POLICY PAPER
4 NOVEMBER 2019
#AI
#TRUST
#INNOVATION

JACQUES DELORS INSTITUTE
BERLIN
Centre for European Affairs at the Hertie School of Governance

ESTABLISHING TRUST IN AN AI-POWERED FUTURE

GUIDING PRINCIPLES FOR A HYPERCONNECTED WORLD

Executive Summary

PAUL-JASPER DITTRICH
Policy Fellow,
Jacques Delors Institute
Berlin

In recent years, the progressing digital transformation has been characterized by the increasing use of "artificial intelligence" technologies. The automation of knowledge and routine tasks as well as the increasing use of automatic decision-making systems, for example in HR, has the potential to fundamentally change the lives of many people and entire societies, for better or for worse. One of the most important foundations for the success and acceptance of these technologies in European societies is trust. Without trust in the security, usefulness and transparency of applications using "artificial intelligence", there will be little scope for the full potential of AI in the EU and, in the worst case, the fabric of European societies will suffer.

Generating trust by empowering citizens should hence be a central policy aim for the next decade of the digital transformation. Only if citizens can understand, control and withstand technological solutions they are also in a position to trust them. But all too often, solutions produced by the market do not pass this test. It is therefore the role of governments to strike the right balance between entrepreneurial freedom, innovation and the protection of fundamental rights of European citizens.

European citizens expect regulatory action to manage the digital transformation towards an AI-powered future. Three areas where AI is currently having very disruptive effects should be front and centre of this regulatory effort because they touch upon the most sensitive areas of citizen's material and social wellbeing:

1. IT-security and product safety with the aim of making AI-powered products as safe as possible to use.

2. The implications of automation and AI for the workplace with the aim of making the transition for workers as smooth as possible.

3. Automated decision-making systems with the aim of making algorithms as non-discriminatory and transparent as possible.

TABLE OF CONTENTS

1 . AI AND TRUST: A BALANCING ACT BETWEEN INNOVATION AND CONTROL IS NEEDED

The digital transformation constantly produces new ways of organizing work, society and politics. Many of them improve our lives and increase our well-being. Many of them, however, have the **potential to trigger crises of trust**. Crises of trust in a new technology or application occur if users and citizens feel powerless and without control of their digital tools but instead controlled by them, for example because of data leaks, security failures, fraud or non-transparent decisions of an algorithm with life-altering consequences. Generating trust by empowering citizens should hence be a central policy aim for the next decade of the digital transformation. Only if citizens can understand, control and withstand technological solutions they are also in a position to trust them.

> "IT IS THE ROLE OF GOVERNMENTS TO STRIKE THE RIGHT BALANCE BETWEEN FREEDOM, INNOVATION AND THE PROTECTION OF FUNDAMENTAL RIGHTS

But all too often, solutions produced by the **market do not pass this test**. It is therefore the role of governments to strike the right **balance between entrepreneurial freedom, innovation and the protection of fundamental rights of European citizens.**

On data protection, the EU has shown determination and assertiveness. In carefully balancing the trade-off between innovation and the right to privacy in the GDPR, the EU has proven that it is willing to protect the interests of European citizens by strengthening their individual agency vis-à-vis data-processors and enabling them to control technology to some degree instead of feeling controlled by it.

This approach should be continued when it comes to regulating "Artificial intelligence". In a broad sense of the term, namely AI as software applications using various techniques like machine learning or neural networks in order to automate tasks or decisions with some degree of autonomy, **AI is already, or will soon be, applied in all sectors of the economy**. This ranges from health care, where AI can help automating the research of new pharmaceutical drugs and drug trials, to autonomous vehicles, better manufacturing and service robots, delivery drones or live translation programs. Carefully regulating AI in a way that allows for as much innovation as possible while thoroughly protecting the fundamental rights of EU citizens has to be one of the main political priorities for the incoming European Commission. France and Germany, as the EU's two largest economies, have to lead in this sensitive policy area and should demand an ambitious commitment from the new Commission to develop concrete rules for a hyper-connected world, which enables its citizens to confidently approach and use the possibilities of new technologies. Where possible, they should also develop bilateral initiatives to speed up the process.

