation exerts its effect mostly on the dispersion of labour income. There are, however, additional channels through which the overall distribution of income (and hence inequality) may change. First, AI increases the share of capital income relative to that of human labour (Korinek and Stiglitz, 2017; Sachs, 2017). Together with the high degree of concentration of AI industries, it may lead to an increase in the inequality of capital income and also total income. Second, the extra wealth created by AI is likely be shared unequally across



BOX 8. AI and digital labour platforms

Digital labour platforms can be defined as digital networks that coordinate labour service transactions in an algorithmic way (Pesole et al., 2018; Gomez et al., 2017). Although they are still marginal in terms of employment, they represent a radical departure from traditional forms of work and thus do not fit well into existing categories or regulations. Their use of digital technologies, such as geolocation, big data analytics and algorithmic management, allow them to break the traditional boundaries between markets and firms, performing functions of both mediation and management for the transactions they coordinate. They also imply a radical change in the nature of work, because their coordination efficiency facilitates the direct provision of labour services in small discrete tasks, breaking up the traditional bundling of tasks into jobs and occupations (Fernández-Macías 2018).

What is the role of AI in digital labour platforms? Since they rely on big data analytics and complex algorithmic rules, **digital labour platforms already make intensive use of AI techniques.** But the potential for advanced AI techniques to continue expanding the coordination efficiency of digital labour platforms is big, and will be crucial for its further development.

Digital labour platforms illustrate the potential effect of AI on employment which goes beyond the direct replacement or augmentation of existing labour input. **The use of AI for the management and coordination of labour has the potential to change the organisation of work significantly,** alter the very nature of employment contracts and further develop the division of labour and task specialisation.

countries. Winner countries will have higher income levels, and more room for domestic redistributive policies (Lee, 2017). Third, as time evolves, AI may enable improved health and skills, leading to forms of human enhancement. If not limited by policies, the ability to ‘purchase’ skills and health will further widen pre-existing inequalities (Harari, 2015; Korinek and Stiglitz, 2017).

It is commonly agreed that policy measures are needed to counteract the negative effects of technological progress on equality. Besides increasing access to high-paying jobs raising the overall skill level through increased education expenditures, policies should further ensure a reasonable standard of living. The literature discusses several types of policies to achieve this goal, such as universal basic income or guaranteed employment (Furman and Seamans, 2018), policies that aim at a redistribution from ‘winners’ to ‘losers’ and policies that shift the taxation of human labour towards the taxation of capital (Korinek and Stiglitz, 2017). In any case, the within-country regional variation of the impact of technological progress suggests that policies should be adjustable to local needs, where local offices can help in the design of targeted policies. It is important to be aware that such policies may lead to inefficiencies (equity-efficiency trade-off), reducing the size of the pie to be distributed. Further empirical analysis is necessary to assess the effectiveness of these policies.

Despite these negative prospects of technological progress on equality, there might be some benefits from the peculiar features of AI. As a general-purpose technology AI could yield equality enhancing effects. For instance, according to Acemoglu and Restrepo (2016), AI can create new tasks which can be performed by high-skilled workers in the short run. Yet, as these tasks become standardised in the long run, low-skilled workers can also benefit from them. In addition, AI shows potential to disrupt the spiral of labour market polarisation. It may be able to perform high-skill tasks that were previously beyond the abilities of technology, such as the classification of case documents for lawyers

or the reading of medical images. Even creative and social tasks are not out of the realms of AI abilities (Brynjolfsson and Mitchell, 2017). In the end, AI may produce depolarising effects.

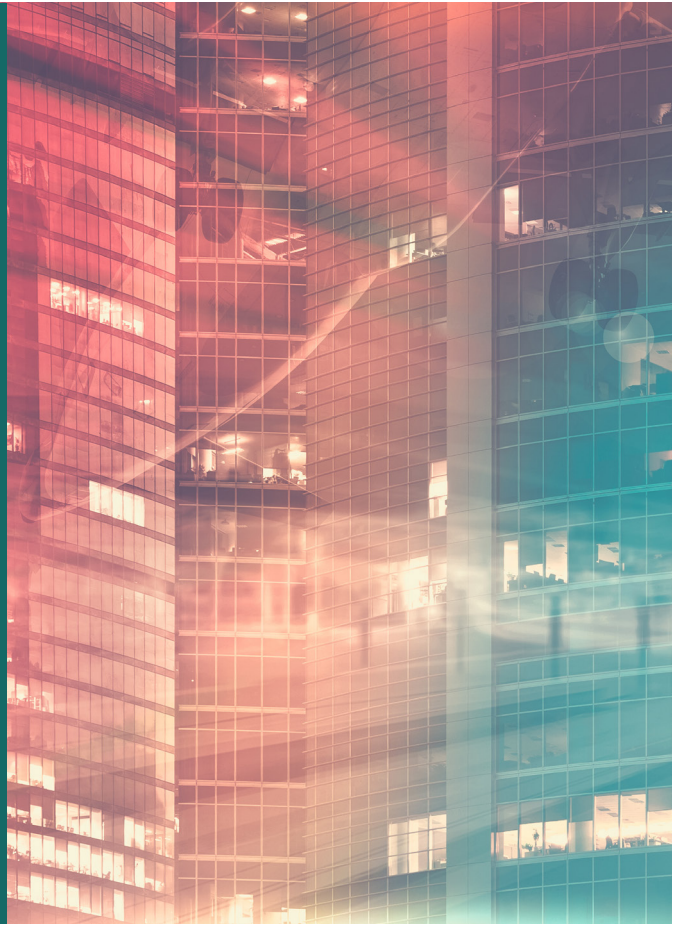
9.4 Summary and conclusions

To date, the theoretical and empirical economic research literature is ambiguous on the growth, employment and wage effects of AI. It can be negative if machines only substitute human labour, but positive if machines instead complement human workers and increase overall productivity. Empirical studies use historical data that go back long before the rise of AI. They can tell us something about technological change in general but little about AI specifically. A recent USA study finds a negative impact of technology on jobs and wages. An EU study finds a positive impact, also because capital gains feedback into overall demand. Studies that use robotics data as a (rival) proxy for (non-rival) AI investments generate inconclusive evidence about the impact of AI. Some forward-looking studies estimate the number of jobs at high risk of AI-driven automation. They produce excessively alarmist results because they only look at substitution and omit complementarity.

Focusing only on employment effects ignores the potential of AI for economic growth which could generate benefits for all. AI as a general-purpose technology can rapidly spread across industry sectors and yield strong positive growth effects. To the extent that ML generates new insights from data, it may contribute to the automation of production of new ideas and innovation itself. The ‘productivity paradox’ becomes more apparent as the rapid growth in ML applications does not seem to be reflected yet in the productivity growth statistics.

The macroeconomic outcomes of AI are likely to be shaped by firm behaviour. In this respect, the issues of market structure, organisation and sectoral reallocation seem particularly relevant. In perspective, policy interventions might be needed precisely in these areas in order to: i) ensure a competitive

“AI as a general-purpose technology can rapidly spread across industry sectors *and yield strong positive growth effects.*”



stance of the market that incentivises innovation; ii) avoid that economies of scale in data handling lead to excessive market concentration, and in general other competitive advantages stemming from differences in national regulations; and iii) ensure that sectoral reallocation takes place with limited social costs, e.g. from job market displacement.

AI can unfavourably affect the distribution of income through many channels. The most discussed are job polarisation, regional differences in AI adoption, or regional, sectoral or individual differences in the speed of adjustment to structural labour market changes. Preliminary empirical evidence seems to confirm these concerns. There is also a strong case for potential inefficiencies in the implied labour reallocation process, which are likely to hurt specific groups disproportionately. It is commonly agreed that policy measures are necessary to counteract the negative effects of technological progress on equality. Strong within-country variation in technological progress

suggests that policies should also be adjustable to local needs.

Whether or not policy will be effective in mitigating the potential adverse effects of automation is an empirical question. **The nature of AI as a general-purpose technology offers some room for hope that the effect of AI on equality may not be all negative.** It can be positive if AI boosts the creation of new tasks for low-skill workers and expands its abilities to perform high-skill tasks.

Given these uncertainties, there is a need for a strong monitoring framework to also detect at regional and subregional levels the potential impacts of AI deployment and to quickly put in place appropriate intervention. We return to this important issue in Chapter 13 on resilience.



Cybersecurity operations will benefit from AI, but also face new challenges from it.

SUMMARY

In this chapter, we reflect on the transformations and impacts brought by the advent of AI to cybersecurity and, conversely, how AI algorithms themselves are affected by security concerns. When it comes to cybersecurity, it is crucial to be aware of the fundamental dual nature of AI. It will empower human operators, professionals and investigators and it will bring benefits to cybersecurity. However, those who seek to attack systems and networks are equally expected to take advantages of AI to carry out faster and smarter cyber-attacks.

Moreover, AI systems based on ML techniques are often not robust enough against malicious attacks, introducing a range of new and poorly understood vulnerabilities. This could pose a danger for the security of the systems using them. To review all these developments in context, we describe a range of recent AI applications in cybersecurity, outline the specific vulnerabilities of AI systems and related attacks, and suggest ways of increasing AI robustness and introduce security and safety by design principles.

CYBERSECURITY PERSPECTIVE

■ 10.1 Background: AI and cybersecurity

Cybersecurity is the field broadly concerned with the security of digital information, systems and services, including the security of cyber-physical⁴⁸ systems. It encompasses notions such as *resilience*, *deterrence* against malicious actors or cybercrime and *defence* (see Chapter 13 for a societal perspective on resilience). In a recent joint Communication to describe the technological and policy dimensions of cybersecurity, the European Union recognised its emerging importance for European societies (EC, 2017b).

The application of partly autonomous algorithms in cybersecurity is not entirely new, although traditionally those systems were usually not referred to as ‘Artificial Intelligence’. Cybersecurity controls capable of functioning autonomously and taking intelligent decisions to protect information systems and services have existed for quite some time. They started to play a role many years ago with the development of smart security applications, for instance for deciding whether or not to allow a certain network communication, to autonomously filter spam messages, or adapt to new circumstances such as the identification of previously unseen forms of cyber-attacks.

Since then, the field of cybersecurity has undergone rapid transformation due to the developments in ML, deep, and reinforcement learning, which have resulted in notable successes in addressing computer vision, NLP and autonomous decision-making (see e.g. Goodfellow et al., 2016; Brundage et al., 2018; Osoba et al., 2017; Mission Villani, 2018). It is likely that these transformations will continue and

AI will bring benefits to cybersecurity and the fight against crime, but will also introduce new challenges in the field.

amplify the importance of cybersecurity even further in the coming years.

The main lines of development we can expect in the coming years include:

- *New capabilities* enabling cybersecurity to do things that were not possible until recently, including those increasing the resilience of systems and services (see Section 10.2 for more details on some recent developments) as well as investigative tools and capabilities to assist in the deterrence of crime, e.g. aiding law enforcement in the prosecution of criminals.
- The integration of AI into widespread digital products and applications, and into the controls of cybersecurity itself, may introduce new, and especially *conceptually different and not yet fully understood*, vulnerabilities that can be exploited by malicious actors (see Section 10.4 for an overview). These

vulnerabilities, given the crucial importance of data for AI systems, also raise the importance of data protection.

- Challenges to private and public security from the possible abuse of AI systems to enhance cyber-attacks and malicious use. Important examples include increased and more extended cyber-attacks, more difficulties to attribute attacks to a specific actor, the targeting of human vulnerabilities through autonomous social engineering, social media and propaganda manipulation, the production of credible fake content in news or reporting, attacks on cyber-physical systems such as autonomous vehicles, or the development of autonomous weapon systems (see, e.g. Brundage 2018).

■ 10.2 Applications of AI in cybersecurity

Early examples of AI systems applied in cybersecurity include systems for network intrusion detection to identify previously unknown cyber-attacks (Paxon, 1998) and antivirus approaches to detect new malware (Sanok, 2005). The explosion in the practical use of ML in the early 2000s brought further tangible benefits to the cybersecurity domain. The most striking examples were probably the first robust adaptive systems to detect and filter spam and malicious messages autonomously (see also Box 1 in Chapter 2).

Most AI-based methods currently in use in operational controls are based on ML approaches, using a host of well-tested algorithms and expert schemes and are usually embedded within a traditional cybersecurity architecture (for a recent example see IEEE 2017). They are largely *supervised learning* methods, i.e. AI systems that evolve their models using available labelled data as examples. Typically, they are performing classification of messages, digital events or computer code into acceptable or non-acceptable classes. These AI-based controls are capable of limited forms of adaptation, mostly through retraining on new samples.

With recent successes in DL, we are starting to see a number of AI methods based on neural networks and DL gradually moving from research into workable cybersecurity applications. They employ deep neural network architectures and big data to devise increasingly accurate cybersecurity learning systems, e.g. for anomaly detection in malware and computer network analysis, message filtering or in assisting human security operators (e.g. Crawford et al., 2015; Radford 2018; Meidan 2018). In all likelihood, these trends will continue, but will also come hand in hand with a proliferation of AI-backed cyber-attacks (Brundage, 2018).

■ 10.3 Deterrence and fight against crime

Recent developments strengthen the resilience of cybersecurity systems, reduce risks and increase their capability to withstand threats and malicious intent reliably. Another element is that research into AI in cybersecurity systems plays a crucial part in developing cyber deterrence capabilities. Firstly, AI systems are being and will increasingly be used by cyber criminals (e.g. Brundage et al., 2018), and thus research into their capabilities and weaknesses (see Section 10.4) will play a crucial part in defending against such malicious usage of AI. Secondly, of course, law enforcement will increasingly engage in active usage of AI systems to reinforce investigative capabilities and to strengthen digital evidence-making in court. Most developments aim at leveraging new tools for three main operative goals, the localisation of a crime, the identification of victims and perpetrators, and the determination of the content of the crime.

The most prevalent examples can be found in digital forensics and in the employment of biometric systems for investigative purposes. The wide field of digital forensics encompasses very different applications, ranging from computer and hard-drive forensics to fraud detection in large databases. The spread of online services and digital devices has led to a proliferation of digital content relevant to the prosecution of criminal activities. In this context, AI systems, for instance for image, audio and

text analysis, or for unsupervised analysis of un-ordered data volumes, have the potential to assist law enforcement, not only in the investigation of cyber-dependent and cyber-enabled crime, but also in traditional forms of crime where digital content can play a key role in the investigation.

10.4 Robustness of AI algorithms against malicious action

With an increasing number of AI systems employed in cybersecurity, not only are we making progress but we are also introducing new vulnerabilities that then open the window for new types of attacks. For this reason, these AI systems are now coming under increased scrutiny.

In classical software development, the developer uses a computer language to create the program that implements a certain algorithm. This constitutes a straight development process where we can

use secure design principles, have a clear view of where vulnerabilities are likely to occur, and put in place effective security controls. In contrast, AI systems are fundamentally more complex, especially if they are based on ML architecture. They are composed of a possibly non-linear feedback loop system between a human-designed algorithm and a human (or even AI) assembled dataset. Together, these constitute a learning model that acts as the actual 'AI' and is basically a new algorithm. These AI models are then embedded into a program, often combined with other pieces of code or software and possibly implemented with a host of different programming tools. The situation is depicted in Figure 12.

Any vulnerability or design flaw in this complex system will likely be of a very different nature than in classical systems. Tracing an event and its cause, such as an attack on the system and the vulnerability exploited becomes a much more complex

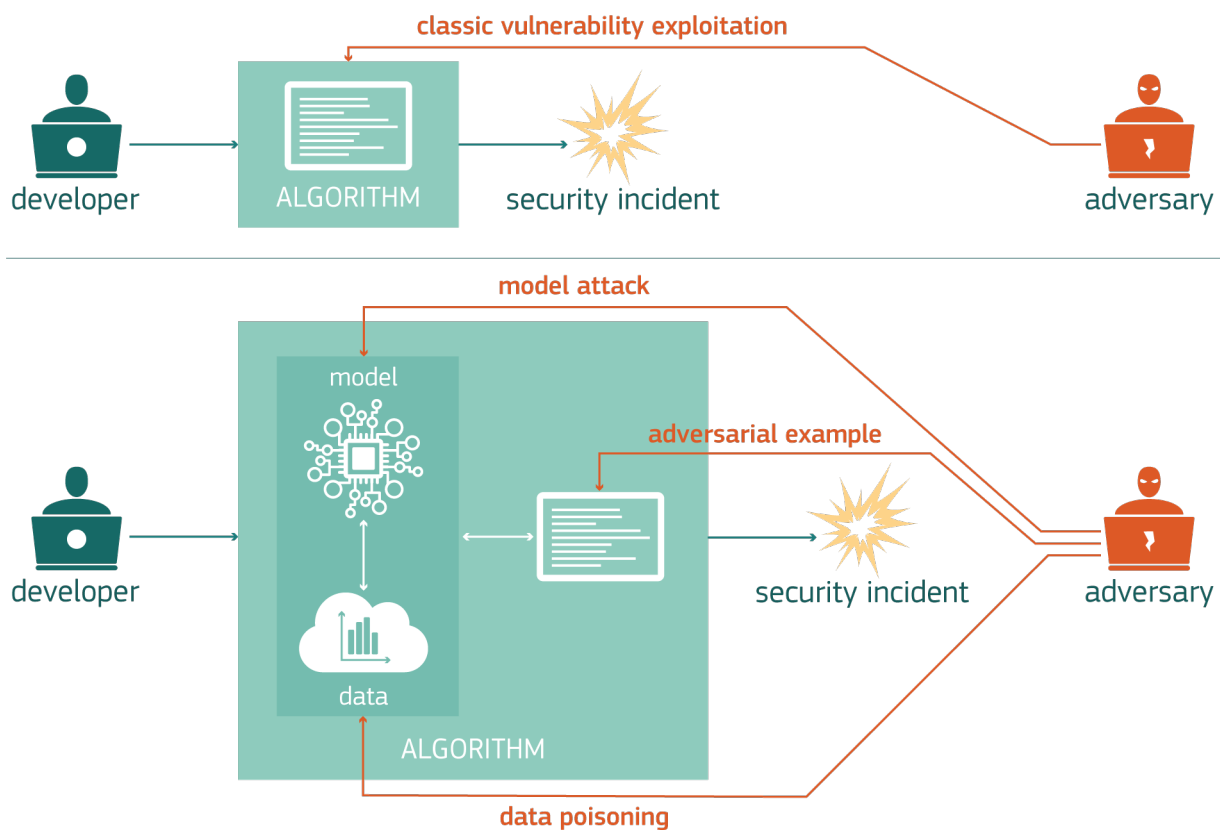


Figure 12. Paradigm shift in the cybersecurity of systems following the introduction of AI components

and demanding task (see Fig. 12). Moreover, once the AI system is embedded into parts of a classical system and different combinations of old and new types of vulnerability come together we may end up with new obstacles altogether. A different angle from which to see the complexity problems displayed in Figure 12 is that of algorithmic explainability, or interpretability: to what extent are the results of a given algorithm still understandable to a human or are even uniquely explainable (e.g. because of non-linear functions employed in many ML models)?

Explainability is a key area of fundamental research because we need explainable and interpretable models to increase the capabilities of AI systems to abstract and transfer knowledge, understand their robustness in terms of cybersecurity, assess possible bias, and crucially, build and retain trust in society (see Chapters 2 and 6 for a discussion on fairness and accountability in the context of AI).

The emerging fields of AI safety and adversarial ML are actively concerned with these new challenges. The former refers to the research branch concerned with designing safe AI systems from the start (see, e.g. Amodei, 2016). The latter is concerned with the new types of vulnerabilities exhibited by the most prevalent AI systems employing ML (see, e.g. Huang et al., 2011; Szegedy et al., 2013). Typical vulnerabilities, exclusive to AI systems, include the following (see, e.g. Biggio et al., 2012; Szegedy et al., 2013; Huang et al., 2017):

- *Data poisoning*, the deliberate introduction of false data in the training of the model, especially dangerous for autonomous AI systems that employ online re-learning procedures. This effectively means that the data itself becomes a major vulnerability of such systems. In Figure 12, this type of vulnerability would conceptually be introduced in the 'data' part of the AI system.
- *Adversarial examples*, input in the trained ML system, which is deliberately designed

to be misclassified or to fool the algorithmic procedure in which the model is being used. In Figure 12, this type of vulnerability would conceptually be introduced in the right part of the AI system, after the 'data' and the 'model' have interacted.

