HOW TO REGULATE AI?

TOWARDS TRUSTWORTHY ARTIFICIAL INTELLIGENCE

AUTHORS: ALJAŽ KOŠMERLJ, IVAN BRATKO, JANEK MUSEK,
KASIA SÖDERLUND, ALEKS JAKULIN, NINA PEJIČ
EDITOR: GREGOR PLANTARIČ
NOVEMBER 2019

Humans today have embraced the comfort of modern technology without really understanding how it works and the abnormal potential of the data accumulation that our devices have. These devices know exactly who we are, where we are, and what we are doing (besides thousands of other things). They achieve this with the help of many sensors and data input vectors controlled by Artificial Intelligence (AI) systems, or rather "smart" algorithms running in the background. These sometimes simple but mostly increasingly complex algorithms can, in a split second, find, sort, define, control and act on a multitude of data to help us with our day-to-day tasks. Otherwise today people would probably be lost in the profusion of information.

For most people algorithms are like some kind of "modern myth"; they wonder: what are they? What exactly are they doing? Can they do anything? Who controls them? Who controls their supervisors? Do we control algorithms or do they control us? What are the consequences? What happens if we give them our most precious possession - our life? And a whole bunch of other moral and ethical issues. Scientists are thus hypothetically wondering, for example, how the algorithm should make a decision in a driverless car in the event of an accident which involves two human lives (when at one point the algorithm has to decide between one life and the other life, who will be "allowed" to live)?
Algorithms (in the world of computer science) are computer programs that "help computers decide things" (basically "a list of steps that leads us to solutions to the problem"). Which sounds pretty straightforward and innocent until thousands, or millions are in action at the same time and then we start to wonder what is actually going on. Algorithms can now also operate autonomously and operate and control machines. Such autonomous robots can already function

and behave independently of humans (or at least with a high degree of independence); however, when one day they are equipped with (highly developed) artificial intelligence, which can understand or learn any intellectual task that a human being can undertake - also known as "General AI"[1]-, we are slowly approaching science fiction scenarios. Of course, we are still very far away from that, but we already have "less smart" machines in use in industry, the military, medicine, space exploration, etc. (some recent projects include Google Car, Atlas, Paro, Skybot F850, Cheetah, BigDog, Spot, Petman, iRobot swarm, autonomous military drones, the social humanoid robot Sophia and the robot Ori Hime). Of course there are always political tendencies to use the AI system and algorithms to control people (the infamous EU project INDECT or the Chinese Social Credit System), or even to predict the criminal behaviour of people based on algorithms in the style of the science fiction film "Minority Report" – those systems have serious flaws and biases (discriminating algorithms).

Since AI systems (respectively algorithms) are increasingly affecting our lives there has been a tendency for legal certainty; on how to regulate them? The regulatory challenges significantly exceed those imposed on autonomous driving or finance, to name just a couple, not forgetting that we are dealing with a distinctly interdisciplinary question. Hence the increasing number of legal and other experts and scholars who are addressing this issue. The regulation of AI or more specific algorithms is only a part of a broader interdisciplinary field of regulation of emerging technologies (nanotechnology, biotechnology, autonomous machines, robotics etc.), but this will prove to act as a signpost and cornerstone for our future regulation of emerging technologies. Allowing algorithms to control our life is often compared to opening Pandora's box. That's why it is of the utmost importance that our next moves are very well thought out, since further developments in this field can only be expected to be exponential.

1 Artificial general intelligence (AGI), strong AI, full AI etc.

SURVIVING AND MANAGING ARTIFICIAL INTELLIGENCE
BY ALJAŽ KOŠMERLJ

Every major technological advance impacting society is naturally met with some level of scepticism, distrust, and even fear. People need time to study a novelty and decide whether and how to accept it. The more involved this novelty is with the basic functions of our lives – for example, how we move around or how we communicate – the greater its impact and thus the tentativeness of acceptance. The field of artificial intelligence (AI), which appears to invade the area of our very thoughts and mental capacities, can, by this reasoning, only be expected to meet with the strongest of resistance. These fears are, however, often based on a poor understanding of the nature of the field. They cloud our ability to capture the true issues, often far more mundane than the sci-fi-quoting headlines, as well as the potential benefits AI can offer in the short and mid-term.