2. ANALYSIS: BREAKING DOWN THE CHALLENGE OF AI TO EUROPEAN SOCIETIES

> "EUROPEAN CITIZENS EXPECT REGULATORY ACTION TO MANAGE THE DIGITAL TRANSFORMATION

Given the **depth (transformation of business models)** and **breadth (number of sectors affected)** of the economic and social consequences of AI, it is no wonder that it is not unequivocally welcomed in the EU. According to a 2017 Eurobarometer survey, six in ten respondents in the EU (61 percent) have a positive view of artificial intelligence, while 30 percent have a negative view.[1] They have, however, very clear expectations towards companies and governments. 88 percent of Europeans agreed that artificial intelligence as a technology required "careful management". This means that European citizens expect regulatory action to manage the digital transformation towards an AI-powered future. Three areas where AI is currently having very disruptive effects should be front and centre of this regulatory effort because they touch upon the most sensitive areas of citizen's material and social wellbeing:

1. IT-security and product safety with the aim of making AI-powered products as safe as possible to use.

2. The implications of automation and AI for the workplace with the aim of making the transition for workers as smooth as possible.

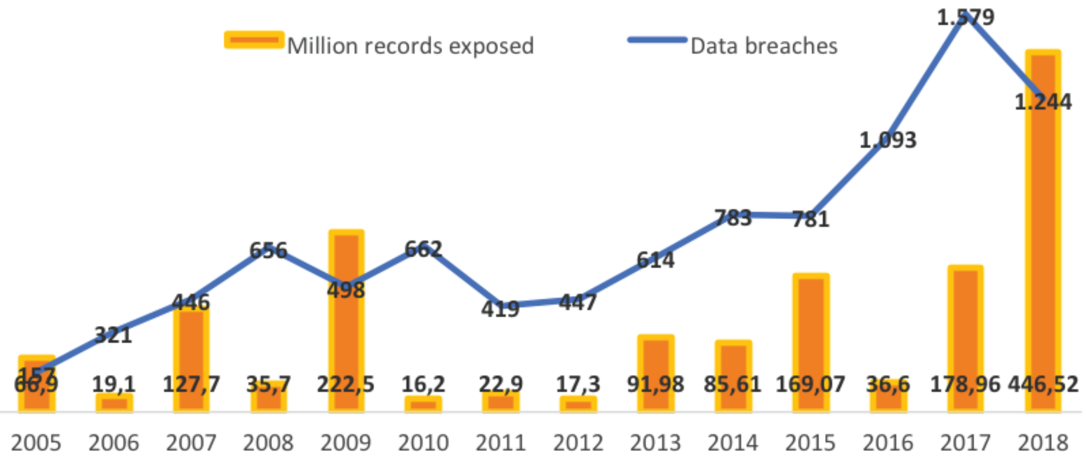
3. Automated decision-making systems with the aim of making algorithms as non-discriminatory and transparent as possible.

2.1. Safety/security

IT-security is not an AI-exclusive problem. Hacking of accounts, scamming, identity theft and ransom attacks have been around since the beginning of the internet. A **record number of 446 million individual records of data subjects were exposed only in American data breaches in 2018**.[2] Cyber-crime is notoriously hard to investigate due to easily available cryptography and anonymization technology and the transnational nature of the internet. Recurrent incidents of mass data leaks have contributed to a rising feeling of distrust in the ability of tech and other companies to keep sensitive consumer, patient or employee data safe with them. Insecure products and services have opened up numerous new opportunities for criminals in the last decades.

[1] Special Eurobarometer 460, Attitudes towards the impact of digitisation and automation on daily life, May 2017.
[2] Identity Theft Resource Center, 2018 End-Of-Year Data Breach Report.