- *Design flaws* in the underlying mathematical procedures that optimise the learning process of the model, leading to model attacks. For instance, the usage of a loss function or likelihood model that is known to be susceptible to certain types of input could create an easy path for adversarial examples. Moreover, insecure systems might allow a malicious user to steal the mathematical structure of a model. In Figure 12, this type of vulnerability would conceptually be introduced in the 'model' part of the AI system.

There is growing evidence of how easily these vulnerabilities can be exploited in real applications, especially in ML-based models. Thus, designing AI-specific security controls is an active, albeit young, field of research. Research on adversarial ML currently focusses on two main measures depending on circumstances, such as the intention of the attack and the type of vulnerability (e.g. Papernot et al., 2015):

- *Data sanitisation*: cleaning the training data of all potentially malicious content before training the model. In most realistic cases, another AI system is employed to act as a filter, but in very critical circumstances human intervention might be inevitable.
- *Robust learning*: redesigning the learning procedure to be robust against malicious action, especially adversarial examples. This either entails explicit training against known adversarial examples, or redesigning the mathematical foundation of the algorithms and employing techniques from statistics, known as regularisation and robust inference. This approach actually follows the



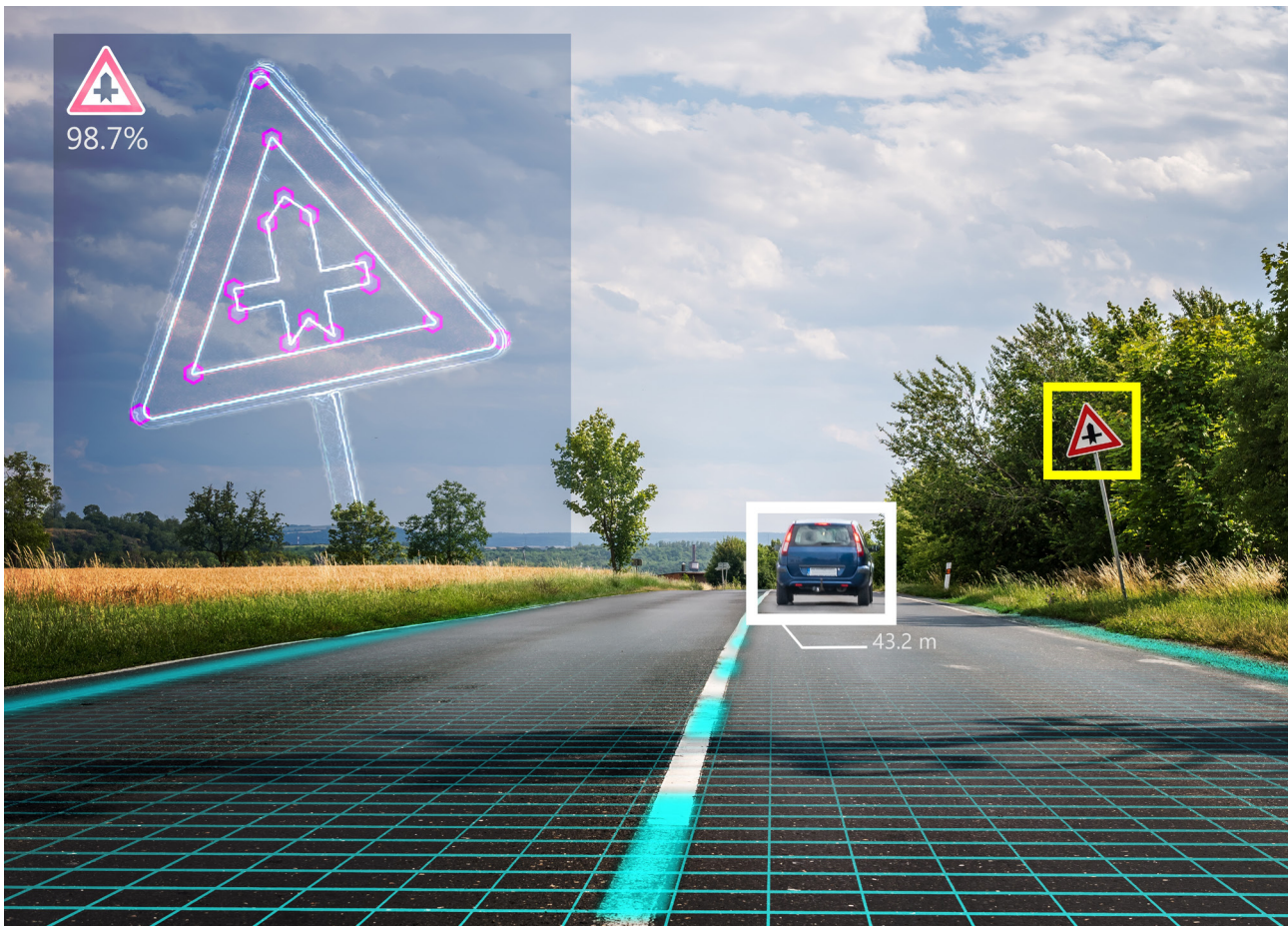
“ AI will not replace humans in cybersecurity anytime soon but rather empower them. *It is important to understand the limits of the new generation of AI algorithms.* ”

well-known security-by-design principle, i.e. taking the security of a software or application into account from the beginning of the design process. It should be noted that for ML based AI systems this approach usually implies a degradation of performance.

■ 10.5 Summary and conclusions

The advent of the AI revolution is expected to have an increasing impact on cybersecurity. The existing landscape will be subject to many transformations in terms of more effective and efficient cyber defences, but at the same time more sophisticated cyber-attacks. Key conclusions from these developments are:

AI will not replace humans in cybersecurity anytime soon but rather empower them. Expectations about AI are high, and for good reasons, given the latest advances in the field and the growing number of successful cases of its application in several domains, including cybersecurity. However, it is important to understand the limits of the new generation of AI algorithms.



We have summarised the key issues above, and we clearly still need human operators to complement the capabilities of AI for cybersecurity.

AI will bring clear benefits to cybersecurity and the fight against crime, but will also introduce new cybersecurity challenges. ML algorithms have altered the classical paradigm of software development. When using them, human developers no longer program an algorithm to tell the computer *how* to solve a given problem but instead they program it to *learn how* to solve the problem. As a result, the AI system behaves as a new algorithm, not directly programmed by the human developer: this is the result of a learning process from training data carried out by the ML algorithm originally developed by the human. Recently, it has been found that, despite their effectiveness, **AI systems based on ML techniques are often not robust against malicious attacks** that exploit the specific way in which the AI system behaves by presenting it with

specially crafted input, manipulating the data that is used to train it, or abusing biases in the statistical algorithms used to perform the training. These are new classes of vulnerabilities, specific to ML algorithms, which need to be further understood and for which specific organisational and technical controls need to be designed and implemented.

Secure software development practices, security certifications, security audits and cybersecurity controls need to consider these new kinds of vulnerabilities specific to the AI components of the digital systems. Further research is needed in the emerging field of adversarial ML in order to better understand the limitations in the robustness of ML algorithms against malicious action and design effective strategies to mitigate these vulnerabilities. This is of particular relevance when AI is used in cybersecurity controls or to assist human actors in the identification, localisation and prosecution of cybercrime.

There is a need to secure the availability of up-to-date high-quality datasets in the domain of cybersecurity. The latest techniques in ML such as DL or reinforced learning can produce impressive results but require huge quantities of training data to work and be effective. As is the case in other domains, there is a scarcity of large enough high-quality datasets in the field of cybersecurity. The availability of such datasets to the scientific community is crucial to ensure future advances in the field.

Moreover, the creation of up-to-date reference datasets in the several areas of the cybersecurity domain can also boost research and innovation by enabling agile and effective benchmarking and evaluation of state of the art AI techniques.

The dual nature of AI. **It is crucial to be aware of the fundamental dual nature of techniques, tools and systems employed in cybersecurity and AI,** especially when both have to be considered together. AI can improve security, but cybercriminals are equally expected to take advantage of the latest advances in AI techniques to carry out faster and smarter cyber-attacks.

This situation does not necessarily require restricting research. An effective defence can only be designed knowing how attacks might take place. Therefore, it is important to ensure that further

public research is conducted in the field of AI and cybersecurity, also considering the offensive side, to study emerging and potentially AI-powered cyber-attacks and devise corresponding countermeasures.

Furthermore, the research landscape on AI is already highly dominated by the USA and China (see Chapter 3) and the ongoing general drain of research and researchers from Europe needs to be reversed, especially in fields already lacking knowledgeable experts in the market, such as cybersecurity and AI. Focusing research on the ethical use of AI and the role of regulation can become an asset to strengthen European leadership in these areas, and retain expertise in AI and cybersecurity.

AI in cybersecurity requires a multidisciplinary approach. On the highly technical matters that surround the application of AI in cybersecurity, strong interactions of policymakers with researchers and engineers seem immensely important. The fields of AI and cybersecurity are rapidly changing, and reports will quickly become outdated, making a constant information flow from the technical level to decision bodies almost mandatory. As an extension of this, we also need to foster a culture of cross-collaboration between AI and cybersecurity developers and, on a broader level, researchers from other related fields such as philosophy, economics or social science.





Intensive computing needs to be energy efficient

SUMMARY

In this chapter, we analyse one of the often-ignored aspects of AI, i.e. that the opportunities created by great data volumes and advanced analytics come at an increasing environmental, economic, and geopolitical cost: energy. The introduction of the IoT and 5G networks will increase the problem significantly. Energy is recognised as a main determinant of the long-term sustainability of AI. Therefore, we need to take early action to address this aspect of the development of AI and its related technologies. While some actions can be done using today's technology, more power-reduction innovation is needed to fulfil future demand. It is necessary to introduce advanced technologies in the hardware, software, and services areas, to develop intensive-computing ecosystems that are 'green'.

These ecosystem solutions must face four important energy-consumption areas: reduce the energy consumed by the existing big data centres; limit data movement reducing the energy consumed from machine-to-machine communications; reduce the embodied energy; reduce the obsolescence of digital technologies (see, for example, the short lifespan of mobile phones, replaced every one to three years), which is perhaps the main energy consumption voice for digital technology.

COMPUTER PROCESSING AND ENERGY PERSPECTIVE

■ 11.1 Introduction

There are several studies (SIA, 2015; Mayer 2009; IEE, 2012) showing that the digital economy is facing an increasing problem: a severe mismatch between the processing and storage needs of the escalating volumes of data available, and a sustainable energy footprint. A report prepared for Greenpeace (2012) claimed that if the cloud were a country, it would have the fifth largest energy demand in the world, while Vidal (2017) suggested that the data tsunami could consume one fifth of global electricity by 2025.

These estimates maybe somewhat off the mark, but there seems to be a consensus that **data growth will largely outpace foreseeable improvements in computational power and the associated energy footprint**. For example, between 2002 and 2009, data traffic grew 56-fold, compared to a corresponding 16-fold increase in computing power (largely tracking Moore's law) as shown in Figure 13 (Short et al., 2011; and Kambatla et al., 2014). The development of the IoT is expected to widen the fork between data and processing power in the coming years with an increasing energy bill. Therefore, the computing and energy consumption generated by AI and IoT (including devices, networks, and data centres to which they are connected)

There is a paradigm shift in computing to bring processing at the level of smart sensors, opening an opportunity for European investment, regulatory frameworks, and standards to shape these developments.

must be optimised significantly. In this chapter, we explore three main approaches to reduce the energy footprint after reporting in more detail the energy challenges of one important part of the data chain: data centres.

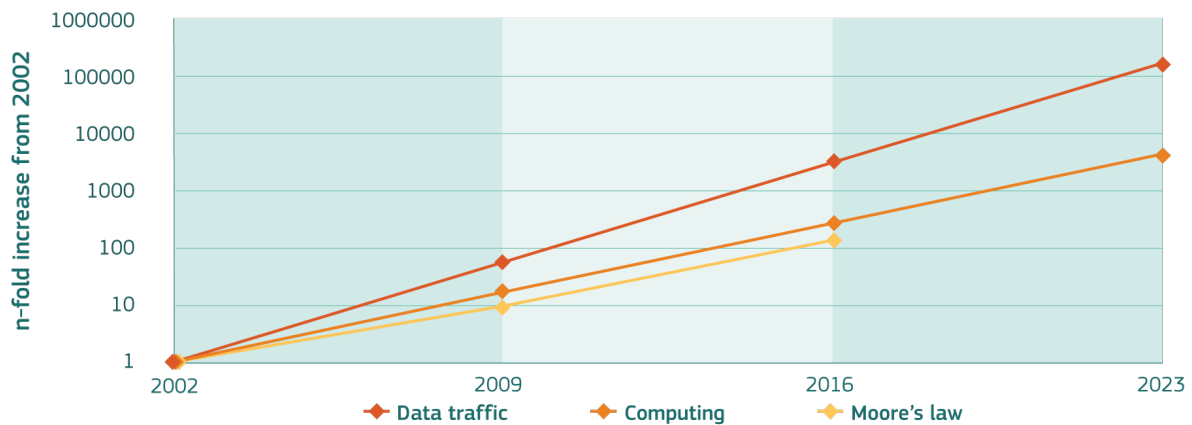


Figure 13. Projection of data and computing growths (logarithmic scale)

Source: JRC based on Kambatla et al., 2014

11.2 Assessment of data centre (DC) energy consumption

AI requires a considerable amount of data and processing power, which often implies the need for dedicated facilities such as data centres (DC). DCs host IT equipment for data storage and processing (e-mail, financial transactions, social media, computing, etc.) including computers (servers) and other components, such as communication and storage equipment. Normally the DC is hosted in a controlled environment requiring air cooling and conditioning, in order to extract the heat generated in processors. In addition, power has to be distributed to the IT equipment, including conversion from AC to DC, battery storage and uninterruptible power supplies (UPS) to cope with very short power disruptions. There is a high level of redundancy in DCs for both power distribution and cooling to increase power availability and reliability, including provision for standby generation (usually diesel-powered). These redundancies and the need for a controlled environment are typically doubling the power and energy consumption of the IT part (i.e. twice the IT power is required by a typical data centre).⁴⁹ In extremely efficient DCs this additional consumption is only 10 %.

The global energy consumption of DC was estimated at 194 TWh in 2014⁵⁰, which is around 1 %

of annual global electricity consumption, i.e. more than the electricity consumption of several EU Member States.

DC energy consumption in the USA (largest market in the world) has been estimated at 70 TWh, representing, about 1.8 % of total US electricity consumption.⁵¹ In Europe, the number of DC has dramatically increased in the last 10 years, particularly with the development of cloud computing. DC energy consumption in the EU (second or third market in the world) was estimated in 2014-2015 to be at between 43 and 49 TWh i.e. between 1.5 % and 1.8 % of EU total consumption. This is a calculation by the JRC based on an assessment of national experts and published reports.

These estimates do not include the energy consumption to transfer the data from one DC to another or to the end-users. This is part of the broadband operators' consumption (e.g. telecoms, over fibre, DSL, cable or mobile-4G) and end-users (e.g. modems, routers, switches at home and offices). The data transmission energy consumption is comparable to the total DC energy consumption. **Altogether, therefore, DCs and data transmission could account for 3-4 % of all EU power consumption.**

DC consolidation, outsourcing and cloud computing are helping to keep energy consumption in DCs flat,

notwithstanding the increase of data and processing, as larger DCs tend to be more efficiently designed and managed.

11.3 Options to improve the energy efficiency of the increasing demand for HPC

There are different approaches to address this problem:

- (a) advance CPU capabilities to process big data, run AI/ML jobs, and consume less energy;
- (b) improve the energy consumption of DCs necessary to cool the systems;
- (c) move the intelligence from the centre to the edge by optimising (cloud-based) computing infrastructures and bring AI closer to the data.

11.3.1 CPU advancements: energy-saving and parallelisation computing

With respect to CPU advancement, there are several lines of action:

- *Graphics Processing Unit (GPU)*: a specialised electronic circuit designed to accelerate the creation of images intended for output to a display device. GPUs are used in mobile phones, personal computers, workstations, and game consoles. Modern GPUs are very efficient at manipulating computer graphics and image processing, and their highly parallel structure makes them more efficient than general-purpose Central Processing Units (CPUs) for algorithms where the processing of large blocks of data is done in parallel, as for ML and DL.
- *AI accelerators, the Tensor Processing Unit (TPU)*: an AI accelerator developed by Google specifically for neural network ML. The chip has been designed to use Google's TensorFlow framework, an open math library which is used for ML applications such as neural networks. Other AI accelerator designs are starting to appear from other vendors aimed, for example, at the robotics markets.
- *Advanced cloud (multi-core) chips* that combine the advantages of CPUs with GP-GPUs, and specialised AI chip. This new breed of chips promises to reduce ten-fold the times the processing power per watt (e.g. through a fine-grained power management) to be capable of running the most complex computing tasks.
- *Embedded processors* for smart sensors by using new microcontrollers designed for AI, which are the brains of billions of 'smart things'. Trained DL models are compiled to generate optimised neural network code that is embedded in advanced microcontrollers. Embedded processors are usually simpler in design and have minimal power requirements. In addition, moving the computation from the centre to the edge of the network contributes to reducing the energy footprint of the IoT.
- *Neuromorphic computing*: a technology based on the use of very-large-scale integration systems that mimic neuro-biological structures present in the nervous system. Recent solutions focus on models of neural systems for perception, motor control, or multisensory integration.
- *Reversible computing*: a model of computing that tries to minimise the amount of energy dissipated as heat during processing. They are based on low power circuits, known as 'adiabatic' circuits, that are designed to conserve energy.
- *Quantum computing*: a form of computing exploiting some of the physical properties of quantum mechanics. Classic computers

store information in bits that can either be 0 or 1. Qubits, which are the quantum equivalent of transistors, are capable of ‘superposition’, i.e. they can be both 0 and 1 at the same time, and are therefore able to store a lot more information. ‘Entanglement’ is the other key property, which means that qubits are connected to each other, so that operating on one affects those that are connected. This property allows for very fast parallel processing. Quantum computing is already used for cryptography, but other applications are starting to emerge.

11.3.2 Innovative and more efficient cooling systems/engineering solutions

Efficient cooling: the first solution is to raise the DC temperature so that there is no need to over-cool servers and other equipment. Another is to use fresh outdoor air (free cooling, direct or indirect) or cold water provided by lakes, rivers, the sea, or underground aquifers (the final cooling to the racks is still provided by air through a heat exchanger). Finally, ‘adiabatic’ or evaporative cooling can also be a good alternative to compressor-based cooling. There are an increasing number of DCs with no-compressor-based cooling system, running all the time on free cooling mode, which is highly efficient. For high-performance computing, the air cooling is limited because of the poor conductivity of air and the very large power densities and heat generated by the processors. The use of water cooling can dramatically reduce the cooling energy consumption.

Other solutions: removing old servers and installing newer and more efficient ones (e.g. with Energy Star), installing more efficient software, and more efficient power distribution, e.g. batteries on servers. Efficiency has also improved thanks to the EU Code of Conduct on Energy Efficiency in Data Centres⁵², adopted voluntarily by more than 350 DCs in Europe.

“The expected growth in data, computing processing, and energy consumption is changing the present paradigm.”

11.3.3 Innovative infrastructures, architectures and configurations

AI and IoT are going to change the nature of the internet as we know it. **The expected growth in data, computing processing, and energy consumption are changing the present paradigm** ‘do everything in the cloud’ to a new one that can be called ‘bring computing to where the data are’.

Edge and fog computing: many use the terms fog computing and edge computing interchangeably, as both involve bringing intelligence and processing closer to where the data are created. The key difference between the two is where the intelligence and computing power is placed.