The majority of the modern field of artificial intelligence deals with very specific and narrow tasks. In this sense the very name can be misleading, as there is no single "intelligence". This field studies and develops methods that solve problems for which humans deem that a certain level of intelligence (whatever that may be) is needed. These methods are tuned to solve a particular task well. They may be able to do so with superhuman efficiency, but their abilities do not generalise to other tasks. A program taught to find cars in photographs will not (without specific extra training) be able to identify houses, dogs, or even buses, and Deep Blue, having beaten Kasparov in the historic 1997 match, would be a poor draughts player (as it doesn't even know the rules, let alone about strategy). This is part of the reason why most AI researchers are not worried about any type of technological singularity scenario, where a computer that is smarter than humanity would design an even smarter computer, etc. Such automated innovation would require great scientific breakthroughs at a very basic level and is unfathomable just through the continuation of the current modelling approaches that may identify relevant advertisements from our online behaviour.

The real problems of AI in today's society are a lot more down-to-earth than news coverage might sometimes imply. Wild speculation only hurts discussion on how to address these. Mentioning an Asimov-inspired[2], all-seeing and all-controlling autonomous AI system, that could take over our government, adds little to the current problem of automatic targeting in political advertising (of Cambridge Analytica fame). Related big problems, such as self-driving cars, automatic mass surveillance systems, or autonomous military weapons, need a lot of thought and debate, but it needs to be level-headed and fact-based.

Better public understanding of artificial intelligence would not only enable better dialogue regarding these big problems but would also alleviate a lot of them by informing peoples' actions. This is driven home even further by the fact that the principles of AI algorithms used today can be understood using high-school-level mathematics. The payoff would arguably be greater than just mitigating existing problems. These methods can be used by individuals as tools to enhance their lives and work. We should strive towards equipping everyone, especially children and young adults, with the skills for being wary of and handling something so impactful. People will adapt to the realities of life with AI, just as they have with email spam or shady marketing practices, but structured efforts can help a lot.

2 Isaac Asimov (https://en.wikipedia.org/wiki/Isaac_Asimov) an American science fiction writer and the inventor of the term "robotics".

HOW TO REGULATE AI?
TOWARDS TRUSTWORTHY ARTIFICIAL INTELLIGENCE

Education alone cannot solve everything as the implications of the use of artificial intelligence are so broad and deep that systemic action is warranted. Most experts agree that some level of regulation and oversight is needed, especially when considering the potential adverse use of this technology (e.g. for crime or misleading the public). However, the actions taken can be formulated to seek opportunities rather than just limiting possibilities. For example, one concern regarding the widespread use of AI in government and the service industry is that it may take on biases against minority groups through the data it analyses. But these are commonly biases that are already present in these systems. By going through a modelling process on this data with appropriate algorithms which can be thoroughly examined gives us a chance to identify these biases and address them. Of course, this requires expertise and a rigorous methodology which can be formally required and inspected. Transparency of the entire process and of the models is key for numerous applications. Thus, it makes sense to strive for certification of AI systems – seeking proof that they work appropriately for the sensitivity of the task they are performing. In this sense there is a clear parallel to the pharmaceutical industry, where the exact function of the drugs may not be clear to lay people, but a system of oversight ensures safety for the public.

There are several international organisations working towards managing AI and the EU has an active role in at least three of these efforts. In the second half of 2019, the Council of Europe (CoE) started work on the legal standardisation of AI and formalising AI certification. Slovenia is supporting this work through its seat in the United Nations and is leading the CoE committee on AI[3]. EU experts have also played a crucial role in the formulation of the OECD AI Principles[4]. These principles have been formulated over eight months of work by 40 people and enshrine the ideals of inclusivity, fairness, transparency and accountability. They are also the only formal international agreement regarding AI on this level, having been voted on and agreed to by 43 countries. Finally, as of when this text is written, UNESCO is confirming the establishment of its International Research Centre for AI at the Jožef Stefan Institute in Ljubljana. The centre will focus on the topics of using AI for the common good, education, and AI-related policy.

In conclusion, as with any ground-breaking technology, artificial intelligence offers many opportunities as well as pitfalls and there must be careful consideration to take advantage of the former while avoiding the latter. These considerations should be based on fact rather than hype, otherwise we risk losing sight of the key issues. There are strong international efforts in play, with Slovenian experts at the forefront of these, to build a formal framework which would ensure that the use of AI follows human values and works towards the common good.