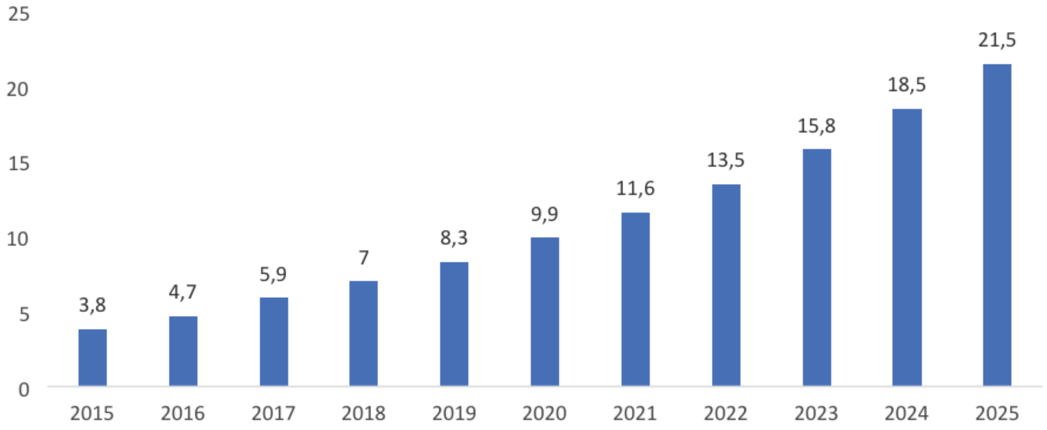
FIGURE 1 ■ Number of breaches and records exposed 2005–2018 in the US



Source: Identity Theft Resource Center

AI and AI-powered devices are likely going to increase them further. Risks loom, for example, in connection with "smart home devices" like AI-powered speakers or video cameras. The advent of connected "smart" devices (speakers like Alexa, locks, video cameras, fridges etc.) which are often accompanied by some form of voice or facial recognition software opens up a whole new entry gate for hackers, fraudsters and other malicious actors. As regards business-to-business relations, the billions of connected devices, machines and sensors that together make up the Industrial Internet of Things (IoT) will be targeted as well. 5G will further speed up the introduction of connected devices, the overall market for Connected IoT devices is projected to have 10-20 % growth rates per year (see figure below). In the next decade **the EU must prepare for challenges related to the safety of connected devices such as a mass take-over of these devices to use them with malicious intent.**

FIGURE 2 ■ Global Number of Connected IoT Devices in billions



Source: IoT Analytics Research 2018.

2.2. Future of work / automation

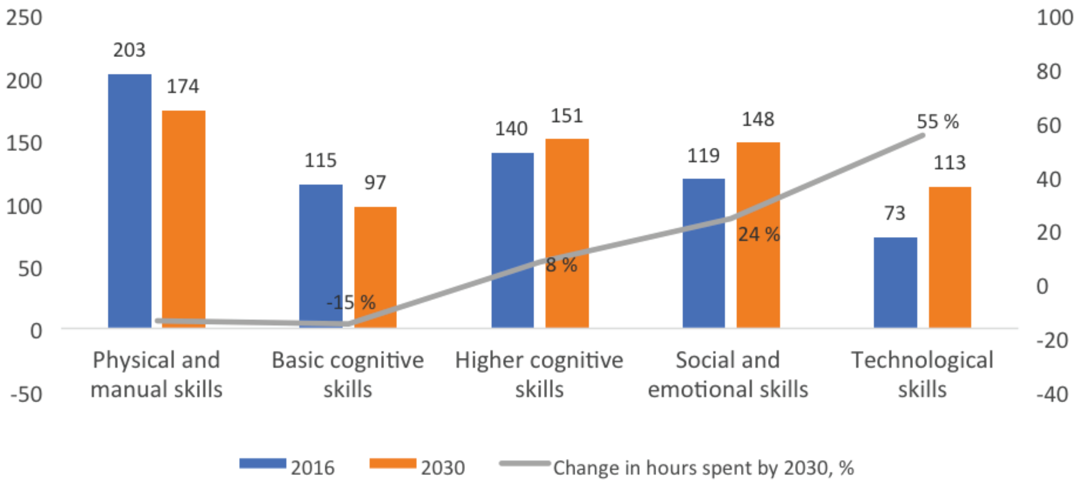
Early estimates of the potential impact of AI on labour markets, as provided by the Osborne Frey study[3] (a European application of it[4] predicted up to 54 percent of European jobs at risk by computerization, read: AI) were probably slightly overblown. Nonetheless, **AI will dramatically change the way we work and collaborate in the EU on all skills levels and in most sectors of the economy.** This change bears an explosive mix of economic and political risks. In a worst-case scenario, workers could experience a rapid devaluation of their skillsets amid fears of automation and a populist backlash against the accelerated introduction of AI.