- *Edge computing:* a methodology for optimising cloud computing systems by **performing data processing at the edge of the network, near the source of the data**. For



example, performing more computation at the level of the sensors capturing the data, or mobile devices like mobile phones. In this way, there is less need to transfer data to centralised servers or clouds, and only the result of the processing is then transferred, reducing the data traffic considerably. This approach can also leverage resources that are not continuously connected to a network, such as laptops, smart phones, tablets and sensors.

- *Fog computing*: implements a decentralised computing infrastructure in which data, computing, storage and applications are distributed in the most logical, efficient place between the data source and the cloud. Fog computing essentially extends cloud computing and services to the edge of the network, bringing the advantages and power of the cloud closer to where data is created

and acted upon. In other words, fog computing acts as midway in the processing chain between the edge (e.g. sensors, mobile phones) and the cloud.

Fog and edge computing architectures have been introduced recently recognising the potentially massive increase in data traffic to be expected with the adoption of the IoT, where billions of devices used on a daily basis will be connected to each other using new-generation wireless technology and embedded AI. The benefits of decentralisation include: scalability (less data moved through the internet), latency (AI closer to data sources), data security (local scope), and cost. Fog and edge computing are different from traditional cloud computing and, hence, they will require different solutions.

They introduce a different cloudification (business) level dealing with more local and less powerful clouds, which are closer to the smart edge devices and to the users – they are called edge clouds or cloudlets. This type of infrastructure, which will be crucial for the future development of IoT and AI in Europe, is only emerging now. The approach and frameworks commonly used for data processing and AI in the cloud are not good for the cloudlets. We have therefore a (narrow) window of opportunity to make sure European industry takes a lead in developing this new infrastructure and collectively (European Commission, Member States, and industry) sets the rules of the game, including:

- A European open reference architecture for fog and edge computing, complementing the traditional cloud (DCs) approach.
- Reinforcing research on transportable DCs, to support cloudlets and edge architectures.
- Respecting European values (transparency, privacy, and ethics) at the fog computing level – i.e. fog nodes and cloudlets.

- As to cloudlets' energy consumption, small amounts of energy are consumed in a plurality of locations. Often the organisation paying for the energy may not be the same one running the infrastructure or the service. Therefore, energy efficiency policies need to consider new business models and possibly new regulations.

11.4 Summary and conclusion

In this chapter, we have analysed one of the often-ignored aspects of AI, i.e. that **the opportunities created by great data volumes and advanced analytics come at an increasing environmental, economic, and geopolitical cost: energy**. The introduction of the IoT, and 5G will increase the problem significantly. Energy is recognised as a main determinant of the long-term sustainability of AI. We need therefore to take early action to address this aspect of the development of AI and its related technologies. While some actions can be taken using today's technology, more power-reduction innovation is needed to fulfil future demand. It is necessary to introduce advanced technologies in the hardware, software, and services areas, for realising intensive-computing ecosystems that are 'green'.

These ecosystem solutions must face four important energy consumption areas:

- (a) Reduce the energy consumed by the existing big data centres, for example, reducing data centralisation and pushing data and computing distribution towards the edge of the cloud. Another promising area is the improvement of the existing cooling technologies.
- (b) Limit data movement reducing the energy consumed from machine-to-machine communications; this can also be achieved by pushing data and computing towards the edge of the network. Important improvements in this direction come from smart sensors (e.g. embedded-AI systems) and fog and edge clouds (cloudlets).
- (c) Reduce the embodied energy (i.e. energy consumed in the production of goods) of digital technology by implementing the Industry 4.0 Smart Factory principles and in particular the smart energy pillar.
- (d) Reduce the obsolescence of digital technologies (see, for example, the short lifespan of mobile phones, replaced every one to three years), which is perhaps the main energy consumption voice for digital technology.

We need to accelerate the speed of energy consumption reduction through energy efficiency, supporting a holistic and multi-disciplinary approach e.g. by introducing and using innovative semiconductor technologies, smart sensors, efficient data transmission and DCs, and advanced computing architectures, to compensate for the speed of data traffic increase. For that we need to carefully estimate the amount of energy required by AI (and the IoT), taking into consideration all the required phases – e.g. data generation, moving, storing, processing, and insight generation and provision.







Good quality shared data is essential to develop socially responsive AI

SUMMARY

In this chapter, we review the key economic characteristics of digital data, including economies, of scale, of scope, and non-rivalry. The first two point to an increase concentration of data, and hence information and power, in the hands of a few actors in the internet economy, while non-rivalry creates potential tensions between opening access to the data so that society benefits the most, and restricting access so that the data holder benefits most. These tensions are reflected in the legal framework, as discussed earlier, and make the development of data strategies in an open and globalised environment a particular challenge. We argue, however, that if we apply the lessons of successful internet companies to the European public sector and develop ecosystems based around public platforms, it is possible to create large pools of shared data that are semantically well-structured and labelled, and can fuel new AI applications in different domains. In this way, we can open access to the data, develop the market, serve the public, and enrich the data commons at the same time.

DATA PERSPECTIVE

■ 12.1 The law and economics of access to data

ML is a data-driven technology, as discussed in Chapter 2. It is necessary for us to make sense of the enormous volumes and range of data being produced daily, but at the same time ML needs large volumes of data to learn. For this reason, as we have seen in Chapter 4, all the countries developing AI strategies put access, and adding value to data as key concerns.

Developing such strategies is not easy because they need to balance the often-competing interests between opening access to data to promote innovation and transparency, and restricting it to protect privacy and commercial confidentiality. Europe is also in a particularly difficult position because the global data market is characterised by a few dominant non-European players and platforms that strengthen their power as they gather more and more data from European customers and governments.

There are many facets to this problem. Therefore, in this chapter, we outline the key economic characteristics of digital data, review the EU regulatory settings and their influence on the trade-offs between openness and restriction, and then put forward the elements for a data-sharing strategy for the public sector.

■ 12.1.1 The economic characteristics of data

Digital data has three main economic characteristics:

Economies of scale: ML shares many similarities with statistical modelling. Statistical estimates become more reliable as the size of the underlying da-

Data is the lifeline of AI. Europe is data rich but the economics and legal framework create complex challenges. Opening access to data and building interactions among participants is key to succeeding.

taset increases and variation in the data decreases. With a high variety of observations and many explanatory variables in the dataset, achieving robust predictions may require very large datasets. Typically, ML algorithms need much more data than humans in order to train for comparable skills. For example, an automated driving algorithm may require millions of kilometres of driving data while a human driver needs only a few thousand kilometres of experience to become a proficient driver. Collecting and annotating large datasets is difficult, the quality of raw data is often poor and needs cleaning and structuring, and a lot of the data are often useless. Investing in high-quality datasets for training ML algorithms is costly and implies high fixed costs. Once they are trained, however, the marginal cost of additional use of the algorithm can be very low. Common ML methods do not perform well with small data sample sizes, especially when data are noisy, i.e. contain many irrelevant variables.

ML models tend to ‘overfit’ in these cases, reducing the efficiency of algorithms once they are applied in real-life settings. One way to address this is to try to understand the underlying mechanisms in the data through causal inference. This potentially increases the interpretability and fairness of algorithms (Kilbertus et al., 2017; Athey, 2017).

Economies of scope: these occur when the benefits of analysing a merged dataset are higher than the analysis of each dataset separately. These benefits only occur when there is a relationship between the two sets, i.e. when they are not completely separable and data that pertain to one situation may also be relevant for another situation. For example, web-surfing data may produce insights on consumer behaviour. Merging these with mobile phone data may produce more insights when compared to studying both datasets separately. Merging with pay data in shops adds further insights, etc. Economies of scope explain why data-driven firms are so data-hungry and collect all the data they can get: the wider the coverage of the available datasets in terms of variety of situations and observations, the more accurate and insightful the predictions that can be made with these datasets.

Non-rivalry: whilst goods and services are rival products, i.e. they can only be used by one person at a time, **an idea, a piece of information or an algorithm are non-rival: they can be used by many people at the same time.** As a result, the benefits derived from data are higher when used by a group of people compared to used by an individual. From a societal perspective, it may therefore be better to share data and algorithms as widely as possible rather than keeping them private. However, the flip side of wide sharing and easy access is the lack of incentives to invest in production of data and algorithms. Data may be a by-product of ongoing activities and require no additional incentives. For instance, e-commerce firms store consumer behaviour data on their websites as part of their online transactions. Public administrations collect loads of data while carrying out their duties. In other cases however, firms may deliberately invest in collecting

data by installing sensors and setting up servers to collect, store, transmit and process the data. Without the ability to monetise revenue from the data there would be no incentive to collect them. This problem was first discovered by authors and publishers of books and by inventors of new technologies. All these innovative ideas and artworks could easily be copied by others, leaving the original creator without any remuneration for his/her efforts. Intellectual property rights (IPR) were introduced to avoid these situations.

Should we apply the same economic principles of IPR to data and assign exclusive rights to data holders, possibly even full ownership rights? As in the case of IPR, these rights should keep a **balance between granting exclusive rights as an incentive for the production of data and ensuring access to data and algorithms in order to produce wider benefits for society.** The applicability of IPR to ML algorithms was discussed in Section 7.3. We come back to legal protection of data rights below in 12.1.3. First, we explore various meanings of ‘data access’.

12.1.2 Applying economic reasoning to data access

The modalities and conditions for access to data vary considerably, with potentially very different economic outcomes for individuals and for society as a whole (Bergmann and Morris, 2017). We should therefore be careful and more precise in what we mean when we use the word ‘access’.

A data holder may not perceive any benefits in directly sharing or selling data to others and decide to keep them private, protected by legal and/or technical protection measures. Rather than selling data directly, they can be used indirectly as input in the production of a commercial service. For example, Google does not directly sell or share consumer data but uses them for selling advertising slots. Consumers benefit from the information services while advertisers pay for the advertising service.

In this way, even non-accessible data can generate benefits for individuals and for society.

Data holders can also make the data directly available. If the data are rather exclusive and valuable, they may fetch a price in the data market. Since data are non-rival they can be sold many times. Monopolistic sellers of exclusive data can either sell at a fixed price or adjust prices according to each buyer's willingness to pay. In that case, everyone has access and there is no welfare loss to society. The data monopolist maximises monetary benefits from the data. A welfare distribution issue may occur when most of the benefits accrue in monetised form to the data holder and less so to data users or the person whose behaviour generated the data. If the data are less exclusive and competing sources for (partial) substitute data are available, the monopoly can be weakened and more benefits will accrue to data users. The choice between direct and indirect data sales is often not straightforward for data holders; some may do both in parallel. For example, many data firms sell data directly and provide data analytics services.

A different situation occurs with open or free access that makes data available to all potential users at zero cost. Benefits are distributed across the widest possible group of users, similar to the perfect price discrimination case above, except that all benefits accrue to data users, not the data holder. If data are a by-product of the data holder's ordinary activities and can be collected at low or zero cost, free access will not affect the welfare of the data holder. Even when publishing the data carries a cost to the data holder, open access may prove beneficial to the holder because greater use of the data can act as an advertisement, and improve data quality through validation and error detection at limited cost for the data holder. However, if data collection is costly and their use is very valuable, the holder may not have an incentive to grant free and open access. Losing revenue may also reduce incentives for data collectors to invest in innovative services that could be produced with the data.

Another scenario occurs when data are extracted from data holders against their will, possibly without their knowledge and at a cost to them, but may be of benefit to society. Examples include mandatory open data access for environmental reporting and access to private data to detect or prevent crime. Individual data holders may incur costs but the net benefit to society is positive. Another example is online firms that test various commercial strategies on a small sample of consumers before deploying them on a larger scale. The consumers in the test sample may gain or lose from the experiment, without knowing it, but the firm may benefit from extracting behavioural information from users.

Trust is an important factor in data sharing and access. A variety of initiatives have been deployed to strengthen trust while sharing data. Experimental personal information management spaces (PIMS) give individuals more control over the use of their personal data. Similarly, 'data trusts' seek to provide trusted exchange mechanisms via legal conditions, storage formats and mutual oversight organisations.

12.1.3 Policy intervention in the data market

Extracting insights from datasets by means of ML or other techniques can be very beneficial for some but the insights can also be exploited to redistribute benefits between parties. Sharing access to data can potentially increase the benefits but may also result in more redistribution. Should policymakers intervene in data ownership and access markets, or should they be left to operate freely? Translated into economic jargon, this raises the question of potential data market failures that reduce social welfare and require policy intervention. Such failures may not only occur in primary data markets but also in downstream markets for goods and services where transactions depend on access to data. The extent of access and competition in upstream data markets may have implications for competition in downstream goods and services markets.

Data may have very beneficial uses for society in terms of innovation and welfare. However, data can also be used to reduce competition in markets and nudge consumers into decisions that are not necessarily to their advantage. There is often a trade-off between short-term price efficiency and longer-term innovation incentives.

Regulators can create exclusive control rights for data holders or they can soften these exclusive rights and punch some holes in them to facilitate access to the data. In reality, regulators often do both at the same time – as we will now illustrate for the case of the EU.

In the EU, there are two general horizontal legal instruments that have something to say on data rights: the General Data Protection Regulation (GDPR) (EC 2016a) for personal data of natural persons and the Data Base Directive (DBD) (EC, 1996) for any data collected by firms (see Section 7.2).

The EU GDPR sets some rules regarding the use of personal data. Any data that can be linked to a natural person is considered to be personal data, irrespective of the source, storage or transmission mechanism. The GDPR does not grant ownership rights to personal data but gives data subjects a number of control rights over their personal data, including the right to obtain consent from the data subject before accessing the data, to the right for data subjects to access their personal data and delete them, to port their data to other uses, etc. Privacy Personal data protection is considered to be a basic and inalienable human right that cannot be traded away. For this reason, personal data are assumed not to be the subject of an economic transaction. However, a frequently used online business model involves the exchange of personal data in return for a data-based service. The data subject can accept or refuse the exchange, depending on whether s/he perceives it as a fair and credible deal that will bring more benefits than potential harms from reduced privacy.

Some provisions in the GDPR reduce access to personal data. For example, they prohibit the use of personal data for purposes other than the originally intended purpose, unless they are anonymised. This makes it difficult for holders of personal data to aggregate them into larger databases or share them with other types of data services for the purpose of aggregation and analysis by means of ML algorithms. Other provisions in the GDPR facilitate access to data. For example, Article 20 mandates portability of personal data. At the request of the data subject, the data holder has to transfer the personal data to a destination of choice of the data subject. The data subject can thus choose to give wider access to his/her data. Anonymous data can be transferred or used for other purposes. A good example is navigation apps. An important source of road traffic data is the movement of mobile phones in cars. Telecom service operators track the positioning of mobile phones as a technical measure to keep the phones connected to the antennas of the mobile phone network. The GDPR allows the transmission of anonymous data about movement of mobile phones. The GDPR puts no restrictions on cross-border transfers of data within the EU. However, **transfers outside the EU can only be made to operators that respect the GDPR.**

The EU Database Directive (DBD) (EC, 1996) contains two provisions: copyright and a *sui generis* right. Copyright protects the structure of databases which, if original, constitutes the author's own intellectual creation. By contrast, the more controversial *sui generis* right protects databases regardless of their originality, as long as there has been 'substantial investment in obtaining, verifying or presenting the contents'. Both the copyright and the *sui generis* right in the DBD put the database owner in a position of exclusive access and therefore reduce access to the data. A recent evaluation of the DBD (EC, 2018g) concluded that it has been effective in harmonising the EU legislation on databases and avoiding fragmentation, and that it still provides an appropriate balance between protection of investment and interests of users. On the other hand, it also notes that as a result of a 2004

decision of the Court of Justice⁵³ clarifying that the *sui generis* right does not apply to databases that are the by-products of the main activity of a firm, it must be assumed that these rights do not apply broadly to the data economy (machine-generated data, IoT devices, big data, AI, etc.). The evaluation concluded that at this stage reforming the *sui generis* rights would be largely disproportionate. The debate among scholars on the benefits of recognising such rights particularly for machine-generated data is still very open (see Section 7.2).

The text and data mining (TDM) exception to copyright protection could also be considered as a measure that opens access to data. In the EU, the TDM exception is limited to non-commercial use of the data by public sector research institutes and organisations. That is an opening combined with a restriction. In the USA, the ‘fair use’ provisions allow for wider use of data obtained through TDM, provided they do not compete with the services offered by the data holder (Section 7.2).

The EU Public Sector Information (PSI) Directive (EC, 2003) seeks to make data held by public-sector organisations more accessible to citizens and firms. To the extent that some of these organisations may be involved in the production of commercial services – for instance public utilities in transport and energy – it may also cover some commercial operations. Providing public access to these commercially sensitive data may entail opportunity costs for the data holders. The Directive suggests that pricing of data access should incur marginal costs. Digital datasets are often characterised by the high fixed costs of creating the dataset and virtually zero marginal costs of replicating and transmitting it. As a result, the marginal cost rule may not cover the actual cost of producing public-sector data but, as argued earlier, much of this data is generated to fulfil a legal mandate in the first place, so its reuse becomes a by-product, and can benefit the public administration much more than keeping the data closed. This is the case, for example, if the data generates new products and services that the public administration can use

to extract additional information and/or serve the community better. A 2015 study funded by the EC⁵⁴ supports the view that public administrations will be by far the largest beneficiaries of opening up the data they hold.

Since data are non-rival and can be transferred at virtually zero cost all over the world, the question of ‘data sovereignty’ has become an issue. This revolves around the idea that data are subject to the laws and governance structures in the nation where it is collected⁵⁵. Within the EU, the GDPR allows transfer of personal data between Member States. The recent EU Free Flow of Data policy initiative aims to extend this to all types of data in order to avoid national fragmentation of data markets. Personal data transfers outside the EU require that the recipient complies with GDPR rules.

“Since data are non-rival and can be transferred at virtually zero cost all over the world the question of ‘Data Sovereignty’ has become an issue.”

■ 12.2 Summary

(Big) datasets are a necessary input into ML algorithms. Access to data is important for the development of AI. Access and wide data sharing can occur in markets for direct and indirect data sales, across a wide spectrum of modalities and economic conditions, even when data holders have exclusive control over their data.

EU regulatory interventions in data markets are caught between two poles: (a) offering more exclusive rights as an incentive for data producers and holders to invest more in data collection & analysis, and **(b) making data more widely available** and accessible to facilitate the extraction of new insights from data, including with AI/ML. There is a wide-open unregulated space between personal data rights under the GDPR and general database ownership rights under the DBD where market-based data exchanges, bilateral contracts and technical protection measures rule. There is an ongoing debate in the EU as to whether this open space contains some market failures that need to be filled up with further regulation. That debate ranges from expanding exclusive data ownership rights (European Commission, January 2017) to facilitating access to data (European Commission, April 2018). The rapid rise of AI/ML as a very promising data-processing technology to extract more insights and value from data is fuelling this debate. When data and data analytics technologies become cheaper and more widely available, the need for protection as an incentive to investors may diminish in favour of more access to stimulate innovation. Finding a new balance between these two opposite poles in the debate requires more societal and policy debate.⁵⁶

In the next section we consider some possible strategies to extract value from the data while sharing them.

■ 12.3 Towards a data strategy for public administrations

■ 12.3.1 Setting the scene

At several points in this report we have highlighted that data is crucial to the development of AI in general and ML in particular. Robust algorithms need large volumes of good-quality data for training, where ‘good’ means data that is well organised, semantically labelled, free of bias and open, as discussed in Chapters 2 and 6.

Public administrations could benefit significantly from the use of data science and AI, for example to design policy that is targeted to groups and geographical areas that are in greatest need of intervention, modelling the potential cumulative impact of policy in different areas, and developing policies that ‘learn’ from feedback loops from policy implementation in analogy to the way ML adapts based on data feeds (Craglia, Hradec and Troussard, 2019). For this reason, most national AI strategies summarised in Chapter 4 consider AI as a great opportunity to modernise public administration, public services and research centres like the Alan Turing Institute in the UK, which has developed an active public policy programme.