3 https://www.coe.int/en/web/artificial-intelligence/cahai
4 http://oecd.ai/

WHY IS EXPLANATION IN ARTIFICIAL INTELLIGENCE SO IMPORTANT?
BY IVAN BRATKO

Trustworthy Artificial Intelligence is one of the declared goals of European Commission. An important component of Trustworthy AI is the ability of the system to explain its decisions. This means that the system has to be capable of explaining its decisions and its reasoning in terms that are understandable to humans. This feature is also referred to as the transparency of an AI system, as opposed to a black box whose decisions are difficult or impossible to understand by humans. Here we analyse why transparency is often so important, and why it may be difficult to achieve.

The field of AI employs a toolbox of computational methods. Some of these methods naturally tend to be transparent. For example, AI reasoning techniques based on mathematical logic are, at least in principle, possible to be traced by a human as a sequence of logical reasoning steps. This is similar to following mathematical proof, one logical inference after another. Logic-based methods therefore tend to be suitable for explanation. On the other hand, some other kinds of AI methods are notoriously hard to follow. The rationale behind the decisions produced by such methods is therefore inaccessible to a human user. Such methods behave like black boxes. Such methods do not provide explicit justification, nor do they enable further enquiry into the problem. The user is then supposed to accept their decisions or recommendations as they are. The type of reasoning methods that naturally support transparency are also called symbolic methods, and the latter type (black boxes) are called sub-symbolic. One famous class of black box methods are neural networks, in particular the kind of networks used in deep learning[5].

It should be noted that some black box methods, deep learning in particular, have recently enjoyed the most rapid progress in terms of performance accuracy. Yet they remain questionable because of the lack of explanation, although considerable research efforts have been made towards improved transparency. Methods of explanation for deep learning remain largely insufficient. Sometimes, to improve the transparency of an AI method, its accuracy is deliberately sacrificed.

It depends on the application as to whether explanation by an AI system is critically needed. The need for explanation increases with the importance of decisions, and also with uncertainty about the correctness of machine decisions. If there is a considerable possibility that a machine's decisions are incorrect or debatable, then the explanation is needed all the more. Some examples of problem domains in which explanation tends to be critically needed are: medicine, finance, legal decisions, and also political decision-making. In the last example, political decisions, the need for explanation might not be entirely obvious. Consider a political party leader who is using an AI advisor. Suppose that in the middle of the election campaign AI recommends a decision that seems to contradict the party's political programme. The party leader is reluctant to implement this advice as there seems to be no obvious reason for this. So she would like an explanation by the AI system, to check whether it is really necessary, or there is an error in the system's judgement. But as the machine cannot explain the reasons, the leader will be tempted simply to accept its advice just because she knows that a machine's decisions are on average more accurate than a human's. In such a case, a lack of transparency may lead to the computer effectively changing the party's political programme and its values.

5 Deep learning is part of a broader family of machine learning methods based on artificial neural networks (ed. note)

In the discussion above, transparency referred to individual decisions or recommendations. In machine learning[6], there is another type of transparency: after the learning system has completed its learning from data, the question is, what has the system actually learned? The result of learning is a general theory of the domain that can be used for decision-making, and which explains the regularities in the domain. Knowing the automatically learned theory would be very useful to a human user in order to improve the user's own knowledge of the domain. In cases where machine learning is used in science to construct new scientific theories from experimental data, understanding the resulting theories is obviously necessary. Again, some machine learning methods, including deep learning, do not provide satisfactory insights into such new theories.

Understanding a machine's decisions and intentions is also essential in cases where a human and a machine cooperate in solving a task. An example is a human and a robot jointly accomplishing a task, each of the two partners carrying out operations for which they are more suitable. For such a cooperation to be effective, the human has to understand the purpose of the actions carried out by the robot.

A further situation when explanation is important arises when an AI system makes a counter-intuitive decision or an obvious mistake. An explanation of such a decision is useful in finding out what the problem is, and how to fix it. There are well-known surprising examples of completely fooling a neural network in image recognition tasks, although the network otherwise seems to be performing with high accuracy. Good and easy to understand explanations of how such mistakes may happen would be very useful for building trust in this technology and assessing risks. This seems to be even more important in cases where the correctness of decisions by AI systems cannot be absolutely guaranteed, but only in probabilistic terms (giving estimates of the probability of the system making an incorrect decision). This is also a reason for difficulties in the certification of automated decision-making systems. An explanation might be a way towards alleviating this problem.