This danger of a loss of trust in technology will probably occur even in the absence of dramatic job cuts. In fact such a process might already have started. Even though AI so far has not triggered an avalanche of job cuts in the EU, its impact on the collective imagination is already quite tangible. **72 percent of respondents fully or somewhat agreed in a 2017 Eurobarometer survey that AI was "stealing people's jobs"**.[5]

> "INSTEAD OF JOBS DISAPPEARING, JOB PROFILES AND SKILLS REQUIREMENTS FOR CERTAIN JOB CATEGORIES ARE GOING TO CHANGE

Yet, employment in the EU is not likely to decrease anytime soon. Instead it is important to analyze and communicate the actual changes of labour markets and skills. AI is in most cases not replacing entire jobs, but instead automates individual aspects or tasks of jobs like evaluating an x-ray picture or scanning through 10.000 pages of litigation documents. That means that **instead of jobs disappearing, job profiles and skills requirements for certain job categories are going to change,** especially for people working in the manufacturing sector or in lower skilled service sectors. As the chart below estimates, physical and manual skills as well as basic cognitive skills will be in slightly lower demand, but social/emotional as well as technological skills will likely see an enormous increase in demand.[6]

FIGURE 3 ■ Total hours worked in Europe and United States 2016 vs 2030 estimate in billions (left axis)



Source: McKinsey Global Institute Workforce Skills Model; McKinsey Global Institute analysis 2018

[3] Carl Benedikt Frey and Michael A. Osborne, The Future of Employment. How Susceptible are Jobs to Computerisation, September 2013 University of Oxford.

[4] Jeremy Bowles, Chart of the Week: 54% of EU jobs at risk of computerisation, Bruegel Blog Post, 24.07.2014.

[5] Special Eurobarometer 460, Attitudes towards the impact of digitisation and automation on daily life, May 2017.

[6] McKinsey Global Institute, Skill Shift. Automation and the Future of the Workforce, Discussion Paper May 2018.

2.3. Ethics of automated decision-making

Automated decision-making software is increasingly used across the EU, for example to automatically process traffic offences in France or match traveler's data at the border to police data bases and criminal files in Slovenia.[7] The problem is that, as these systems are being installed everywhere, few guidelines exist about how to use them in an ethical way and overall transparency on their inner workings (for example the weighting that lead to a certain decision) is low. This is especially problematic as decision-making software is not neutral or impartial. Training data for algorithms reflects a social reality, for example one that has a gender bias or discriminates towards minorities. This "mirror" of society that algorithms create generates many non-trivial problems. Racism, material privileges or "wrong ZIP-codes" become perpetuated in models for algorithmic decision-making, contribute to skewed and discriminatory decisions, and eventually perpetuate existing social inequalities. The problem is aggravated by the often opaque and non-transparent nature of the parameters of the automated decision-making process.

How can we **minimize the unethical and non-transparent use of algorithmic decision-making systems** for example in HR or interactions between citizens and administrations and ensure their fair use where they bring real benefits to individual citizens and European societies? How do we achieve a consensus on the essence of these ethical guidelines? How can we develop a policy process, which ensures that these questions are continually addressed and as many voices as possible are being heard? The current debate on the ethics of algorithms is still very theoretical and confined to a rather small circle of intellectuals, activists and some government officials on national and EU-level. None of these questions are even close to being sufficiently answered at the moment and need to be addressed urgently by the new Commission.

2.4. The problem: The EU also needs to innovate faster on AI

“
THE EU HAS A SERIOUS PROBLEM WHEN IT COMES TO TECHNOLOGY IN GENERAL AND NEEDS TO CATCH-UP

Balancing the need to address these risks with the need to stay innovative will not be easy: The EU already has a serious problem when it comes to technology in general and needs to catch-up. Economic competitors of the EU in AI development like the US, China, Canada or Israel are already far more advanced in some sectors like facial recognition software. The US also features a much higher number of AI-startups overall. Israel alone has more AI-startups than France and Germany combined.[8]