In the previous section, we have highlighted the tensions between opening up data and restricting access and use, from both an economic and legal perspective. These tensions are exacerbated by the international dimension: in the global competition between the USA and China, and among the key industrial players (Google, Amazon, Facebook, Apple and Microsoft on the one hand, and Baidu, Alibaba, Tencent and Xiaomi on the other), Europe is notably absent. As noted by the Mission Villani (2018), whilst most data captured in the USA and China, and to an extent Russia, stays in their respective countries because of market dominance, policy, or both, in Europe the vast majority of data provided by European citizens, companies and governments ends up outside Europe. Potentially, the more the public and research sectors pursue an open data policy, the more

they feed this flow of data outside Europe. In this context, we face a real challenge, if we want to stay open, in developing a data strategy for Europe that increases the pool of European data, adds value to it by structuring and semantically annotating it, and trains AI applications that benefit European society. Many countries are grappling with this challenge, as indicated in Chapter 4. One country that has better articulated a possible response is France where the report by the Mission Villani (2018) put forward a three-pronged data strategy (see Box 9).

- 3) Exploring new forms of partnership with the public to share personal data, leveraging their increased control on access to such data provided by the GDPR.

Opening and sharing data across all key stakeholders – public sector, commercial sector and the public – is very important and already a challenge. Adding value to the data while sharing it is an additional crucial dimension to support the development of AI beneficial to Europe. In this respect, we can learn from the private sector as outlined in the next section.

BOX 9. Summary data strategy from the Mission Villani Report (2018)

1. Encourage companies to pool and share their data

The government must encourage the creation of data commons and support an alternative data production and governance model based on reciprocity, cooperation and sharing. The goal is to boost data sharing between actors in the same sector. The government must also encourage data sharing between private actors, and assist businesses in this respect. It must organise for certain data held by private entities to be released on a case-by-case basis, and support data- and text-mining practices without delay.

2. Create data that is in the public interest

Most of the actors heard by the mission were in favour of progressively opening up access to some data sets on a case-

by-case and sector-specific basis for public interest reasons. This could be in one of two ways: by making the data accessible only to the government, or by making the data more widely available, for example, to other economic actors.

3. Support the right to data portability

The right to data portability is one of the most important innovations in recent French and European texts. It will give any individual the ability to migrate from one service ecosystem to another without losing their data history. This right could be extended to all citizen-centred artificial intelligence applications. In this case, it would involve making personal data available to government authorities or researchers.

This proposed strategy is interesting because it builds on three key dimensions:

- 1) Using the EU legal framework, including the GDPR, as an opportunity not just to safeguard European interests but also to nurture European companies.
- 2) Framing the opening of public-sector data in a two-way relationship of reciprocity with the commercial sector to gain increased access to some commercial data of public interest.

12.3.2 About platforms and APIs

From an economic perspective, a (digital) platform is where two or more types of users (consumers, suppliers, advertisers, software developers, etc.) come together to exchange goods, services and information. Successful platforms make use of network effects to attract more users on several sides of the market: more consumers attract more suppliers and vice versa. They also leverage the data that they collect on user behaviour on the platform to reinforce their own position. Economies of scope in data collection and analysis give them an

informational advantage over individual suppliers and consumers (see Martens, 2016).

From a computation perspective, a (web) platform is an evolution of the web from a medium to publish information to an infrastructure to build applications. One of the major components of web digital platforms is the use of web application programming interfaces (APIs). APIs allow third-party developers to access selected data and functionalities available on the platform, and build new applications. APIs let the data provider keep better control of how the data is used, and let the third-party developer build applications without the need to download the entire set of data.

A good example is GoogleMaps, which is accessible to developers via an API. In this way, the developer, for example, of a hotel's website does not need to download the map data showing where the hotel is located or how to get there. It embeds the GoogleMaps API in the application to call its functionalities, and shows the locations at different scales or directions to get there from anywhere in the world. This provides a clear advantage to the developer, but also to the Google platform because it can read every call on the application and see how popular the hotel is, from where it gets its customers, how it is associated with other facilities, and so on. In other words, the platform gains intelligence by monitoring the use of the API. This links to the economic perspective above: the more users of the APIs, the more intelligence the platform gathers, the more it reinforces its position, and the more it attracts new developers and customers.

Public administrations across Europe are starting to use APIs to provide access to their data in machine-readable form. When they do, they often see a significant increase in the use of their data compared to publishing it as downloadable datasets (see Box 10).

BOX 10: Examples of APIs in local government

Denmark's Addresses Web API (DAWA): DAWA is a service allowing public authorities, businesses and citizens to use authoritative addresses in their IT systems. In 2017, it received 1.5 billion requests and approximately 350 000 unique users per week. The number of API requests is limited to 100 requests per second. There are approximately 5 000 IT systems drawing data concerning Danish addresses through DAWA.

Madrid MobilityLabs: this is an open and interoperable platform from the city council allowing third parties to develop applications in the field of urban mobility. It receives 480 million requests per year. 2 500 developers are registered in the system, and around 100 apps have already been developed.

Source: ELISE project, Joint Research Centre.

12.3.3 From opening data to smart sharing

Using APIs to make the data available to third parties is an essential step towards developing the data economy, but is not enough to underpin an AI strategy. **What is necessary is for the European public sector, in particular, to move from pushing the data out – a *broadcasting* model – to drawing the users in, both third-party developers and the public, and to develop interaction, the *interactive* model.**

We see from the example of commercial platforms that users interact in different ways:

- Providing new content.
- Annotating/tagging/classifying the content already available on the platform.
- Using the services offered.
- Building/using new apps and functionalities based on the platform APIs.

Users' behaviour (interaction) and content generated (or exchanged) contribute to form the enhanced data that is so important for AI. Raw data as such is frequently of little value, but this increases when it is contextualised and cross-referenced. Facebook is the biggest content manager in the world, but does not produce any content. Its users do it.

If we want to take advantage of the enormous diversity and richness of European data, we should learn from the examples of successful internet companies and develop European data ecosystems bringing together the public sector, the commercial sector, academia and the third sector and the general public. Developing such ecosystems requires a new partnership, for example, in the form of a social

contract among all the key stakeholders so that each agree on the sharing and use of their data for clear purposes and clear benefits.

We have seen in the review of national AI strategies in Chapter 4 that several countries are already reflecting on how to share data between the public sector, private sector and the public, for example through data trusts (UK), social contracts (France) or MyData (Finland). Each country may have its own model but we need to share the best practices and synergies and agree a framework to then enrich this data for the public good.

There is general agreement that European AI applications must be based on, and reinforce the European values enshrined in the Treaties. These values and the European legal framework must be seen as a source of strength and global competitive advantage, not a limitation for AI. They must permeate all the facets of a European 'datafication' process, i.e. the process of transforming the information we generate into new forms of value or insights. From the existing scientific literature (e.g. Tan et al., 2016; Roy et al., 2018; Opher et al., 2017) we see that this process can be organised into three main phases, (see Fig. 14): data collection and aggregation, data transformation/enhancement, and insight generation and provision.

Data collection and aggregation phase

The first datafication phase involves setting up efficient and long-term data collection and storage solutions. This is essential to provide a stable foundation for any analytical process. Infrastructure is necessary to facilitate and support the formation of collaborative ecosystems, at different levels (national, European, and community or sectoral ones), to constitute a reliable, flexible, and scalable supply-chain system for data, and processing. Valuable examples could include public administration data ecosystems, industrial data ecosystems, for example, in the space sector, Industry 4.0, health, etc. In the latter, data collected and provided by the public would also be crucial.

“ We should learn from the examples of successful internet companies and develop European data ecosystems bringing together the public sector, the commercial sector, academia and the third sector, and the general public. ”

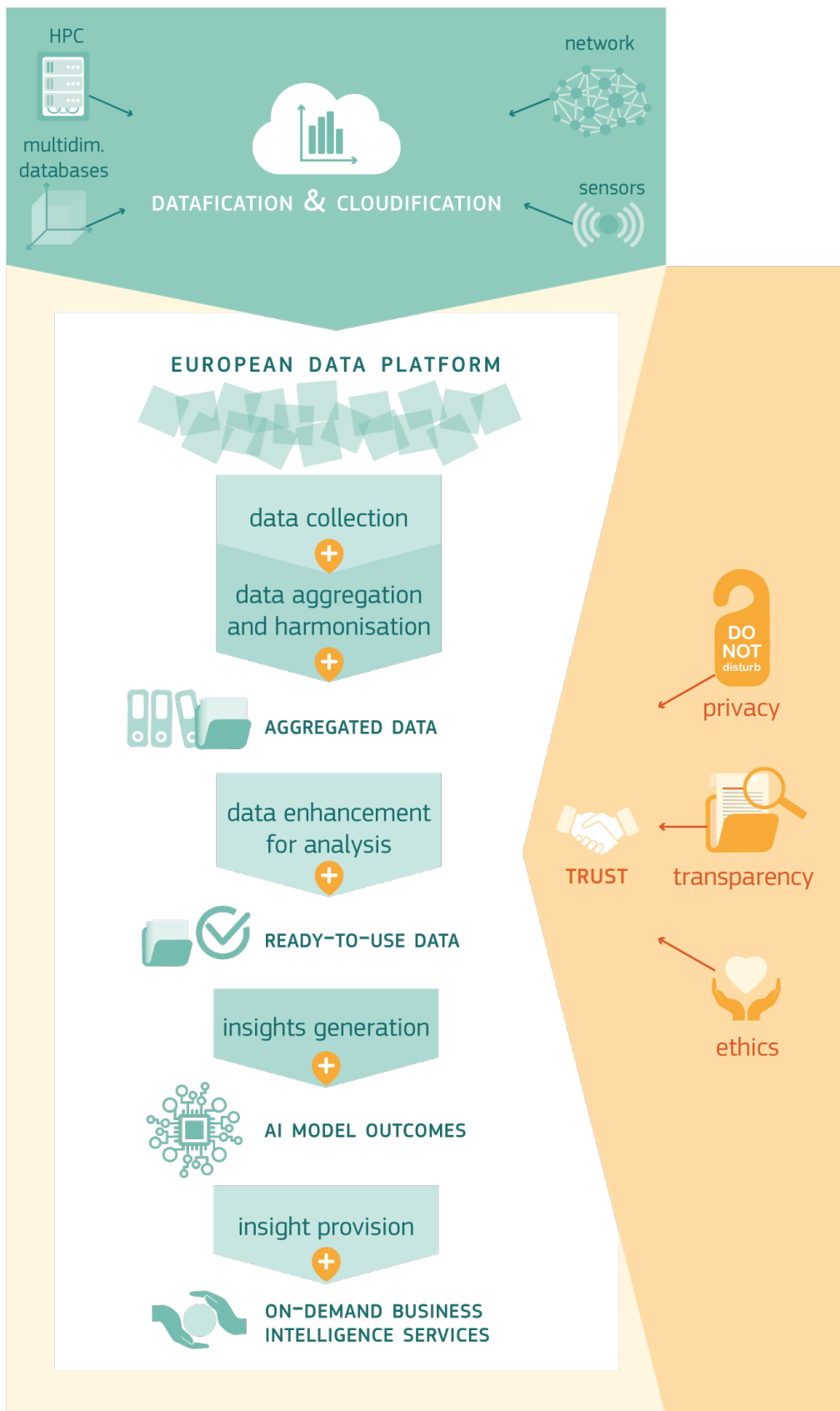


Figure 14. Datafication process enabling the AI revolution.

As discussed in Chapter 13, if we want to use the opportunities afforded by AI to build competences, skills, and resilience at the local level, particularly in peripheral regions, we also need to consider local-level ecosystems bringing together public administrations, local firms, academia, and local communities to share and enrich local data and build applications addressing local needs.

Data transformation/enhancement phase

In this phase, different data streams and types need to be transformed, harmonised and aggregated to increase the value of the data collected and make it useful for AI application. This transformation phase requires domain expertise, which may not be available everywhere in the same way. Therefore, it is important to develop the transformation algorithms from the outset so that they are transparent and reusable. This is the key level where we need to address the concerns of transparency, explainability, accountability and fairness expressed in Chapter 2.3 and also also discussed in Chapters 6. To do so, requires the adoption at the European level of protocols to document, share and reuse algorithms, and agreed evaluation frameworks for algorithms.

Insight generation and provision through the platform

This last development phase is essential to match demand and supply and create real economic value. In this context, a platform is the place that enables interactions between customers and service providers, where the latter can be the public administration or third-party developers. In this phase, ML algorithms are developed and trained with the pool of data available to respond to user demand or local needs, tested, and deployed to provide a service. Important aspects to achieve a European-level strategy for data services and boost the European AI economy would include developing a framework for sharing and reusing AI/ML training datasets and algorithms, defining minimum quality of service, and ensuring platform interoperability and portability of data.

■ 12.4 Summary and conclusions

In this chapter, we review the key economic characteristics of digital data, including economies of scale, of scope, and non-rivalry. The first two point to an increase in the concentration of data, and hence information and power, in the hands of a few actors in the internet economy, while non-rivalry creates potential tensions between opening access to the data so that society benefits the most, and restricting access so that the data holder benefits most. These tensions are reflected in the legal framework, as discussed in Chapter 7, and make the development of data strategies in an open and globalised environment a particular challenge.

To address this challenge, we have put forward some key elements of a possible **strategy to make the best use of the vast richness of European society and data**. We suggest that the tensions and dilemmas between openness and restrictions outlined in Section 12.1 can be navigated by learning from the most successful companies on the internet. We need to add value to the data we have, particularly in the public sector, by moving from the broadcasting model to the interactive one in which users contribute to enrich the data whilst using it, and providers gain intelligence from the feedback they receive. This means **developing European data platforms with a clear datafication strategy articulated around data collection and aggregation, transformation and enhancement, insight generation and provision**. Through these phases we can also address the concerns over transparency, accountability, and fairness expressed in Chapters 2 and 6 and find a European way to AI development that is consistent with our values.



*'In winter, snow falls on the bamboo trees. They bend but never break, and come spring, the trees start to grow straight again.**

Address by Norio Mitsuya, Parliamentary Senior Vice-Minister, Ministry of Foreign Affairs, Japan (MCM Chair); OECD Southeast Asia Regional Forum, 26 March, Bali, Indonesia, 2014. <https://www.mofa.go.jp/mofaj/files/000033886.pdf>

SUMMARY

In this chapter, we introduce the concept of resilience which is useful to frame a European approach to AI addressing different phases: prevention, anticipation, preparation, but also adaptation and transformation to bounce back from the effects of a shock. It is important to both measure and monitor resilience, and we introduce some examples of how this could be done even though there are a number of methodological and data-related limitations that need to be addressed.

As we show in Chapter 9, we do not know what impacts AI will have on work and society, but we do know we need to increase our resilience to AI, particularly in regions of Europe already more vulnerable. We put forward some elements of a possible approach centred on the network of Digital Innovation Hubs. There are already several hundred hubs at the present time which are expanding rapidly with a target of one for each region in Europe, many of which are focusing on AI. Their purpose is to facilitate access to technology and know-how to public administrations and firms, SMEs in particular, in combination with the planned AI-on-demand-platform. We suggest putting these hubs at the centre of local ecosystems which could create local pools of shared data so that the AI skills developed/upgraded locally could be put to good use in developing AI algorithms and solutions trained on local data to address local needs.

SOCIETAL RESILIENCE PERSPECTIVE

■ 13.1 The need for resilience

As we discussed in the previous chapters of this report, AI offers many potential opportunities to improve our lives but also raises a number of challenges to society. We are preparing ethical guidelines for the development of AI systems that are accountable, responsible and transparent but we must also recognise that future change may be unpredictable and sudden. Therefore, we should not only try and shape the development of technology so that it benefits our society, but also **prepare our institutions, policies, people and society at large to become more flexible, adaptable and ready to transform**. Basically, we need to become more resilient.

Interest in resilience has been rising rapidly during the last 20 years as a response to increasing challenges like climate change, globalisation, immigration, and digital transformation that test our coping capacities as individuals, regions, countries and institutions.

Resilience is a key concept in the policy discourse of the EU⁵⁷, and the JRC has developed a framework (Manca et al., 2017) for societal resilience that includes not only prevention, anticipation and preparation but also adaptation and transformation, i.e. the need to reform and potentially bounce back after a shock. According to this framework, a system (or society) is resilient if it can face shocks and persistent structural changes without losing its ability to deliver societal well-being in a sustainable way (i.e. deliver societal well-being without compromising that of future generations).

Harnessing the richness and diversity of European regions and data is crucial to build resilience and develop AI applications responsive to European and local needs.

This framework may prove very useful in addressing the multiple challenges and opportunities of AI. For example, many workers will need both protection (absorptive capacity), and support for their potential retraining (adaptive capacity). This also highlights the centrality of individual resilience capacities. Preparation is also a major element, including the right policy frameworks (as advocated in the context of globalisation by Baldwin and Vihriala, 2017, or Giles, 2018 for antitrust policies) and the necessary public debates about the ways forward as a society (as raised by, among others, Guillén and Reddy, 2018; University of Melbourne, 2018).

We also need to consider that the effects of AI may transcend geographies and sectors. It is expected to give a new boost to globalisation (Baldwin and

Vihriala, 2017), but could also lead to a repatriation of jobs previously delocalised in cheaper countries. In any case, job losses and reallocations will reshape communities (Alexander, 2018). As an example, self-driving cars will not just enable driving without a steering wheel, but will change a major ecosystem built on car ownership and use (Finn, 2018). Inequality may have an even stronger global component (Lee, 2017). This calls for a *holistic and complex approach (the system view)*.

Moreover, the development and use of AI is set to rise and continue for a long time. We need to look into the future, **assess its effects on people, communities, countries and humanity (societal well-being), and we need to influence its future development (the dynamic reaction)** wisely. As argued by Melvin Kranzberg, ‘technology is neither good nor bad; nor is it neutral’, so it is crucial to actively steer the development of AI to ‘improve the quality of our products and our standard of living’ (Goolsbee, 2017).

Finally, we need not only to become more resilient in the face of AI but also to use AI to become more resilient, for example by extending the use of AI-enabled predictive analytics in public policy, infrastructure maintenance, disaster risk management, agriculture, crime prevention or traffic control (Prakash, 2017). AI can also provide improved technologies to foster citizen engagement in building up a resilient EU society.

■ 13.2 Measuring resilience to AI

It is important to both measure and monitor resilience. Alessi et al. (2018) provide one of the first examples of a full measurement strategy aligned with the JRC conceptual framework focusing on the 2008-12 financial and economic crises. The two-step approach adopted by Alessi and colleagues creates resilience indicators based on the properties of the system’s components, and then looks for statistically significant and robust predictors (resilience characteristics)⁵⁸ of resilient behaviour. The more widespread approach is to identify, typically by expert assessments, such characteristics

directly. One such example is JRC’s INFORM Index for Risk Management (De Groeve et al., 2014).

In the case of AI, expert-based indicators are already being developed. One example is Pau et al. (2017) who looked at a set of Asian countries, and prepared two composite indicators, for AI preparedness and resilience. Preparedness is ‘the ability of companies and talent to capitalise on opportunities brought about by AI’, while resilience – in their approach – is the ‘ability to adapt to and withstand structural changes brought about by AI’.

The main ingredients for preparedness are start-up activity, AI funding (venture capital), students enrolled in science, technology, engineering and mathematics (STEM) subjects at top-ranked universities, and AI publication volume. Resilience is assessed using relevant government policies and the job structure. In our terminology, preparedness can be viewed as a part of resilience, as it signals the ability to respond well to AI penetration. Below we present three illustrative examples of similar indicators for the EU.

As an attempt to assess the penetration of, and the dynamic response to AI in Europe, Figure 15 shows the shares of manufacturing versus healthcare in employment, comparing the USA and the EU (unlike the Asian example, we look at the structure by sectors and not by job types). A major decline in manufacturing and a rise in healthcare can be seen in both entities. The EU still has a larger share of manufacturing and a smaller share of healthcare. This indicates the possibility of further declines in manufacturing, but also in the absorption capacity of healthcare. The degree of relative shift was larger in the USA (6.8 percentage points, as opposed to 5.6), which may explain the higher degree of pessimism (and also the change in inequality) in the USA. Although the underlying mechanisms require further analysis, the decline in manufacturing is informative about the penetration of automation, while the increase in healthcare and its further absorption capacity can be linked to resilience.