Although the problem of explanation in AI has been known for a long time, it was only recently recognised generally as a critical component in many applications of AI, in particular in automated prediction and decision-making. Methods for automated explanation generation are being developed in the area of AI now popularly called Explainable Artificial Intelligence (abbreviated XAI). It should be noted that despite the significant attention that XAI is now attracting, technical progress towards practical and effective methods for generating good explanations is slow. Given the importance of explanation in building Trustworthy AI, more research in XAI should be supported.

6 Machine learning (ML) is the scientific study of algorithms and statistical models that computer systems use to perform a specific task without using explicit instructions, relying on patterns and inference instead (ed. note)

THE IMPORTANCE OF ETHICS IN THE AI DEVELOPMENT
BY JANEK MUSEK

The "importance of ethics" sounds like an overused phrase. Yet, in truth, we are not fully aware of this importance. All major societal, personal and professional risks are the consequences of the gap between values and ethical standards, on the one hand, and value-aligned, ethical behaviour, on the other. War, violence, corruption, crime, exploitation, terrorism, poverty, poor interpersonal relations, low working efficiency, all of these are at least to some extent connected to the difference between ethical standards and actual behaviour.

A revolutionary development of AI is provided in all serious scenarios of our future. According to the most important of them, like the "Super smart society" or "Society 5.0", the future, unprecedented development of AI should be based on values, including new or newly adapted values. However, the values are the basis of ethics. It is very clear, therefore, that proper ethical and moral standards should be the basic guidelines for AI development and its practical use. AI is the most powerful tool invented in the history of humanity. In comparison to other tools, the development of AI produces augmented potential benefits and potential harms. The ethical consideration of AI development and use is therefore even more important.

Thus, successfully coping with the major challenges of AI, societal, economic, psychological and even evolutionary, must be directed and managed by clear ethical and moral rules. Unethical practices in the development, use and implementation of AI should be prevented. Ethical regulation should also include the integration of ethical principles into the development of learning, creative in self-programming AI systems. Indeed, the integration of ethics into AI systems can be very promising for another simple reason. We know very well that human beings cannot be perfectly ethical. However, fully ethical AI systems are possible: they could exceed humanity even in terms of the ethical point of view.

Yet how do we deal with the ethical challenges of AI when we don't have clear solutions to many other ethical problems? In the EU, the proposed schedule of ethical AI development regulation is focused on the concept of Trustworthy Artificial Intelligence. According to the proposed EU guidelines (Ethics Guidelines for Trustworthy Artificial Intelligence, 2019), a trustworthy AI should be:

- Lawful (AI must respect all applicable laws and regulations),
- Ethical (AI must respect ethical principles and values), and
- Robust (AI must be technically secure, and it must take into account the social environment at the same time).

From the ethical point of view, Trustworthy AI should follow four basic principles:

1. Respect for human autonomy (development and use of AI systems must respect human rights, freedom and autonomy; human oversight over AI must be secured);
2. Prevention of harm (AI systems should not cause harm or adversely affect human beings in any way);

3. Fairness (the development and use of AI systems must be fair and unbiased);
4. Explicability (the development and use of AI systems must be explainable; they should be based on understandable principles that can be communicated to those directly or indirectly affected by AI).

Further, a Trustworthy AI could only be realised considering seven key demands or requirements:

1. Human agency and oversight (AI should not hamper EU fundamental rights, users should be able to understand and interact with AI systems to a satisfactory degree, the rights of end users should not be subject to a decision based solely on automated processing, AI systems and machinery should be fully under human control);
2. Technical robustness and safety (AI must comprise secure and reliable hardware and software; cybersecurity must be provided);
3. Privacy and data governance (privacy and data protection must be ensured);
4. Transparency (data sets and processes used in building AI systems must be clearly established, fully explainable to human agents, documented and traceable);
5. Diversity, non-discrimination and fairness (all unfair biases should be avoided in the use, development and processing of AI);
6. Societal and environmental wellbeing (AI systems should be used for the benefit of preserving democracy and a sustainable environment);
7. Accountability (the accountability of and responsibility for the use and outcomes of AI systems should be ensured).