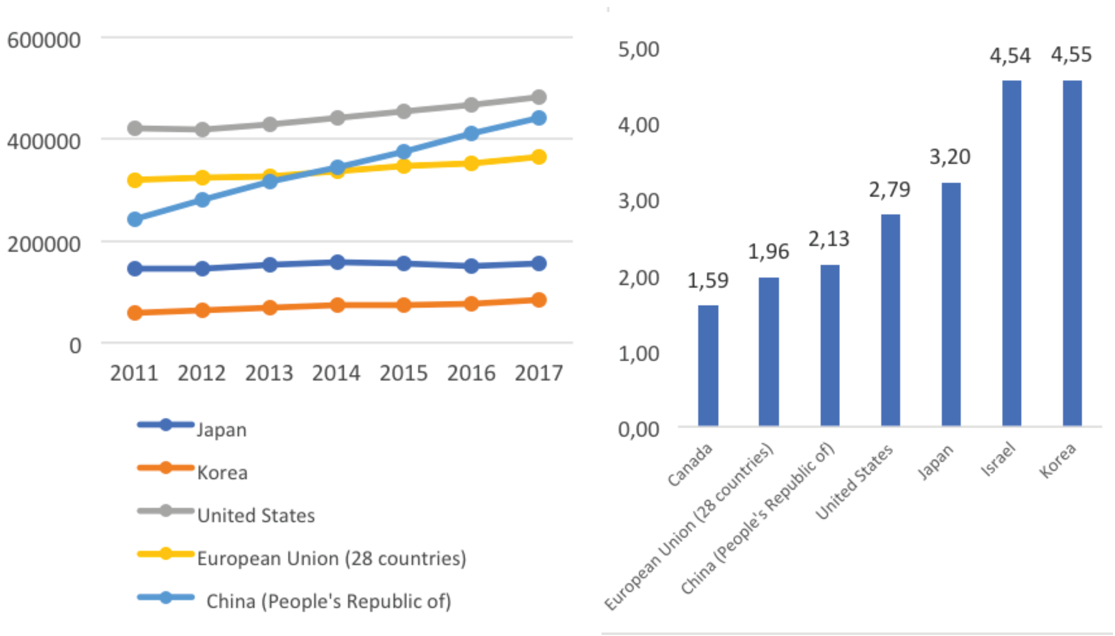
To make matters worse the best-placed European country, the UK, is bound to leave on October 31st. Many countries are also much more ambitious in their public investment targets for research and development for AI. This is following a wider trend: **In terms of expenditures in Research and Development China has surpassed the EU in absolute numbers (2014) and also had higher per capita R&D investments in 2017.**

[7] Algorithm Watch, Bertelsmann Foundation, Automating Society. Taking Stock of Automated Decision-making in the EU, 2019.

[8] Roland Berger, Asgard, Artificial Intelligence – A Strategy for European startups, 2018.

FIGURE 4 ■ GERD at constant prices and PPP $

GERD as % of GDP, 2017



Source: OECD Main Science and Technology Indicators 2018 GERD= Gross Domestic Expenditure on Research and Development.

The Chinese region of Tianjin alone announced to launch an Artificial Intelligence Fund with a budget of about 16 billion, which is almost as much as the EU promised to invest (20 billion Euros) by the end of 2020. The EU will need to invest more public and private resources in R&D and especially in AI if wants to keep up with the leaders in the field in the future.

The following three recommendations are thus written in the spirit of reconciling the need for trust in new technologies on the hand, and the imperative to speed up innovation and diffusion of technology on the other.

3 . RECOMMENDATIONS

3.1. Security and safety: Product Liability for Connected Devices

One of the reasons for the low IT-security of connected devices is an incentive problem: currently there are no strict product liability rules in place, for example for household appliances or routers.[9] A producer of an Internet-connected video camera is not liable for any hacks into his product. Hardware and software might often be from two different producers and without clear rules none of them has an incentive to provide software updates for their devices even in the case of known vulnerabilities. As a result many connected devices currently hit the market without any built-in option for security updates. In the next years, the amount of AI-powered connected devices will increase dramatically. Introducing **product liability for connected devices in the EU** would be one possible contribution to a strategy to increase trust in future AI-applications and should become a top priority for the upcoming Commission.

[9] Jan-Peter Kleinhans, Improving IoT Security in the EU, August 2018 SNV Working Paper.