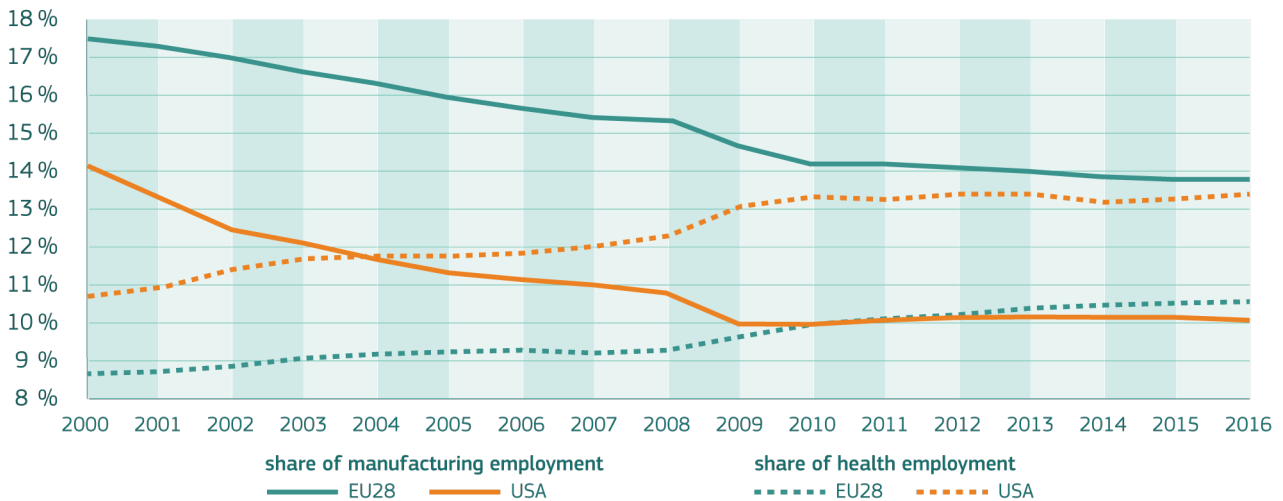


Figure 15. Employment in manufacturing and healthcare: EU and USA

Source: Eurostat for the EU, OECD for the USA

Figure 16 shows a second measurement attempt, this time looking mostly at preparedness at the Member State level. It combines various aspects of AI-oriented government efforts (based on the 2016 edition of the Open Data Barometer) and data on the share of relevant fields in higher education.⁵⁹ The figure shows that for about half of the countries all variables point to the same direction⁶⁰, so government efforts and education patterns portray a coherent picture of preparedness.

Finally, Figure 17 looks at preparedness at the level of EU regions, showing the share of people employed in science and technology⁶¹, as a percent of the economically active population. Some regions in the Southern and Eastern part of the EU look particularly unprepared. **This calls for upskilling, but also action to reduce the brain drain from the South and East to the North.** The network of Digital Innovation Hubs, which is currently expanding from few hundreds to potentially over 1 000, could offer an opportunity to bring together the training supply from academia and technical centres, with potential employers in the private and public sector, and local banks of shared data. In this way, **local AI skills and applications could be developed together to address local needs, and strengthen resilience at the same time.** It is worth noting that at the present time, some of the countries with the low science and technology employment regions, like Bulgaria and Romania, still lack such hubs, suggesting directions

for possible future action. In Figure 18, we bring together both Digital Innovation Hubs and current academic supply of study topics on AI, as qualified in Chapter 8. In this way, we can start reflecting on potential synergies, gaps and priorities to develop such local ecosystems.

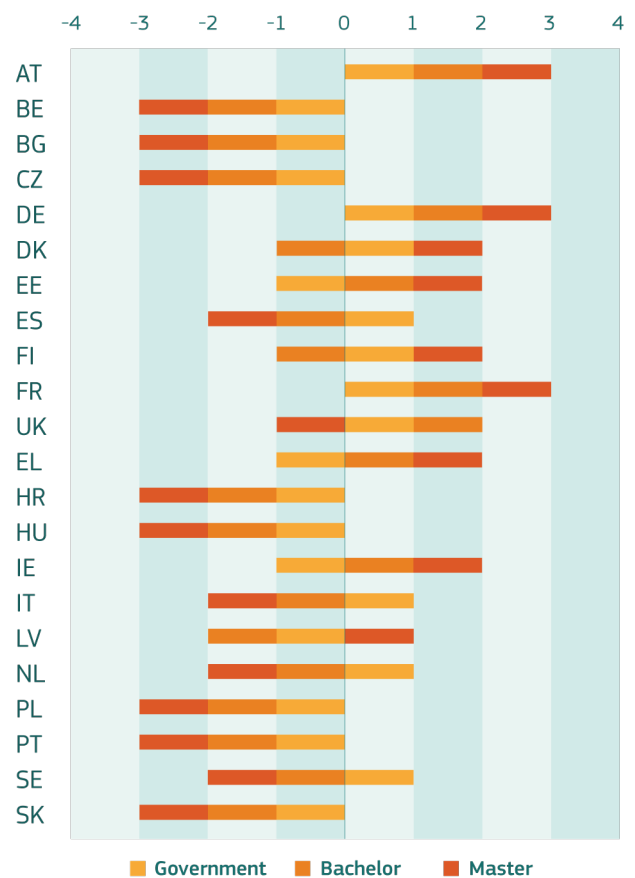


Figure 16. AI preparedness in EU Member States

Source: Source Open Data Barometer and Eurostat.

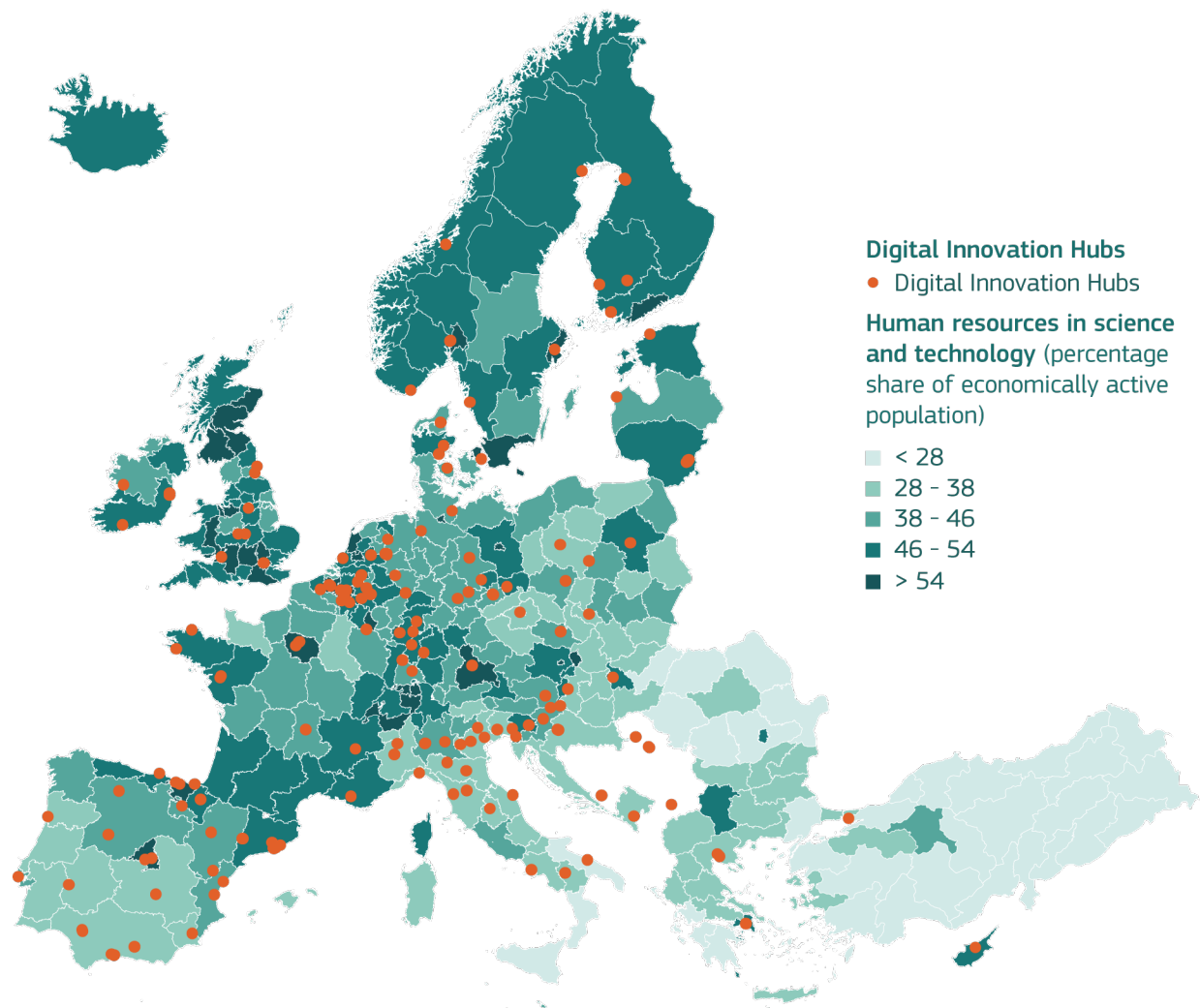


Figure 17. Human resources in science and technology and Digital Innovation Hubs
Source: Eurostat regional yearbook 2018 and JRC.

While we are making some progress in understanding how to measure resilience to AI, there are some difficulties in applying to AI the causal, two-step approach that Alessi et al. (2018) used to measure the response to the financial crisis. For the first step of the approach, which involves assessing the strength of the phenomenon, we lack a long-term time series of AI adoption in different countries. **The AI Index and its planned EU extension by the JRC may offer a starting point to address this limitation.** Another possibility is to use corporate- and citizen-level surveys: Morikawa (2017) presents such examples for Japan, while various special issues of the Eurobarometer survey (SP382 in 2012, SP427 in 2014 and SP460 in 2017, for example) are indicative of public opinion in different European countries. For example, the three surveys above show that only a small share of EU citizens (around 22 %, quite stable in time) would be comfortable having a medical

operation by a robot, while a somewhat bigger share (47 % in 2012 and 2014, but declining to 35 % in 2017) would be comfortable with robots at their workplace. On a more positive note and looking at the broader public perception of digital transformations, the 2017 Eurobarometer survey (SP460), indicates that three quarters of the respondents think the digital transformation has a positive impact on the economy. Two thirds (67 %) think that the impact is also positive on their quality of life, while 64 % also find a positive impact on society. Taken together, these results make it possible to start performing some first resilience analysis while building up the time series on AI adoption.

Another set of challenges relates to the second step of the Alessi et al. model, which involved robust statistical predictors of resilient behaviour. Here the main problem is that AI is already pervasive across

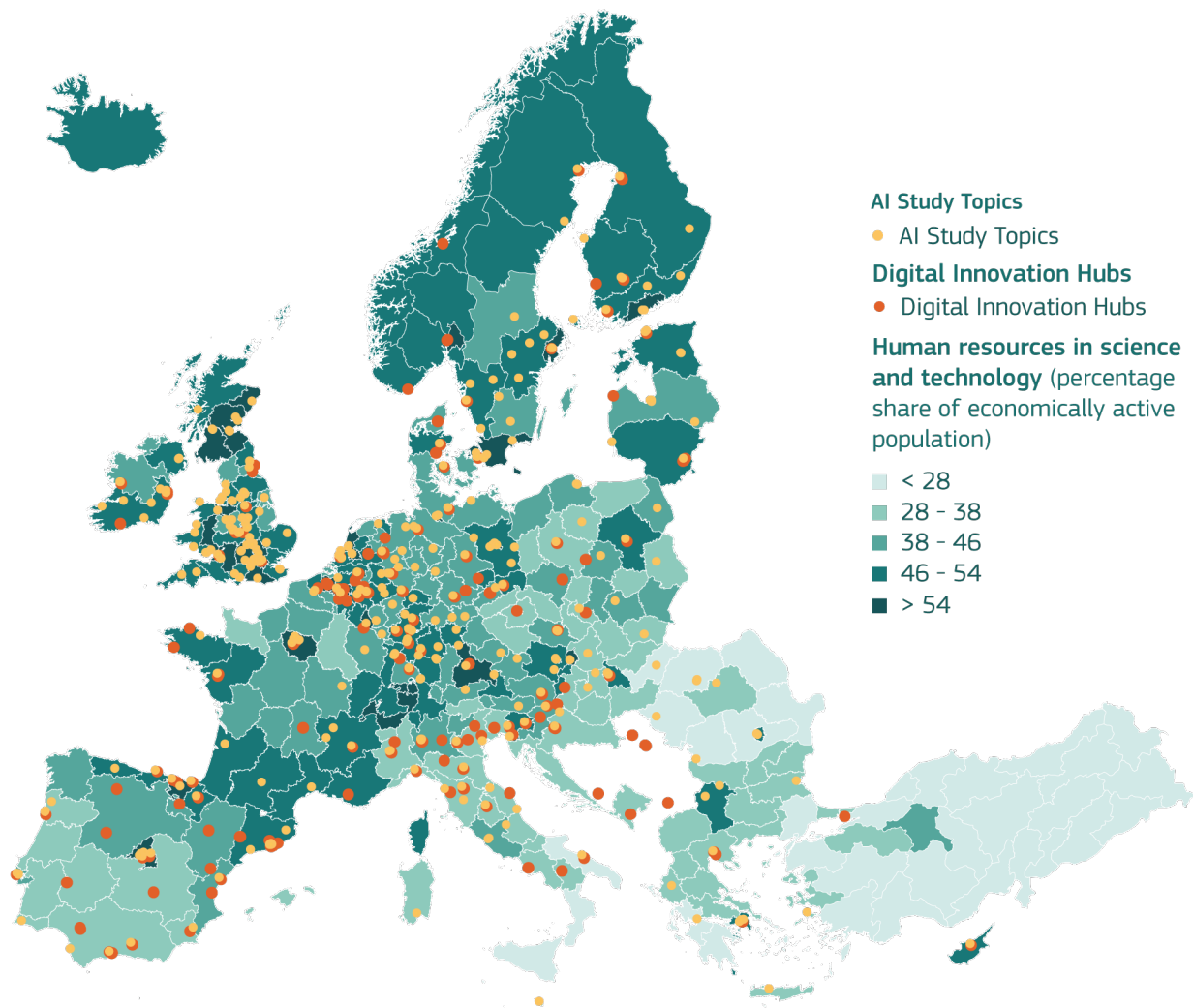


Figure 18. Human resources in science and technology, Digital Innovation Hubs, and academic supply in study topics related to AI
Source: Eurostat regional yearbook 2018 and JRC.

many sectors of the economy and society and therefore it is very difficult to disentangle process from response, and attribute causality. The way we respond to the effects of AI may influence and steer the extent and severity of the effects themselves. This steering process is a key aspect of resilience. One possibility for measuring the effectiveness of steering is to keep track of predictions and then compare them to realised outcomes some time afterwards. For example, the predictions of job displacements and replacements, which are based on current forecasts of technology developments, could be recorded. Systematic differences between the realised and forecast numbers may be indicative of changes in the automation process. Other possibilities may emerge as we apply the resilience framework to AI and progress in practice.

13.3 Lessons from the analysis of AI for understanding resilience

Looking at AI from a resilience perspective does not only enrich our understanding of, and preparedness for AI, but also advances the analysis of resilience. The case of AI is one of the first examples at the JRC of considering a persistent structural change, which calls for an adjusted conceptual and measurement approach. It provides an interesting example of the need to think about different capacities and interventions at different time horizons. It emphasises the role of power relationships, and the issue of potential updates of the 'social contract'. Finally, it calls for steering the future path of AI to the benefit of humankind, and thereby coming out stronger from the adoption of AI.

SUMMARY AND CONCLUSIONS

We are only at the beginning of a rapid period of transformation of our economy and society due to the convergence of many digital technologies. Artificial Intelligence (AI) is central to this change and offers major opportunities to improve our lives.

The recent developments in AI are the result of increased processing power, improvements in algorithms and the exponential growth in the volume and variety of digital data. Many applications of AI have started to appear in our everyday lives, from machine translations, to image recognition, and music generation, and are increasingly deployed in industry, government and commerce. Connected and autonomous vehicles, and AI-supported medical diagnostics are areas of application that will soon be commonplace.

There is strong global competition on AI among the USA, China and Europe. The USA leads for now but China is catching up fast and aims to lead by 2030. For the EU, it is not so much a question of winning or losing the race but of finding a way to embrace the opportunities offered by AI in a way that is human-centred, ethical, secure, and true to our core values.

The EU Member States and the European Commission are developing coordinated national and European strategies, recognising that we can only succeed together. We can build on our areas of strength, including excellent research, leadership in some industrial sectors like automotive and robotics, a solid legal and regulatory framework, and very rich cultural diversity also at the regional and sub-regional levels.

It is generally recognised that AI can flourish only if supported by a robust computing infrastructure and good quality data:

- With respect to computing, **we have identified a window of opportunity for Europe to invest in the emerging new paradigm of computing distributed towards the edges of the network**, in addition to centralised facilities. This will also support the future deployment of 5G and the IoT.
- With respect to data, we argue in favour of learning from successful internet companies, opening access to data and developing



interactivity with the users rather than just broadcasting data. In this way, we can **develop ecosystems of public administrations, firms, and civil society, enriching the data to make it fit for AI applications responding to European needs.**

We should embrace the opportunities afforded by AI but not uncritically. The black box characteristics of most leading AI techniques make them opaque even to specialists. AI systems are currently limited to narrow and well-defined tasks, and their technologies inherit imperfections from their human creators, such as the well-recognised bias effect present in data. We should challenge the shortcomings of AI and work towards strong evaluation strategies, transparent and reliable systems, and good human-AI interactions.

Ethical and secure-by-design algorithms are crucial to build trust in this disruptive technology, but we also need the broader engagement of civil society in the values to be embedded in AI and the directions for future development.

This social engagement should be part of the effort to strengthen our resilience at all levels from local, to national and European, across institutions, industry and civil society. Developing local ecosystems of skills, computing, data and applications can foster the engagement of local communities, respond to their needs, harness local creativity and knowledge, and build a human-centred, diverse, and socially driven AI.

“We need
to act together
and act fast.”

We still know very little about how AI will impact the way we think, make decisions, relate to each other, and how it will affect our jobs. This uncertainty can be a source of concern but is also a sign of opportunity. **The future is not yet written. We can shape it based on our collective vision of what future we would like to have.** But we need to act together and act fast.



FOOTNOTES

- 1 We define players as research centres, academic institutions, and companies which have participated in one or more of the following economic activities: R&D processes, industrial production and marketing, specific AI-related services (see Box 5).
- 2 Technically, this diagram mainly refers to supervised learning.
- 3 Neural network models for computing were introduced in the 1950s and 60s inspired by neurobiology. Learning is modelled as a gradual change in the (computing) network connections in analogy to the synaptic modifications taking place in the biological neural network.
- 4 <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>
- 5 <https://www.reuters.com/article/us-usa-artificialintelligence/trump-administration-will-allow-ai-to-freely-develop-in-u-s-official-idUSKBN1IB30F>
- 6 <https://www.darpa.mil/news-events/2018-09-07>
- 7 <https://www.nasdaq.com/article/us-stock-market-is-biggest-most-expensive-in-world-but-us-economy-is-not-the-most-productive-cm942558>
- 8 <https://www.cNBC.com/2018/02/28/technology-now-makes-up-a-quarter-of-the-stock-market.html>
- 9 <https://www.statista.com/statistics/216657/market-capitalization-of-us-tech-and-internet-companies/>
- 10 https://www.whitehouse.gov/wp-content/uploads/2018/02/ap_18_research-fy2019.pdf
- 11 <https://www.statista.com/statistics/216573/worldwide-market-share-of-search-engines/>
- 12 <http://www.internetlivestats.com/google-search-statistics/>
- 13 <https://www.statista.com/statistics/264810/number-of-monthly-active-facebook-users-worldwide/>
- 14 <https://www.slideshare.net/AIFrontiers/jeff-dean-trends-and-developments-in-deep-learning-research>
- 15 <https://www.datacenterknowledge.com/google-alphabet/google-switching-self-driving-data-center-management-system>
- 16 <https://www.contino.io/insights/whos-using-aws>
- 17 <https://www.cifar.ca/ai/pan-canadian-artificial-intelligence-strategy>
- 18 In this version, the TES analysis of the AI ecosystem is based on data coming from more than 20 data sources, which include Arxiv, Asgard, CBInsights, Crunchbase, DowJones, EU CORDIS, EPO Patstat, Bureau Van Dijk Orbis, Statista, and other minor ones.
- 19 <http://ec.europa.eu/digital-single-market/en/news/eu-member-states-sign-cooperate-artificial-intelligence>
- 20 http://europa.eu/rapid/press-release_IP-18-4043_en.htm
- 21 On this subject, see also Chapter 11 in this report.
- 22 A unicorn is a privately-held start-up company valued at over US\$ 1 billion.
- 23 <https://www.statista.com/statistics/255778/number-of-active-wechat-messenger-accounts/>
- 24 <https://mp.weixin.qq.com/s/2X0IRqV01SKJSLBLwimL7A>
- 25 Floridi (2018) makes a distinction between hard and soft digital ethics, where hard precedes the formulation of new regulations or challenges the existing ones, and soft comes after the compliance by considering what could be done 'over and above the existing regulation'. Thus, soft ethics is a good approach where a normative framework, such as respect of human rights, is already in place.
- 26 This section has considered strategies and public policy documents released by EU institutions and public entities in France and the UK.
- 27 As codified in Art. 2 TEU: 'The Union is founded on the values of respect for human dignity, freedom, democracy, equality, the rule of law and respect for human rights, including the rights of persons belonging to minorities.