Finally, the legislation should provide ethical solutions for cases of conflicts between basic AI ethical principles, for example for the conflicts between human autonomy and the prevention of harm. These solutions should, for example, comprise when and to which extent the use of AI in preventing crime should interfere with individual freedom and privacy.

The Trustworthy AI presupposes full human control of AI development. Yet in the AI development perspective, there is another much darker outcome. In approximately two human generations or sooner, AI can reach so-called "technical singularity"[7] and achieve the criteria of superintelligence (defined as "any intellect that greatly exceeds the cognitive performance of humans in virtually all domains of interest")[8]. Therefore, it can reach the level of absolute emancipation, absolute independence of our species. This development can radically alter the ethical state of affairs. The emancipated, independent, self-conscious and self-sufficient AI can (secretly or not) create its own ethics, plan its own future and the future of the world. Thus, AI will be able to put the very further existence of humankind under question. We can imagine that an AI "government" can decide that humanity is a threat or an obstacle to the AI planned world or, at least, that humanity is not necessary for the further development of the AI governed world. Can we omit this evil scenario of singularity at all?

7 Chalmers, 2010; Eden & Moor, 2012; Kurzweil, 2005; Legg, 2008; Shanahan, 2015; Vinge, 1993; von Neumann, cit. by Ulam, 1958
8 Chalmers, 2010; Eden & Moor, 2012; Kurzweil, 2005; Legg, 2008; Shanahan, 2015; Vinge, 1993; von Neumann, cit. by Ulam, 1958

A possible solution is offered in the possibility of bio-digital fusion in the process of evolution[9]. This solution depends on the capacity of humanity to integrate super AI with the human brain in order to retain development of super AI under the control of the human mind and ethics. In this perspective, the future man or woman is considered as a "cyber-man", equipped with the most advanced superintelligence. Thus, no branches of superintelligence can escape the control of cyber-humanity.

Returning to the less remote perspectives of AI development, contemporary societies should take less distant AI challenges very seriously and an ethical consideration of these challenges is imperative. The creation of AI is probably the biggest event in human history. In principle, the use of AI should be beneficial (for example, we can use artificial intelligence to eradicate poverty and hunger from the human race), yet it will change the world by great and complex challenges, which must be successively coped with. On the other hand, the unethical use of AI systems can cause unprecedented destruction and damage. Thus, respect for ethical standards and the maintenance of human control must be accomplished in the future of an AI-enriched society.

9 Kemp, Hilbert & Gillings, 2016

PERSONAL DATA AND PRIVACY IN THE AGE OF AI
BY KASIA SÖDERLUND

The pace of development of artificial intelligence (AI) and machine learning (ML) increased rapidly in recent years and many of us wonder when we actually started to use technology from the future and when science fiction became reality. We see countless ways of how useful AI can be and how many global problems it may help us to solve. But in all the excitement about the new technology we must not forget that we need to take all the precautions we can to mitigate the risks that development of AI poses. The discussions concerning regulation of AI in the European Union are tense and our lawmakers do not have an easy task to set the rules for all the rights and interests involved.

IS PRIVACY ONLY A PRIVATE MATTER?

One of the growing concerns with AI is its intrusiveness into our privacy in the online sphere. Smart devices that we use daily provide a constant flow of data, which is duly collected and analysed elsewhere with surgical precision by sophisticated algorithms. The result is that the receivers of such streams of personal data are able to measure us in many dimensions and create very precise projections for us. These profiles are mostly used for commercial purposes to show us personalised advertisements or more relevant content on social media.

Yet is it really safe that such precise projections of ourselves are at the disposal of hundreds of organisations around the world? How can we trust that profiling will not be used to exploit or manipulate us in ways of which we are not aware?

Indeed, there are considerable risks involved with this massive outflow of personal data. Datasets containing personal data are often targeted in cyberattacks and are subject to leaks. But, perhaps most worryingly, digital profiles of large groups of people can be used in shaping public opinions and for political purposes. This might sound too dystopian and unrealistic until we realise that this is precisely what happened in the Cambridge Analytica scandal, with its involvement in more than 100 election processes worldwide. Since, probably, we are not aware of many other cases of the misuse of our data, there is no doubt that we should take the issues of privacy and personal data protection seriously.