3.2. Future of work/automation: A new Strategy for Re-Skilling

> "THE EU HAS SHOWN IN THE PAST THAT IT CAN MAKE CONTRIBUTIONS TO RE-SKILLING EFFORTS

Changing job profiles means that workers must get a fair chance to train and improve their individual skills in order to compensate for the loss of certain aspects of their jobs and accompanying skills. **A priority for the new Commission should accordingly be programs for learning and re-skilling** with the aim of mitigating the shift into automation and AI and generate trust in the technology. Education, and hence re-skilling, is of course not an EU competence but a national or sub-national one. Nevertheless, the EU has shown in the past that it can make contributions to re-skilling efforts in the EU, by creating and sustaining large partnerships with private companies. The European Coalition for Digital Skills for example has enabled hundreds of thousands of Europeans to acquire basic coding skills paid for by large American and European technology companies. A similar coalition could, for example, be developed by French and German manufacturing companies together with the European Commission with the aim of preparing as many workers as possible for the transformation of their sectors currently subsumed under the labels "Industrie 4.0" or "Industrie du Futur".[10] By exchanging best practices and sharing some of the costs related to re-skilling the EU could address the gaps in preparedness for the digital transformation of core industries across the EU.

3.3. Algorithmic bias and ethics: Finding the right principles

Even though there is no consensus on tools and methods and concrete policies a common idea of necessary meta-conditions and guiding principles for the ethical use of algorithms is emerging:

1. Fairness: minimizing individual bias in data and models. This can be done, for example, by putting more mandatory scrutiny on training data upfront and forcing the users to have their models scrutinized by impartial outside actors, for example by civil society.

2. Robustness and security: It must be ensured that the models of an algorithmic decision making system have been thoroughly tested to ensure robustness of the results, not least to guarantee that these systems do not pose security risk for society.

3. Interpretability and explanation of results: Individuals should have the possibility to get an explanation when, for example, their loan request or job application was turned down by an automated decision making system. The GDPR has already started to open up some possibilities around explanation of results which have to be specified by the new Commission.

4. Governance and accountability. Algorithmic decision-making systems and the institutions running them need to be publicly scrutinized in their use of these systems and be held accountable for failures to prevent avoidable discriminatory patterns. To ensure such accountability will require new forms of intra-company governance structures for example with the introduction of algorithmic transparency officers (like data protection officers) and also new governance and oversight structures for outside scrutiny.

[10] See for more details on this proposal Paul-Jasper Dittrich, Re-skilling for the Fourth Industrial Revolution, November 2016 Jacques Delors Institut Berlin Policy Paper No. 175.

The outgoing Commission has launched various expert dialogues on algorithmic bias and AI and is currently mulling its next steps for the governance of algorithms.[11] France and Germany should be particularly active in lobbying for an ambitious proposal for algorithmic governance including the set-up of a new oversight agency and demand decisive involvement from civil society in the process of oversight and control. If the EU manages to erect a transparent and accountable structure for the governance of algorithmic decision-making systems it will have taken a major step towards their acceptance. Because technological solutions survive and thrive in the long-term only if they generate trust.

ON THE SAME TOPIC

- Paul-Jasper Dittrich, Better Together? Franco-German Cooperation on AI. Policy Brief, 18.12.2018
- Dr. Nicole Koenig, Dr. Valentin Kreilinger, Paul-Jasper Dittrich, Aachen Treaty: A Second Look. Policy Position, 24.01.2019
- Paul-Jasper Dittrich, Jan Krewer, The European Answer to the Digital Revolution. How to Ensure Europe's Competitive Advantage? Policy Brief, 20.09.2018

[11] See for example Ethics Guidelines for Trustworthy AI, downloadable on the Commission Homepage, April 2019.

Managing Editor: Henrik Enderlein ▪ The document may be reproduced in part or in full on the dual condition that its meaning is not distorted and that the source is mentioned ▪ The views expressed are those of the author(s) and do not necessarily reflect those of the publisher ▪ Jacques Delors Institut – Berlin cannot be held responsible for the use which any third party may make of the document ▪ Original version ▪ © Jacques Delors Institut – Berlin, 2019

NOTRE EUROPE
JACQUES DELORS INSTITUTE
Hertie School
of Governance

Friedrichstraße 194, D – 10117 Berlin
info@delorsinstitut.de
www.delorsinstitut.de