These values are common to the Member States in a society in which pluralism, non-discrimination, tolerance, justice, solidarity and equality between women and men prevail.' The European Charter of Fundamental Rights, recognises, among others, the right to human dignity, privacy, protection of personal data, freedom of expression and information, equality, no discrimination, the rights of the child and the elderly and persons with disabilities as well as the rights to consumer protection, good administration and a fair trial.

- 28 Note that even if Art. 22.2 GDPR recognises exceptions to this right, the data subject will always keep the right “to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision”.
- 29 <http://www.algoaware.eu/>
- 30 Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products, OJ L 210.
- 31 Among the problematic categories are the scope (reference to products and B2C relations) as well as the requirements upon the consumer to benefit from the strict liability regime (notably the burden of the proof concerning the defect of the product, the actual damage and the causal link between those).
- 32 Call for applications for the selection of members of the Expert Group on Liability and New Technologies.
- 33 For a detailed overview of the legal challenges concerning liability, see Staff Working Document on liability for emerging digital technologies (SWD(2018) 137) accompanying the Communication on Artificial intelligence for Europe COM(2018) 237; EC 2018b. Directly connected to the discussion on liability, it must also be noted that, in its Resolution, the EP invites the EC to consider a specific legal status for robots, which could lead to recognising legal personality for autonomous robots (EP 2017).
- 34 At the sectoral level, the EP has called on the Commission to publish a legislative proposal on access to in-vehicle data (EP 2018a). With the release of the Third Mobility Package, the Commission has announced a Recommendation that, among other things, will deal with ‘a data governance framework that enables data sharing, in line with the initiatives of the 2018 Data Package, and with data protection and privacy legislation.’
- 35 The proposal is now being discussed by the Commission, the Council and the Parliament. It must be noted that both the EP and the Council have proposed to enlarge the scope to cultural institutions and to create a new optional exception in addition to that proposed by the EC.
- 36 See Council of Europe, Explanatory report, Protocol amending the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (ETS No. 108), Ad hoc Committee on Data Protection (CAHDATA), 128th Session of the Committee of Ministers (Elsinore, Denmark, 17-18 May 2018). The Exploratory Report reads: “The notion of ‘identifiable’ refers not only to the individual’s civil or legal identity as such, but also to what may allow to ‘individualise’ or single out (and thus allow to treat differently) one person from others. This ‘individualisation’ could be done, for instance, by referring to him or her specifically, or to a device or a combination of devices (computer, mobile phone, camera, gaming devices, etc.) on the basis of an identification number, a pseudonym, biometric or genetic data, location data, an IP address, or other identifier. The use of a pseudonym or of any digital identifier/digital identity does not lead to anonymisation of the data as the data subject can still be identifiable or individualised. Pseudonymous data is thus to be considered as personal data and is covered by the provisions of the Convention. The quality of the pseudonymisation techniques applied should be duly taken into account when assessing the appropriateness of safeguards implemented to mitigate the risks to data subjects.”
- 37 https://www.epo.org/law-practice/legal-texts/html/guidelines2018/e/g_ii_3_3_1.htm.
- 38 See UK Government’s response to the Lords Select Committee AI report, June 2018.
- 39 Approaches designed to test the boundaries of computer programming language design, as a proof of concept.
- 40 <https://www.bloomberg.com/news/articles/2018-02-07/just-how-shallow-is-the-artificial-intelligence-talent-pool>
- 41 <https://www.elementsofai.com>
- 42 See, for example, Hoekstra and van Sluijs (2003).

- 43 Readers may also be interested in ‘HUMAIN’T’, an interdisciplinary JRC project aiming to understand the impact of machine intelligence on human behaviour, with a focus on cognitive and socio-emotional capabilities and decision-making (see <https://ec.europa.eu/jrc/communities/community/humaint>).
- 44 For a more comprehensive overview and assessment of the relevant research literature and an empirical investigation into changing occupational patterns in Europe, see the JRC’s forthcoming report on the future of work.
- 45 Although the scale of redistribution for economic growth to continue at a steady pace would have to be truly enormous, so enormous as to fully compensate the (almost) complete substitution of human labour by machines.
- 46 Singularity refers to the scenario of self-learning and self-programming machines starting a runaway spiral of improvements that overtake human intelligence with catastrophic consequences.
- 47 A general-purpose technology refers to a key enabling technology bringing major changes to society, such as electricity or IT.
- 48 The term cyber-physical systems refers to physical objects, often interconnected over digital networks, that are partly controlled by software algorithms and are capable of interacting with the physical world (e.g. autonomous vehicles).
- 49 In ISO standards, the ratio between the total data centre input power and the IT power is defined as power utilisation effectiveness (PUE). Typical data centres have a PUE of 2, while the most efficient ones have a PUE of 1.1
- 50 <http://www.iea.org/publications/freepublications/publication/DigitalizationandEnergy3.pdf>
- 51 <https://eta.lbl.gov/publications/united-states-data-center-energy>
- 52 This is a European initiative managed by the JRC, <https://ec.europa.eu/jrc/en/energy-efficiency/code-conduct/datacentres>
- 53 *Fixtures Marketing Ltd v. Oy Veikkaus Ab* (C-46/02, 9/11/2004), *Fixtures Marketing Ltd v. Svenska Spel Ab* (C-338/02, 9/11/2004) *British Horseracing Board Ltd v. William Hill* (C-203/02, 9/11/2004) *Fixtures Marketing Ltd v. OPAP* (C-444/02, 9/11/2004)
- 54 https://www.europeandataportal.eu/sites/default/files/edp_creating_value_through_open_data_0.pdf
- 55 See https://en.wikipedia.org/wiki/Data_sovereignty
- 56 This chapter is based on Duch-Brown, Martens and Mueller-Langer (2017). For more in-depth discussion on the issues raised please refer to that publication.
- 57 Examples include the Rome Declaration of 2017, the Reflection paper on Harnessing Globalisation, the Joint Communication on ‘A Strategic Approach to Resilience in the EU’s External Action’.
- 58 They are usually indicated as resilience indicators. The forthcoming report on Resilience refers to them as ‘expert based indicators’, and offers a more in-depth discussion.
- 59 Government policies were evaluated by using the average ranking across the 10 sub-aspects of the three main dimensions (readiness, implementation and impact, see [https://opendatabarometer.org/leadersedition/methodology/ for details](https://opendatabarometer.org/leadersedition/methodology/for%20details)). The education fields proxy the concept of STEM by considering natural sciences, mathematics and statistics, information and communication technologies, database and network design and administration, and software and applications development and analysis. The shares were averaged over the years 2013-16.
- 60 For each variable, a value of one indicates that the underlying measure is above the cross-country average, while a minus one means that it is below. The series are then stacked next to each other. Therefore, a value of -3 means being below average on all three variables.
- 61 Defined as having either successfully completed a tertiary-level education, or not formally qualified but employed in an S&T occupation where tertiary level qualifications are normally required.
- 62 <http://s3platform.jrc.ec.europa.eu/digital-innovation-hubs>

REFERENCES

- Acemoglu, D. and Autor, D., 2011. 'Skills, Tasks and Technologies: Implications for Employment and Earnings'. In *Handbook of Labor Economics*, 4:1043-1171. NBER Working Paper Series, Elsevier.
- Acemoglu, D. and Restrepo, R., 2016. *The Race Between Machine and Man: Implications of Technology for Growth, Factor Shares and Employment*. Working Paper 22252. National Bureau of Economic Research. <https://doi.org/10.3386/w22252>
- Acemoglu, D. and Restrepo, R., 2017. *Automation and Local Labor Markets*. <https://doi.org/10.3386/w23285>
- Acemoglu, D. and Restrepo, R., 2018b. 'The Race between Man and Machine: Implications of Technology for Growth, Factor Shares, and Employment.' *American Economic Review*; 108.6, pp. 1488-1542. <https://www.aeaweb.org/articles?id=10.1257/aer.20160696>
- Acemoglu, D. and Restrepo, R., 2018c. *Automation and new tasks: the implications of the task content of technology for labour demand*, mimeo, NBER, August 2018.
- Acemoglu, D. and Restrepo, P., 2018a. 'Artificial Intelligence, Automation and Work'. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3098384>
- Aghion, P., Jones, B. J. and Jones, C., 2017. 'Artificial Intelligence and Economic Growth'. *National Bureau of Economic Research*, no. w23928.
- AI Now 2017 Report. https://ainowinstitute.org/AI_Now_2017_Report.pdf
- AI Now 2018. Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability. <https://ainowinstitute.org/aiareport2018.pdf>
- Aldewereld, H. M., Dignum, V. and Tan, Y., 2015. 'Design for values in software development'. In van den Hoven J., Vermaas P. E. and van de Poel, I. (Eds.) *Handbook of Ethics, Values, and Technological Design, sources, theory, values and application domains*, pp. 831-846. Berlin: Springer.
- Alessi, L., Benczur, P., Campolongo, F. et al. 2018. *The resilience of EU Member States to the financial and economic crisis What are the characteristics of resilient behaviour?* JRC Science for Policy Report http://publications.jrc.ec.europa.eu/repository/bitstream/JRC111606/jrc111606_resilience_crisis_pilot_withidentifiers.pdf
- Alexander, B., 2018. 'Why robots helped Donald Trump win'. *MIT Technology Review* <https://www.technologyreview.com/s/611422/why-robots-helped-donald-trump-win/>
- Amodei, D., Olah, C., Steinhardt, J. et al. 2016. *Concrete Problems in AI Safety*. <https://arxiv.org/abs/1606.06565>
- Arntz, M., Gregory, T., and Zierahn, U. 2016. 'The Risk of Automation for Jobs in OECD Countries: A Comparative Analysis'. *OECD Social, Employment and Migration Working Papers 2* (189): 47-54. <https://doi.org/10.1787/5jlz9h56dvq7-en>
- Athey, S. (2017). 'The impact of machine learning on economics', *Economics of Artificial Intelligence*. University of Chicago Press.
- Autor, D. H., Levy, F. and Murnane, R. J. 2003. 'The Skill Content of Recent Technological Change: An Empirical Exploration'. *The Quarterly Journal of Economics*. <https://doi.org/10.1162/003355303322552801>
- Autor, D., and Salomons, A. 2017. 'Robocalypse Now—Does Productivity Growth Threaten Employment?' *ECB Forum on Central Banking* 2142 (June 2017): 1-74. <https://conference.nber.org/confer/2017/AI17/Autor.pdf>
- Baccala, M., Curran, C., Garrett, D., Likens, S., Rao, A., Ruggles, A., and Shehab, M. 2018. *2018 AI predictions: 8 insights to shape business strategy*. PwC. <https://www.pwc.com/us/en/advisory-services/assets/ai-predictions-2018-report.pdf>
- Bakardjieva, M. and Gehl, R. 2017. 'Critical Approaches to Communication Technology: The Past Five Years'. *Annals of the International Communication Association* 41 (3-4), Online first. <https://doi.org/10.1080/23808985.2017.1374201>
- Baldwin, R. and Vihriala, V. 2017. 'Globalisation may soon accelerate again – time to get domestic policies right' <https://voxeu.org/article/globalisation-may-soon-accelerate-again-time-get-domestic-policies-right>

- Barton, D., Woetzel, J., Seong, J., and Tian, Q. 2017. *Artificial Intelligence: Implications for China*. McKinsey Global Institute (Vol. Discussion). Retrieved from www.mckinsey.com/mgi.
- Baslandze, S. 2016. 'The Role of the IT Revolution in Knowledge Diffusion, Innovation and Reallocation'. In 2016 *Meeting Papers* (No. 1509). Society for Economic Dynamics.
- Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., and Tanaka, F. 2018. 'Social robots for education: A review'. *Science Robotics*, 3(21), eaat5954.
- Bergmann, D. and Morris, S. 2017. *Information Design: a Unified Perspective*. Cowles Foundation Discussion Paper. <https://cowles.yale.edu/sites/default/files/files/pub/d20/d2075.pdf>
- Biggio, B., Nelson, B., and Laskov, P. 2012. *Poisoning Attacks against Support Vector Machines*. <https://arxiv.org/abs/1206.6389>
- Bijker, W. et al. 1987 (Eds.) *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge: MIT Press.
- Borgmann, A. 1987. *Technology and the Character of Contemporary Life: A Philosophical Inquiry*. Chicago: University of Chicago Press.
- Botsman, R. 2017. 'Big data meets Big Brother as China moves to rate its citizens'. *Wired* October 31. Retrieved from <https://www.wired.co.uk/article/chinese-government-social-credit-score-privacy-invasion>
- Boucher, P., Nascimento, S., Vesnic-Alujevic, L. and Guimaraes Pereira, A. 2014. *Ethics Dialogues: Experiencing Ethics Through 'Things' – Open IoT, Civil Drones and Wearable Sensors*. Luxembourg: Publications Office of the European Union.
- Brundage, M., Clark, J., Allen, G. C., Flynn, C., Farquhar, S., Crootof, R., and Bryson, J. 2018. 'The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation'. Information Society Project. Future of Human Institute. Retrieved from <https://www.repository.cam.ac.uk/bitstream/handle/1810/275332/1802.07228.pdf?sequence=1>
- Brynjolfsson, E., Mitchell, T., and Rock, D. 2018. 'What Can Machines Learn, and What Does It Mean for Occupations and the Economy?'. *AEA Papers and Proceedings*. Vol. 108. <https://www.aeaweb.org/articles?id=10.1257/pandp.20181019>
- Brynjolfsson, E., Rock, D. and Syverson, C. (2018) *The productivity J-curve: how intangibles complement general purpose technologies*, NBER working paper no. 25148, October 2018.
- Brynjolfsson, E., Rock, D., Syverson, C., et al. 2017. *Artificial Intelligence and the Modern Productivity Paradox: A Clash of Expectations and Statistics*. <http://www.nber.org/papers/w24001>
- Cath, C., Wachter, S., Mittelstadt, B. et al. 2018. 'Artificial Intelligence and the 'Good Society': the US, EU, and UK approach'. *Science and Engineering Ethics*, 24 (2): 505-528.
- Charisi, V., Liem, C., Gómez, E. 2018. 'Novelty-based cognitive processes in unstructured music-making play settings in early childhood'. *Proceedings 8th Joint IEEE International Conference of Development and Learning and Epigenetic Robotics ICDL-EpiRob2018*, pp. 218-223.
- Chiacchio, F., Petropoulos G., and Pichler D.. 2018. *The Impact of Industrial Robots on EU Employment and Wages: A Local Labour Market Approach*. Bruegel Working Papers. Bruegel.
- Chollet, F. 2017. *Deep Learning with Python*. Manning
- Chouldechova, A. 2017. 'Fair prediction with disparate impact: A study of bias in recidivism prediction instruments'. *Big Data*, 5 (2): 153-163.
- Cockburn, I. M., Henderson, R., and Stern, S.. 2018. *The Impact of Artificial Intelligence on Innovation*. 24449. NBER Working Paper Series.
- Coeckelbergh, M. 2018. 'Technology and the good society: A polemical essay on social ontology, political principles, and responsibility for technology'. *Technology in Society* (52), 4-9.
- Colombo, F., Murru, M. F., Tosoni, S. 2017. 'The post-intermediation of truth. Newsmaking from Media Companies to Platform'. *Comunicazioni sociali*, 3: 448-461.
- Commission Nationale Informatique et Liberte (CNIL) 2017. *How can humans keep the upper hand? The ethical matters raised by algorithms and artificial intelligence*. Report on the public debate led by the French data protection authority (CNIL) as part of the ethical discussion assignment set by the digital republic bill. Retrieved from https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_ai_gb_web.pdf

- Council of Europe (CoE) 2017. *Algorithms and Human Rights*. Study on the human rights dimensions of automated data processing techniques and possible regulatory implications. Retrieved from <https://rm.coe.int/study-hr-dimension-of-automated-data-processing-incl-algorithms/168075b94a>
- Council of Europe (CoE) 2017. *Technological convergence, artificial intelligence and human rights*. Retrieved from <http://www.assembly.coe.int/nw/xml/XRef/Xref-DocDetails-EN.asp?FileID=24236&lang=EN>
- Craglia, M., Hradec, J. and Troussard, X. 2019. 'Big Data, AI, and Public Policy' in Sucha, V., Mair, D., Sienkiewicz, M. et al. (Eds.) 2019 *Science and Evidence in the Policy Ecosystem Handbook*, Elsevier (forthcoming).
- Crawford, M., Khoshgoftaar, T. M., Prusa, D. J. et al. 2015. 'Survey of review spam detection using machine learning techniques'. *Journal of Big Data* (Springer Nature) 2. doi:10.1186/s40537-015-0029-9.
- Danaher, J. 2017. *Algocracy as Hypemudging: A New Way to Understand the Threat of Algocracy*. Retrieved from <https://ieet.org/index.php/IEET2/more/Danaher20170117>
- Darvas, Z., and Wolff, G. B. 2016. *An Anatomy of Inclusive Growth in Europe*. Bruegel Blueprint 26, October 2016.
- Dauth, W., Findeisen, S., Suedekum, J., and Woessner, N. 2018. *Adjusting to Robots: Worker-Level Evidence*. Federal Reserve Bank of Minneapolis, Working Paper no. 13. <https://www.minneapolisfed.org/institute/institute-working-papers/adjusting-to-robots-worker-level-evidence>
- De Groeve, T., Karmen, P., and Vernaccini, L. 2014. *Index for Risk Management - InfoRM: Concept and Methodology*, Version 2014. European Commission, Joint Research Centre.
- De Prato, G., Lopez Cobo, M. et al. 2018. *The AI Techno-Economic Segment Analysis. Selected Indicators*. :European Commission, Joint Research Centre, Seville, Spain.
- Dearn, J. Patterson, D. and Young, C. 2018. 'A New Golden Age in Computer Architecture: Empowering the Machine-Learning Revolution'. *IEEE Micro* 38(2): 21-29.
- Delage, F. (2017). 'China: diplomacia económica, consecuencias geopolíticas'. *Cuadernos de Estrategia*, (187), 55-92.
- Delforge, A., Gerard, L., 'Notre vie privée est-elle réellement mise en danger par les robots ? : Étude des risques et analyse des solutions apportées par le GDPR', in *L'intelligence artificielle et le droit*, Larcier, Brussels: 2017. 143-188.
- Deloitte 2018. *Study on emerging issues of data ownership, interoperability, re-usability and access to data, and liability*, prepared for the European Commission, available at: http://ec.europa.eu/newsroom/dae/document.cfm?doc_id=51486
- Deming, D. J. 2017. 'The Growing Importance of Social Skills in the Labor Market.' *Quarterly Journal of Economics*, no. January: 1593-1640. <https://doi.org/10.1093/qje/qjx022>
- Dix-Carneiro, R. 2014. 'Trade Liberalization and Labor Market Dynamics'. *Econometrica*, 82(3), 825-885.
- Drexler, J., 2017a. 'Designing Competitive Markets for Industrial Data'– *Between Propertisation and Access*, 8, JIPITEC, E 2017, 257.
- Drexler, J., 'On the Future EU Legal Framework for the Digital Economy: A competition-based Response to the Ownership-Access Debate', in Lohsse, S., Schulze, R. and Staudenmayer, D., (eds.), *Trading Data in the Digital Economy: Legal Concepts and Tools*, Noms, Baden-Baden, 2017, p.221-244
- Drigas, A. and Ioannidou, R. E. 2013. 'Special Education and ICTs'. *International Journal of Emerging Technologies in Learning (IJET)* 8 (2): 41-47.
- Duch-Brown, N., Martens, B., Muller-Langer, F., 2017. *The economics of ownership, access and trade in digital data*, JRC Technical report, JRC Digital Economy Working Paper 2017-01, European Commission, Joint Research Centre, Seville, Spain.
- Dutton, T. 2018. AI strategies. <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>
- Economist*, 1995. 'The Death of Distance, survey of Telecommunications', *Economist*, September, vol. 336 no. 7934.
- Edwards, L. and Veale, M. 2017. 'Slave to the Algorithm? Why a 'Right to an Explanation' is Probably not the Remedy You are Looking For'. *Duke Law & Technology Review*, 16 (1): 18-84.
- European Commission (EC), 1996. Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases. <https://eur-lex.europa.eu/lexUriServ/LexUriServ.do?uri=CELEX:31996L0009:EN:HTML>
- European Commission (EC), 2003. Directive 2003/98/EC of the European Parliament and of the Council on the re-use

of public sector information of 31/12/2003, revised by Directive 2013/37/EC of 17/7/2013. <https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information>

European Commission (EC), 2016a Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119.