GDPR VS MACHINE LEARNING ALGORITHMS

The EU legislator recognised the need to update the law governing personal data processing and adopted the General Data Protection Regulation (GDPR), which entered into force in May 2018. The Regulation unified the law governing personal data processing and generally increased the level of protection of personal data in the European Union. And possibly most importantly, it introduced a mechanism for imposing high fines for infringements of the Regulation, which has already affected dozens of organisations.

However, many businesses applying the more complex and sophisticated ML algorithms are still struggling with certain requirements of the GDPR. For example, the Regulation requires that processing of personal data be limited only to what is necessary to process in relation to the clearly specified purposes and no longer than is necessary (principles of data minimisation, purpose limitation, and storage limitation). These principles are directly in opposition to what such algorithms need to become robust, efficient and accurate: vast amounts of data to learn and to perform optimally.

Another difficulty in compliance is that opaque ML algorithms are not transparent for humans in the way they operate. While transparency is one of the central principles in the Regulation, insisting on simplification of such algorithms inevitably leads to deterioration of their quality. Transparency is particularly important in respect of algorithmic decision-making and profiling which has legal or other significant consequences for individuals. In such cases, according to the guidelines to the Regulation, individuals should be provided with the possibility to review such decisions by someone who can assess all the relevant data and who has the appropriate authority and capability to change the decision. Conformity with this principle therefore seems to be a considerable challenge for the ML industry, as the complexity of ML algorithms is the very reason why they are so effective. However, there are some promising solutions proposed, such as intelligible or explainable AI (XAI), which might provide enough safeguards to satisfy the requirement of transparency.

BIASES IN ALGORITHMS AND WHAT WE CAN DO ABOUT IT

Since algorithms are increasingly used in many fields, news concerning biased or discriminating algorithms is unfortunately a regular occurrence. Contrary to what had been expected, algorithms emulate human biases and make discriminatory decisions on grounds of race, gender, age, religion or sexual orientation.

However, we should first consider the reasons behind their bad decision-making. Since algorithms are usually trained on the historical data, what such results demonstrate is in fact our own past history of biased decisions. However, in contrast to the human decision-makers, algorithms can be subject to scrutiny and rigorous testing in respect of biases before they are used. In fact, data scientists have already created special algorithms that examine other algorithms for their possible biases. Thus, instead of accusing algorithms of discrimination and perpetuating social inequalities, we should rather use algorithms to improve our own impartiality in decision-making. All in all, the GDPR does not prohibit the use of ML algorithms, it only requires that personal data be processed fairly and in a transparent manner, regardless of means.

CONCLUSIONS

The widespread application of artificial intelligence and machine learning has imposed a considerable strain on rights to privacy and the protection of personal data. Many of us find it disappointing that it is virtually impossible to use smart technologies without compromising our online privacy more than is necessary.

While the adoption of the General Data Protection Regulation was a step in the right direction, there is still a long way to go. The good news is that setting the compliance bar higher did not result in stalling the development of machine learning in the EU. Some provisions of the GDPR seemed to be a dead-end for many applications of ML algorithms, but the industry worked hard to meet the requirements of the Regulation and many solutions to the existing problems have been found.

Still, the EU legislator faces the challenge of calibrating the legal environment in such a way that the digital technologies are developed and applied, but at the same time the exercise of fundamental human rights is not precluded. After all, the technology is here for us, not the other way round.

THE AI REGULATION WE NEED IS ABOUT CHECKS AND BALANCES, NOT ABOUT POLITICAL AGENDAS

BY ALEKS JAKULIN

AI RISKS

In popular fiction, AI is often portrayed as enemy robots out to get us. That's an incorrect interpretation of the technology. The real interpretation is that billions of people, our institutions and companies, are vulnerable to mass surveillance. Furthermore, mass automation gives huge power to a very small number of people, creating a vulnerability to mass control. There are as yet no checks and balances that could prevent a disaster.

Back in 1958, the American comedian Bob Hope told a version of this joke after a visit to Moscow: "In America, you listen to the radio. In Soviet Russia, the radio listens to you." Except that now, 3.3 billion people have a smartphone, each with a microphone controlled by software and connected to the Internet. Each smartphone is a listening device. Business, personal data, even government data are stored on the same infrastructure.

The majority of smartphones are controlled by software from just two companies. We know very little about what is happening there: the software is proprietary, incredibly complex, and can change at a moment's notice without anyone knowing. There are virtually no checks and balances yet. Instead, our AI regulations merely define penalties after the fact. By then, it would be too late.