European Commission (EC), 2016b, Proposal for a Directive of the European Parliament and of the Council on copyright in the Digital Single Market, COM(2016)593 final, Brussels.

European Commission (EC), 2016c. Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure, OJ L 157.

European Commission (EC), 2017a, Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 'Building A European Data Economy' COM(2017) 9 final Brussels.

European Commission (EC), 2017b. Joint Communication to the European Parliament and the Council: Resilience, Deterrence and Defence: Building strong cybersecurity for the EU. Tech. rep., EU. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017JC0450&from=ES>

<https://ec.europa.eu/digital-single-market/en/news/staff-working-document-and-executive-summary-evaluation-directive-969ec-legal-protection>

European Commission (EC), 2018a. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Artificial Intelligence for Europe COM(2018) 237 final Brussels.

European Commission (EC), 2018b. AI Factsheet <https://ec.europa.eu/digital-single-market/en/news/factsheet-artificial-intelligence-europe>

European Commission (EC), 2018c. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, 'Towards a common European data space' COM(2018) 232 final Brussels.

European Commission (EC), 2018d. Proposal for a Directive of the European Parliament and of the Council on the re-use of public sector information (recast) COM/2018/234 final, Brussels.

European Commission (EC), 2018f. Commission Staff Working Document Guidance on sharing private sector data in the European data economy Accompanying the document Communication from the Commission to the European Parliament, the Council, the European economic and social Committee and the Committee of the Regions 'Towards a common European data space', SWD(2018)125 final, Brussels.

European Commission (EC), 2018g Commission Staff Working Document Evaluation of Directive 96/9/EC on the legal protection of databases, SWD(2018) 146 final, Brussels.

European Commission (EC) 2018f: Council recommendation of 22 May 2018 on key competences for lifelong learning, OJ C 189, 4.6.2018, p. 1–13 <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32018H0604%2801%29>

European Commission (EC), 2018i. Commission Recommendation (EU) 2018/790 of 25 April 2018 on access to and preservation of scientific information, Official Journal of the European Union, L 134/1.

European Data Protection Supervisor (EDPS). 2018. *Towards a digital ethics, Report from EDPS Ethics Advisory Group*. Retrieved from https://edps.europa.eu/sites/edp/files/publication/18-01-25_eag_report_en.pdf

European Economic and Social Committee (EESC) 2016. EESC Opinion on Artificial Intelligence. <https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/artificial-intelligence>

European Group on Ethics (EGE). 2018. 'Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems. European Group on Ethics in Science and New Technologies.' Brussels: European Commission. https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf

European Parliament (EP) 2016. European Civil Law Rules in Robotics. [http://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU\(2016\)571379_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf)

- European Parliament (EP) 2017. Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)).
- Faliagka, E., et al. 2012. 'Application of machine learning algorithms to an online recruitment system'. In *Proceedings International Conference on Internet and Web Applications and Services*.
- Fearon, J. 1999. What is identity (as we now use the word)? Retrieved from <https://web.stanford.edu/group/fearon-research/cgi-bin/wordpress/wp-content/uploads/2013/10/What-is-Identity-as-we-now-use-the-word-.pdf>
- Feenberg, A. 1991. *Critical Theory of Technology*. New York: Oxford University Press.
- Feenberg, A. 1999. *Questioning Technology*. New York: Routledge.
- Feijoo, C. et al. 2019. *The Industrial Innovation Ecosystem of Artificial Intelligence in China: Current status and prospects*, European Commission, Joint Research Centre, Seville, Spain (forthcoming).
- Fernández-Macías, E. 2012. 'Job polarization in Europe? Changes in the employment structure and job quality', 1995-2007. *Work and Occupations*, 39(2), 157-182.
- Fernández-Macías, E., and Hurley, J. 2016. Routine-biased technical change and job polarization in Europe. *Socio-Economic Review*, 15(3), 563-585.
- Fernández-Macías, E., and Bisello, M. 2017. *Measuring The Content and Methods of Work: A Comprehensive Task Framework*. <http://ibs.org.pl/app/uploads/2017/10/D1.-Martina-Bisello.pdf>
- Fernández-Macías, E., Gómez, E., Hernández-Orallo, J., Loe, B. S., Martens, B., Martínez-Plumed, F., and Tolan, S. 2018. 'A multidisciplinary task-based perspective for evaluating the impact of AI autonomy and generality on the future of work'. *arXiv preprint arXiv:1807.02416*.
- Finnish Government (2017), Publications of the Ministry of Economic Affairs and Employment Ministry, 47/2017, *Finland's Age of Artificial Intelligence Turning Finland into a leading country in the application of artificial intelligence Objective and recommendations for measures*. http://julkaisut.valtioneuvosto.fi/bitstream/handle/10024/160391/TEMrap_47_2017_verkkojulkaisu.pdf?sequence=1&isAllowed=y
- Floridi, L. (ed.) (2015). *The Onlife Manifesto: Being Human in a Hyperconnected Era*. London: Springer Open.
- Floridi, L. (2018). *Soft ethics and the governance of the digital*. *Philosophy and Technology*, 31: 1. <https://doi.org/10.1007/s13347-018-0303-9>
- France Intelligence Artificielle (FIA) 2017. *Rapport de synthèse*. https://www.economie.gouv.fr/files/files/PDF/2017/Rapport_synthese_France_IA_.pdf
- France Intelligence Artificielle (FIA) 2017. *Rapport de synthèse. Groupe de travail*. https://www.economie.gouv.fr/files/files/PDF/2017/Conclusions_Groupes_Travail_France_IA.pdf
- Frey, C. B., and Osborne, M. A. 2013. *The Future of Employment. Working Paper*. Oxford: Oxford Martin Programme on Technology and Employment. <https://www.oxfordmartin.ox.ac.uk/downloads/academic/future-of-employment.pdf>
- Frey, C. B., and Osborne, M. A. 2017. 'The Future of Employment: How Susceptible Are Jobs to Computerisation?' *Technological Forecasting and Social Change* 114: 254-80. <https://doi.org/10.1016/j.techfore.2016.08.019>
- Furman, J., and Seamans, R. 2018. *AI and the Economy*. NBER Working Paper no. 24689.
- Garfinkel, S., Matthews, J., Shapiro, S. S., and Smith, J. M. 2017. 'Toward algorithmic transparency and accountability'. *Communications of the ACM*, 60(9), 5-5. <https://doi.org/10.1145/3125780>
- Geiger, C., Frosio, G. and Bulayenko, O., 2018. 'Crafting a text and data mining exception for machine learning and big data in the Digital Single Market' in Seuba, X., Geiger, C. and Penin, J., (Eds.) 2018. *Intellectual Property and Digital Trade in the Age of Artificial Intelligence and Big Data*, p.95-112, available at: <https://www.ictsd.org/themes/innovation-and-ip/research/intellectual-property-and-digital-trade-in-the-age-of-artificial>
- Giles, M., 2018. 'It's time to rein in the data barons'. *MIT Technology Review*. <https://www.technologyreview.com/s/611425/its-time-to-rein-in-the-data-barons/>
- Goldfarb, A., Trefer, D., 2018a 'How artificial intelligence impacts international trade?'. *World Trade report 2018, The future of world trade: how digital technologies are transforming global commerce*, p. 140.
- Goldfarb, A., and Trefer, D. 2018b. 'AI and International Trade' in Agrawal, A., Gans, J. S., and Goldfarb, A. (Eds.) 2018. *The economics of artificial intelligence*. University of Chicago Press.

- Gómez, E. (ed.) 2018. 'Assessing the impact of machine intelligence on human behaviour: an interdisciplinary endeavour', 1st HUMAINT workshop, Barcelona. https://ec.europa.eu/jrc/communities/file/3071/download?token=_RDTpI0f
- Gómez, E., Blaauw, M., Bonada, J. et al. 2018. *Deep Learning for Singing Processing: Achievements, Challenges and Impact on Singers and Listeners*. <https://arxiv.org/abs/1807.03046>
- Goolsbee, A. 2018. *Public Policy in an AI Economy* (No. w24653). National Bureau of Economic Research.
- Goos, M., Manning, A. and Salomons, A. 2014. 'Explaining job polarization: routine-biased technological change and offshoring'. *American Economic Review*, 104(8): 2509-2526.
- Graesser, A. C., P. Chipman, B. Haynes C., and Olney, A. 2005. 'AutoTutor: An Intelligent Tutoring System with Mixed-Initiative Dialogue'. *IEEE Transactions on Education* 48(4): 612-18. <https://doi.org/10.1109/TE.2005.856149>
- Greenpeace 2012. *How Clean is Your Cloud?* <https://www.greenpeace.org/archive-international/Global/international/publications/climate/2012/iCoal/HowCleanisYourCloud.pdf>
- Gregory, T, Salomons, A. and Zierahn, U. (2018) *Racing with or against the machine, evidence from Europe*, CESifo working paper no. 7247, September 2018.
- Guadamuz, A. 2017. 'Do androids dream of electric copyright? Comparative analysis of originality in artificial intelligence generated works', *Intellectual Property Quarterly*, 2. <https://ssrn.com/abstract=2981304>
- Guillén, M. and Reddy, S. 2018. 'We know ethics should inform AI. But which ethics?' *World Economic Forum note* <https://www.weforum.org/agenda/2018/07/we-know-ethics-should-inform-ai-but-which-ethics-robotics/>
- Hall, W., and Pesenti, J. 2017. *Growing the Artificial Intelligence Industry in the UK*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/652097/Growing_the_artificial_intelligence_industry_in_the_UK.pdf
- Harari, Y. N., 2015. *Homo Deus*. Harvil Secker.
- Higgs, E., Light, A. and Strong, D. (Eds.) 2000. *Technology and the Good Life?* Chicago: The University of Chicago Press.
- Hoekstra, H. A., and Van Sluijs, E. 2003. *Managing Competences: Implementing Human Resource Management*. Assen: Koninklijke Van Gorcum.
- Homok, C., and Koren, M. 2017. 'Winners and losers of globalisation: Sixteen challenges for measurement and theory'. *Economics without borders: economic research for European policy challenges*. Cambridge University Press, Cambridge, UK, 230-273. <https://www.cambridge.org/core/books/economics-without-borders/winners-and-losers-of-globalization-sixteen-challenges-for-measurement-and-theory/E5BFE7843A887EB875AB8AC8CC179C5A>
- Huang, L., Joseph, A., Nelson, B. et al. 2011. 'Adversarial Machine Learning'. In *Proceedings of 4th ACM Workshop on Artificial Intelligence and Security*, October 2011, pp. 43-58. https://people.eecs.berkeley.edu/~tygar/papers/SML2/Adversarial_AISEC.pdf
- Huang, S., Papernot, N., Goodfellow, I. et al. 2017. *Adversarial Attacks on Neural Network Policies*. <https://arxiv.org/abs/1702.02284>
- Hugenholtz, B. 2018. 'Against Data Property' in Hanns, U., Drahos, P. and Ghidini, G. (Eds.) 2018. *Kritika, Essays on Intellectual Property*, Vol. III, Elgar. <https://www.e-elgar.com/shop/kritika-essays-on-intellectual-property-15154>
Earlier version at: https://www.ivir.nl/publicaties/download/Data_property_Muenster.pdf
- IEEE 2018. *Ethically Aligned Design*. Retrieved from http://standards.ieee.org/develop/indconn/ec/autonomous_systems.html
- IEEE. 2017. *Special Issue: Hacking without Humans*. IEEE Security & Privacy. <https://publications.computer.org/security-and-privacy/tag/hacking-without-humans/>
- Institute for Energy Efficiency (IEE) 2012. *Scalable, Energy-Efficient Data Centers and Clouds*. https://iee.ucsb.edu/sites/iee.ucsb.edu/files/data_center_report.pdf
- Jakobsson, J. 2017. 'Nya miljoner ska ta deras dyslexi-startup till USA.' Di Digital, March 29, 2017. <https://digital.di.se/artikel/nya-miljoner-ska-ta-deras-dyslexi-startup-till-usa>. For English version, see <http://www.lexplore.com/>
- Jasanoff, S. 2013. Technologies of humility: Citizen participation in governing science. *Minerva*, 41(3): 223-244.
- JIIP, TNO & Technopolis 2018. *Study in support of the evaluation of Directive 96/9/EC on the legal protection of data bases*, European Commission. <https://publications.europa.eu/en/publication-detail/-/publication/f10f8ee0-597e-11e8-ab41-01aa75ed71a1/language-en>

- Kahn, J. 2018. 'Just How Shallow Is the Artificial Intelligence Talent Pool?' *Bloomberg.Com*, February 7, 2018. <https://www.bloomberg.com/news/articles/2018-02-07/just-how-shallow-is-the-artificial-intelligence-talent-pool>
- Kambatla, K, Kollias, G, Kumar, V., Grama, A. 2014. 'Trends in big data analytics', *Journal of Parallel and Distributed Computing*, 74. (7) 2561-2573.
- Kania, E. 2017. 'China Is On a Whole-of-Nation Push for AI. The US Must Match It.' *Cyber Warfare*, 10-12. <https://doi.org/10.1016/B978-0-12-803843-7.00083-1>
- Karabarbounis, L. and Neiman, B. 2013. *The Global Decline of the Labor Share*, NBER Working Papers 19136, National Bureau of Economic Research, Inc. <https://ideas.repec.org/p/nbr/nberwo/19136.html>
- Karlgard, R. and Malone, M. 1995. 'City vs. country: Tom Peters & George Gilder debate the impact of technology on location'. *Forbes ASAP Technology Issue*, February, 27, pp. 56-61.
- Kerber, W. 2016. *A new intellectual property right for data: an economic analysis*. GRUR 989.
- Kerber, W. 2016. 'Governance of Data: Exclusive Property vs. Access'. in *International Review of Intellectual Property and Competition Law*, 47: 7, 759-762, available at: <https://link.springer.com/article/10.1007/s40319-016-0517-2>
- Kerber, W. and Frank, J.S. 2017. *Data Governance regimes in the Digital Economy: the Example of Connected Cars*. https://www.tilburguniversity.edu/upload/5c3e205b-fc0e-42ef-a1e8-8d72c0e16025_Kerber_Frank_01102017_Data%20Governance%20and%20Connected%20Cars.pdf
- Kilbertus, N., Carulla, M. R., Parascandolo, et al. 2017. *Avoiding Discrimination through Causal Reasoning*, NIPS <https://papers.nips.cc/paper/6668-avoiding-discrimination-through-causal-reasoning.pdf>
- Knight, W. 2017. 'Progress in AI isn't as Impressive as You Might Think', *MIT Technology Review* https://www.technologyreview.com/s/609611/progress-in-ai-isnt-as-impressive-as-you-might-think/?utm_source=MIT+Technology+Review&utm_campaign=ab120fee85-weekly_roundup_2017-12-07_edit&utm_medium=mail&utm_term=0_997ed6f472-ab120fee85-156491921&goal=0_997ed6f472-ab120fee85-156491921&mc_cid=ab120fee85&mc_eid=dafbd41dd1
- Knight, W. 2018. 'Inside the Chinese lab that plans to rewire the world with AI'. *MIT Technology Review* <https://www.technologyreview.com/s/610219/inside-the-chinese-lab-that-plans-to-rewire-the-world-with-ai/>
- Korff, D. Wagner, B. and Powles, J. et al. 2017. *Boundaries of Law: Exploring Transparency, Accountability, and Oversight of Government Surveillance Regimes*. University of Cambridge Faculty of Law Research Paper no. 16/2017. Retrieved from <https://ssrn.com/abstract=2894490>
- Korinek, A., and Stiglitz, J. 2017. *Artificial Intelligence and Its Implications for Income Distribution and Unemployment*. <https://doi.org/10.3386/w24174>.
- Lee, K. F. 2017. 'The real threat of AI'. *NY Times* <https://www.nytimes.com/2017/06/24/opinion/sunday/artificial-intelligence-economic-inequality.html>
- Manca, A. R., Benczur, P. and Giovannini, E. 2017. *Building a Scientific Narrative Towards a More Resilient EU society*. JRC Science for Policy Report http://publications.jrc.ec.europa.eu/repository/bitstream/JRC106265/jrc106265_100417_resilience_scienceforpolicyreport.pdf
- Markham, A., Tiidenberg, K. and Herman, A. 2018. 'Ethics as Methods: Doing Ethics in the Era of Big Data Research – Introduction'. *Social Media and Society*, July-September 2018: 1-9.
- Marquart, S. 2016. *South Korean Government Announces Nearly \$1 Billion in AI funding*. <https://futurism.com/south-korean-government-announces-nearly-1-billion-ai-funding/>
- Martens, B. 2016. *An economic policy perspective on online platforms*. JRC Digital Economy working paper 2016-05. Seville: JRC.
- Martens, B. and Muller-Langer, F. 2018. *Access to car data and competition in aftersales services markets*, JRC digital economy working paper, European Commission, Joint Research Centre, Seville, Spain (forthcoming)
- Mayer, M. 2009. *Innovation at Google: The physics of data*, available at <https://huffduffer.com/marshallkirkpatrick/10340>
- McKinsey Global Institute (MGI). 2018. *Notes from the AI frontier: Modeling the impact of AI in the world economy* <https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-modeling-the-impact-of-ai-on-the-world-economy>

- Meidan, Y., Bohadana, M., Mathov, Y. et al. 2018. 'N-BaloT: Network-based Detection of IoT Botnet Attacks Using Deep Autoencoders'. *IEEE Pervasive Computing*, July/Sep 2018.
- Mezzanotte, F. 2017. 'Access to Data: the Role of Consent and the Licensing Scheme', in Lohsse, S., Schulze, R. and Staudenmayer, D. (Eds.) 2017. *Trading Data in the Digital Economy: Legal Concepts and Tools*, NOMOS: p.159-188.
- Michaux, B. 2018. 'Singularité technologique, singularité humaine et droit d'auteur', in De Terwangne, C., Degrave, E. and Dusollier, S. (Eds.) 2018. *Laws, Norms and Freedoms in Cyberspace / Droits, norms et libertés dans le cybermonde*, Liber Amicorum Yves Poulet, Larcier, p.401-416.
- Miller, C., Ohrvik-Stott, J., Coldicutt R., 2018. *Regulating for Responsible Technology: Capacity, Evidence and Redress: a new system for a fairer future*. London: Doteveryone. <https://doteveryone.org.uk/project/regulating-for-responsible-technology>
- Mission Villani sur l'intelligence artificielle. 2018. *For a meaningful artificial intelligence. Towards a French and European Strategy*. https://www.aiforhumanity.fr/pdfs/MissionVillani_Report_ENG-VF.pdf
- Mitcham, C. 1980. 'Philosophy of Technology'. In Durbin, P.T. (ed.) 1980. *A Guide to the Culture of Science, Technology and Medicine*. The Free Press.
- Montreal Declaration 2017. *For a Responsible Development of Artificial Intelligence*. Retrieved from <https://www.montrealdeclaration-responsibleai.com/the-declaration>
- Morikawa, M. 2017. *Assessing the impact of AI and robotics on job expectations using Japanese survey data* <https://voxeu.org/article/who-fears-losing-their-job-ai-and-robots-japanese-survey-data>
- Mubayi, P., Cheng, E., Terry, H. P., Tilton, A., and Hou, T. 2017. *China's Rise in Artificial Intelligence*. Goldman Sachs. Retrieved from <https://www.cnbc.com/2017/09/01/goldman-says-china-has-talent-data-and-infrastructure-to-embrace-ai.html>
- Nascimento, S., Pólora, A., Paio, A., Oliveira, S. et al. 2016. 'Sustainable Technologies and Transdisciplinary Futures: From Collaborative Design to Digital Fabrication', *Science as Culture*, Vol. 25 No 4, pp. 520-537.
- National Science and Technology Council (NSTC) 2016c. *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
- National Science and Technology Council (NSTC). 2016b. *Artificial intelligence, automation and the economy*. <https://www.whitehouse.gov/sites/whitehouse.gov/files/images/EMBARGOED%20AI%20Economy%20Report.pdf>
- National Science and Technology Council, (NSTC). 2016a. *Preparing for the future of artificial intelligence*, Executive Office of the President, (2016a). <https://info.publicintelligence.net/WhiteHouse-ArtificialIntelligencePreparations.pdf>
- Nedelkoska, L. and Quintini, G. 2018. *Automation, Skills Use and Training*. OECD Social, Employment and Migration Working Papers. Paris: OECD.
- Nkambou, R., Azevedo, R. and Vassileva, J. (Eds.) 2018. *Intelligent Tutoring Systems: 14th International Conference, ITS 2018, Montreal, QC, Canada, June 11–15, 2018, Proceedings*. Programming and Software Engineering. Springer International Publishing.
- O'Neil, C. 2016. *Weapons of math destruction: How big data increases inequality and threatens democracy*. New York: Crown Publishers.
- OECD 2018. *Job Creation and Local Economic Development: Preparing for the Future of Work*. OECD Publishing, Paris, <https://doi.org/10.1787/9789264305342-en>
- Office parlementaire d'évaluation des choix scientifiques et technologiques (OPECST) 2017. *Toward a Controlled, Useful and Demystified Artificial Intelligence*. <http://www.senat.fr/rap/r16-464-1/r16-464-1-syn-en.pdf>
- Opher, A. , Chou, A., Onda, A. et al. 2016. *The Rise of the Data Economy: Driving Value through Internet of Things Data Monetization*. IBM Corporation. <https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=WWW12367USEN>
- Osaba, O. and Welser, W. 2017. *The Risks of Artificial Intelligence to Security and the Future of Work*. Rand Corporation <https://www.rand.org/pubs/perspectives/PE237.html>
- Osborne Clarke LLP 2016. *Legal study on Ownership and Access to Data*, European Commission. <https://publications.europa.eu/en/publication-detail/-/publication/d0bec895-b603-11e6-9e3c-01aa75ed71a1/language-en>

- Papernot, N., McDaniel, P., Jha, S. et al. 2016. The Limitations of Deep Learning in Adversarial Settings. *Proc. IEEE European Symp. Security and Privacy (EuroSP)*. 372–387. doi:10.1109/EuroSP.2016.36.
- Pau, J., Bajer, J. and Houston, N. 2017. *Artificial intelligence in Asia: Preparedness and Resilience*, Asia Business Council http://www.asiabusinesscouncil.org/Research17_AI.html
- Paxon, V. 1998. 'Bro: A System for Detecting Network Intruders in Real-Time'. *Computer Networks*. https://www.usenix.org/legacy/publications/library/proceedings/sec98/full_papers/paxson/paxson.pdf
- Porayska-Pomsta, K. 2015. 'AI in Education as a Methodology for Enabling Educational Evidence-Based Practice'. In *Seventeenth International Conference on Artificial Intelligence in Education (AIED 2015)*, 52–61. Madrid.
- Pouillet, Y., 2018a. 'Is the general data protection regulation the solution?', *Computer Law & Security Review: The International Journal of Technology Law and Practice* (forthcoming).
- Pouillet, Y., 2018b. 'Sharing Data: Some critical reflections about a Policy asserted by the EU Commission'. Presentation delivered at the Conference Data sharing and re-use, organised by Crids and Cullen International and hosted in Brussels, on 3rd October 2018 (to be published).
- Powles, J. and Hodson, H. 2017. 'Google DeepMind and healthcare in an age of algorithms'. *Health and Technology*, 7 (4): 351–367.
- Prakash, S., 2017. *The Time Has Come to Use Artificial Intelligence In Governance* <https://towardsdatascience.com/the-time-has-come-to-use-artificial-intelligence-in-governance-9782257e265b>
- Radford, B. J., Apolonio, L., Trias, A. J. and Simpson, J. A. 2018. *Network Traffic Anomaly Detection Using Recurrent Neural Networks*. <https://arxiv.org/abs/1803.10769>
- Ramalho, A., 2017a. 'Data Producer's Right: Power, Perils & Pitfalls', *Better Regulation for Copyright* presentation at the European Parliament, (Brussels: September 6, 2017), p.51-58, available at: https://juliareda.eu/wp-content/uploads/2017/09/2017-09-06_Better-Regulation-for-Copyright-Academics-meet-Policy-Makers_Proceedings.pdf
- Ramalho, A. 2017b. *Will Robots Rule the (Artistic) World? A Proposed Model for the Legal Status of Creations by Artificial Intelligence Systems*, available at SSRN: <http://dx.doi.org/10.2139/ssrn.2987757>
- Ramalho, A. 2018. *Patentability of AI-generated inventions: is a reform of the patent system needed?* Institute of Intellectual Property, Foundation for Intellectual Property of Japan, p.1-32, available at: <https://ssrn.com/abstract=3168703>
- Rathenau Institute 2017. *Human rights in the robot age*. Council of Europe Report. <https://www.rathenau.nl/sites/default/files/2018-02/Human%20Rights%20in%20the%20Robot%20Age-Rathenau%20Instituut-2017.pdf>
- Ritter, S., Anderson, J. R., Koedinger, K. R. and Corbett, A. 2007. 'Cognitive Tutor: Applied Research in Mathematics Education'. *Psychonomic Bulletin & Review* 14 (2): 249–55. <https://doi.org/10.3758/BF03194060>
- Rodrik, D. 2018. 'Populism and the economics of globalization'. *Journal of International Business Policy*. https://drodrik.scholar.harvard.edu/files/dani-rodrik/files/populism_and_the_economics_of_globalization.pdf
- Rosé, C. P., Martínez-Maldonado, R., Hoppe U. et al. (Eds.) 2018. *Artificial Intelligence in Education: 19th International Conference, AIED 2018, London, UK, June 27–30, 2018, Proceedings, Part I*. Lecture Notes in Artificial Intelligence. Springer International Publishing.
- Roy, S. S., Samui, P., Deo, R. et al. (Eds.) 2018. 'Big Data in Engineering Applications', *Studies in Big Data*. Vol. 44. Springer: Singapore.
- Sachs, J. D. 2017. R&D, 'Structural Transformation, and the Distribution of Income'. In Agrawal, A., Gans, J. S., and Goldfarb, A. (Eds.) 2017. *The economics of artificial intelligence*. McKinsey Quarterly, April. <http://conference.nber.org/conferences/2017/AIf17/summary.html>
- Samuelson, P. 2018. 'The EU's Controversial Digital Single Market Directive – Part II: Why the Proposed Mandatory Text- and Data-Mining Exception Is Too Restrictive', *Kluwer Copyright Blog*. <http://copyrightblog.kluweriplaw.com/2018/07/12/eus-controversial-digital-single-market-directive-part-ii-proposed-mandatory-text-data-mining-exception-restrictive/>
- Sanok, D. 2005. 'An analysis of how antivirus methodologies are utilised in protecting computers from malicious code'. *Proceedings of the 2nd annual conference on Information Security curriculum development*.

- Scassellati, B., Boccanfuso, L., Huang, C. M. et al. 2018. 'Improving social skills in children with ASD using a long-term, in-home social robot'. *Science Robotics*, 3(21), eaat7544.
- Schönberger, D. 2018. 'Deep Copyright: Up - And Downstream Questions Related to Artificial Intelligence (AI) and Machine Learning (ML)' in De Werra, J. (ed.) 2018. *Droit d'auteur 4.0 / Copyright 4.0*, Schulthess Editions Romandes, p.145-173. <https://ssrn.com/abstract=3098315>
- Selbst, A. D. and Powles, J. 2017. 'Meaningful information and the right to explanation'. *International Data Privacy Law*, 7 (4): 233–242.
- Semiconductor Industry Association (SIA), 2015. *International technology roadmap for semiconductors*. <https://www.semiconductors.org/resources/2015-international-technology-roadmap-for-semiconductors-itrs/>
- Schafer, B., Komuves, D., Zatarain, J. M. N. and Diver, L. 2015. 'A fourth law of robotics? Copyright and the law and ethics of machine co-production', *Artificial Intelligence and Law*, 23:3, 217–240.
- Short, J. E., Bohn, R. E., Baru C. 2011. *How much information? 2010–report on enterprise server information*. UCSD Global Information Industry Center.
- Sleeman, D., and Brown, J. S. 1982. *Intelligent Tutoring Systems*. New York: Academic Press.
- Stevenson, B. 2017. 'AI, Income, Employment, and Meaning' in Agrawal, A., Gans, J. S., and Goldfarb, A. (Eds.) 2017. *The economics of artificial intelligence*. University of Chicago Press. <http://www.nber.org/chapters/c14026>
- Stilgoe, J. 2017. 'Machine learning, social learning and the governance of self-driving cars'. *Social Studies of Science*, 48 (1): 25-56.
- Szegedy, C., Zaremba, W., Sutskever, I. et al. 2013. *Intriguing properties of neural networks*. <https://arxiv.org/abs/1312.6199>
- Tan, W., Fan, Y., Ghoneim, A. et al. 2016. 'From the Service-Oriented Architecture to the Web API Economy', *IEEE Internet Computing*. Jul-Aug: 64-68.
- The Future Society. 2017. *Making the AI revolution work for everyone*. Retrieved from <https://www.tuftsgloballeadership.org/sites/default/files/images/resources/Miaillhe%20Reading.pdf>
- Tollenaar, N., and Van der Heijden, P. G. M. 2013. 'Which method predicts recidivism best? A comparison of statistical, machine learning and data mining predictive models'. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 176.2: 565-584
- Tuomi, I. 2018. *The Impact of AI on learning, teaching, and education: Policies for the future*. Cabrera, M, Vuorikari, R. and Punie, Y. (Eds.) Joint Research Centre Science for Policy Report, European Commission, Joint Research Centre, Seville, Spain <http://publications.jrc.ec.europa.eu/repository/handle/JRC113226>
- UK Government Office for Science, 2016a. *Robotics, automation and artificial intelligence*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/592423/Robotics_automation_and_artificial_intelligence_-_cst_letter.pdf
- UK Government Office for Science, 2016b. *Artificial intelligence: opportunities and implications for the future of decision making*. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/566075/gs-16-19-artificial-intelligence-ai-report.pdf
- UK Government, 2018, *Industrial Strategy: Artificial Intelligence Sector Deal, Policy Paper, (April 26, 2018)*: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/702810/180425_BEIS_AI_Sector_Deal_4_.pdf
- UK House of Commons Science and Technology Committee, 2016. *Robotics and Artificial Intelligence* <https://www.publications.parliament.uk/pa/cm201617/cmselect/cmsctech/145/145.pdf>
- UK House of Lords, 2017. *AI in the UK: ready, willing and able?* Select Committee on Artificial Intelligence Report of Session 2017-19, HL Paper 100.
- University of Melbourne, 2018. *Biometric Mirror highlights flaws in artificial intelligence* https://www.eurekalert.org/pub_releases/2018-07/uom-bmh072318.php
- Van Asbroeck, B., Debussche, J. and Cesar, J., 2017. *Data ownership in the context of the European data economy: proposal for a new right*, White Paper, Bird & Bird.

- Van den Hoven, J. 2017. 'Introduction'. In Van den Hoven, J., Miller, S. and Pogge, T. (Eds.) 2017. *Designing in Ethics*, pp. 1-10. Cambridge University Press.
- Van Dijck, J. 2014. 'Datafication, dataism and dataveillance: Big data between scientific paradigm and ideology'. *Surveillance & Society*, 12 (2): 197-208.
- Verbeek, P. P. 2011. *Moralizing Technology: Understanding and Designing the Morality of Things*. Chicago and London: University of Chicago Press.
- Vesnic-Alujevic, L., Breitegger, M. and Guimaraes Pereira, A. 2015. 'Do-It-Yourself' Healthcare? Quality of Health and Healthcare Through Wearable Sensors'. *Science and Engineering Ethics*, 24 (3): 887-904.
- Vidal, J. 2017. 'Tsunami of data' could consume one fifth of global electricity by 2025'. Climate home news. <http://www.climatechangenews.com/2017/12/11/tsunami-data-consume-one-fifth-global-electricity-2025>
- Wachter, S., Mittelstadt, B., Floridi, L. 2017. 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation?', *International Data Privacy Law*.
- Weber, R., 2017. 'Improvement of Data Economy through Compulsory Licences?' in Lohsse, S., Schulze, R. and Staudenmayer, D. (Eds.) 2017. *Trading Data in the Digital Economy: Legal Concepts and Tools, Nomos*, pp.137-158.
- Woetzel, J., Seong, J., Wang, K. et al. 2017. *Digital China: Powering the Economy to Global Competitiveness*. Retrieved from <https://www.mckinsey.com/featured-insights/china/digital-china-powering-the-economy-to-global-competitiveness>
- Woolf, B. P. 2009. *Building Intelligent Interactive Tutors: Student-Centered Strategies for Revolutionizing e-Learning*. San Francisco, CA: Morgan Kaufmann.
- Xi, J. 2014. *The governance of China (1st ed.)*. Beijing: Foreign Languages Press.
- Yeung, K. 2016. 'The Forms and Limits of Choice Architecture as a Tool of Government'. *Law & Policy*, 38 (3): 186-210.
- Zech, H. 2016. 'A legal framework for a data economy in the European Digital Single Market: rights to use data' *Journal of Intellectual Property Law & Practice*, 11(6) 460-470.

LIST OF BOXES

Box 1.	Early success case of machine learning in cybersecurity: spam filtering	19
Box 2.	Turning point in image recognition	20
Box 3.	DL for singing processing	21
Box 4.	Some examples of AI developments in the USA	26
Box 5.	The Techno-Economic Segment (TES) approach in a nutshell	30
Box 6.	Some examples of AI applications in China	50
Box 7.	The impact of AI on working conditions	82
Box 8.	AI and digital labour platforms	83
Box 9.	Summary data strategy from the Mission Villani Report (2018)	109
Box 10.	Examples of APIs in local government	110

LIST OF FIGURES

Figure 1.	Paradigm shift in machine learning	20
Figure 2.	Percentage of AI players detected by the most representative keywords (2000-18)	30
Figure 3.	AI players in the world by geographical zones (% over world total), 2009-2018	31
Figure 4.	Top 10 AI players in the world and relation with GDP by geographical zones, 2009-2018	32
Figure 5.	AI players in the EU and relation with GDP by country, 2009-2018	32
Figure 6.	R&D versus non-R&D players in top 10 countries by number of AI players, 2009-2018	33
Figure 7.	Number of AI players in selected types of activity (% over world total), 2009-2018	33
Figure 8.	Distribution of EU players (%) in EU-funded AI research projects, 2009-18	34
Figure 9.	Occurrence of AI topics (%) by geographical zone, 2000-2018	35
Figure 10.	AI thematic profiles in academic programmes (as % of total programmes)	74
Figure 11.	Distribution of universities offering AI study topics over human resources in science and technology 2017	75
Figure 12.	Paradigm shift in the cybersecurity of systems following the introduction of AI components.	89
Figure 13.	Projection of data and computing growths (logarithmic scale)	96
Figure 14.	Datafication process enabling the AI revolution	112
Figure 15.	Employment in manufacturing and healthcare: EU and USA	117
Figure 16.	AI preparedness in EU Member States	117
Figure 17.	Human resources in science and technology and Digital Innovation Hubs	118
Figure 18.	Human resources in science and technology, Digital Innovation Hubs, and academic supply in study topics related to AI	119

ACKNOWLEDGEMENTS

The main contributors to this report were: Alessandro Annoni, Peter Benczur, Paolo Bertoldi, Max Craglia, Blagoj Delipetrev, Giuditta De Prato, Enrique Fernandez Macias, Emilia Gomez, Maria Iglesias, Henrik Junklewitz, Montserrat López Cobo, Bertin Martens, Susana Nascimento, Stefano Nativi, Alexandre Polvora, Ignacio Sanchez, Songul Tolan, Lucia Vesnic Alujevic from the JRC, as well as Claudio Feijóo (Universidad Politécnica de Madrid), and Ilkka Tuomi (Meaning Processing).

Other contributors were: Amanda Anderberg, Marcelino Cabrera Giraldez, Francesca Campolongo, Melisande Cardona, Vicky Charisi, Nestor Duch-Brown, Serena Fatica, Elisabeta Florescu, Jiri Hradec, Maciej Krysztofowicz, Marius Miron, Alberto Moro, Andrea Pagano, Andrea Perego, Yves Punie, Riccardo Righi, Sofia Samoili, Fabiana Scapolo, Lorenzino Vaccari, Lukazs Zemba from the JRC and Jean-Paul Simon (JPS Public Policy Consulting).

We are particularly grateful to the International Editorial Board members for their comments and support:

- Dr. Bob Bishop, President and Founder of the International Centre for Earth Simulation Foundation.
- Prof. Virginia Dignum, Professor Social and Ethical Artificial Intelligence, Umeå University.
- Dr. Cosmina Dorobantu, Deputy Director of the Public Policy Programme and Policy Fellow at the Alan Turing Institute for Data Science and AI.
- Dr. Nuria Oliver, Director of Research in Data Science at Vodafone and Chief Data Scientist in Data-Pop Alliance.
- Prof. Barry O’Sullivan, Director, Insight Centre for Data Analytics, University College Cork.
- Prof. Yves Pouillet, Professor at the Faculty of Law University of Namur, and Co-Chair of the Namur Digital Institute.
- Dr. Andrea Renda, Senior Research Fellow and Head of the Global Governance, Regulation, Innovation and the Digital Economy at the Centre for European Policy Studies.

GETTING IN TOUCH WITH THE EU

IN PERSON

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

ON THE PHONE OR BY EMAIL

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: **00 800 6 7 8 9 10 11** (certain operators may charge for these calls),
- at the following standard number: **+32 22999696**, or
- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

ONLINE

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU PUBLICATIONS

You can download or order free and priced EU publications from EU Bookshop at: <https://publications.europa.eu/en/publications>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU LAW AND RELATED DOCUMENTS

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: <http://eur-lex.europa.eu>

OPEN DATA FROM THE EU

The EU Open Data Portal (<http://data.europa.eu/euodp/en>) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.



EU Science Hub
ec.europa.eu/jrc



@EU_ScienceHub



EU Science Hub - Joint Research Centre



Joint Research Centre



EU Science Hub