A realistic metaphor is that we're living in a city made of paper. It just takes one match. We need a fire code that limits and prevents casualties in case of a fire. We do not need an AI ethics discussion that regulates how the paper should be ethically and inoffensively folded. It's the flammability of paper that's the problem. We need to limit the amount of paper in one place, ensure sufficient space between paper objects, put fire extinguishers, hydrants, and hoses in place, create fire brigades, and implement fire inspections. To understand how this applies to AI, replace paper with data. In addition to a fire code, another example of a regulation is a constitution: it allows a country to continue operating even if a particular politician or political party goes breaks down. It slows down the fire by means of the checks and balances required by someone to exert power.

To apply this thinking to AI, we should establish the separation of data, control, and access: along with audit trails and approval mechanisms that operate at a societal level.

The two countries that have the most advanced AI control are Russia and China. While available information is limited and to some extent a confidential matter for each state, it appears that Russia is testing separating its internet from other countries, a fundamental check and a balance mechanism. Russia also has a point of control for international communications, Roskomnadzor.

China has both of those, and has made an extra step. It has placed government officers within companies that process large quantities of data and it implements active information control and surveillance. The inspectors/regulators are

thus full-time and embedded in organizations, in contrast to external inspections and audits. Moreover, staffing and process are external and defined by the government, in contrast to relying on internal company staffing.

Western societies have historically relied on market competition to establish sufficient diversity to prevent catastrophic failures. The market has instead been allowed to become overly consolidated within a small number of so-called Big Tech companies. Since the market is no longer competitive, the regulators are moving in.

AI REGULATION PATTERNS

It is of the utmost importance that the development of regulations is led by staff who are intimately familiar with the technology. Nonetheless, there are a few cognitive tools that can help regulators evaluate regulatory proposals. This section will introduce a number of best practice patterns.

Security Thinking

When developing important systems, we need to assume that an external adversary will attempt to benefit by attacking the system. Any control measure can be expected to be broken.

Regulations based on liability will come in too late: after the damage has already been done. Moreover, executives often shield themselves from information that would make them liable – or they are simply not compensated well enough to assume the maximum level of responsibility.

Thus, regulations should seek to limit the maximum impact of an adversarial attack. For example, India set a maximum reach of any individual message sharing to prevent the spread of false information. Unfortunately, this measure is easily defeated by manually or automatically generating variants of the same message, making the regulation ineffective from the beginning.

Security thinking is rarely developed among regulators but is a basic tenet of military strategy and cybersecurity.

Enforceability

Even if regulations might be well-intentioned, they need to be enforced effectively. We need to assume that an adversary will defy regulation and seek to profit nonetheless.

A regulation that is regularly violated can be seen as unethical: it places a cost on all the compliant parties, who bear the cost. At the same time, adversaries pay no cost of compliance, while actively benefiting from violations. When violations are not easily discovered, when evidence isn't collected, and when violators aren't penalised, such regulations hurt society.

An example of limited enforceability is the GDPR Directive. For example, an EU company is not allowed to privately collect and process personal images of EU citizens. A foreign company can do this with fewer consequences if it is operating outside EU jurisdiction. As a result, EU citizens are not protected from foreign data processors, while EU companies are less competitive. The EU has few ways to even audit such a matter.

Cost Consciousness

Some rules are so complex to understand and implement that their net benefit is negative for society.

An example of the failure to consider cost is the EU Cookie Directive: internet users have to provide endless consent solicitation even for the harmless uses of cookies that include shopping carts, affiliate links, or for counting the number of unique visitors. As an unintended consequence, the independent internet advertising industry in Europe has been decimated, giving an enormous advantage to a small number of foreign companies that were able to force users to consent.

Implementing GDPR has also proven to be exceedingly costly, where many compliant news organisations have simply decided that they will terminate their service to Europeans faced with the complexity of implementing full compliance, even though very little actual information collection would have taken place. It is an attractive regulation for which to lobby: it creates a lot of opportunity for those who "help" organisations to comply.

Matching Modes

Digital systems operate at the speed of light, whereas humans move at the speed of our minds and feet. It's not too different from a policeman trying to control road traffic on foot. The police need fast cars and means to monitor and control traffic.

If we want regulatory systems to be effective for AI systems, we cannot rely on traditional legal review. Instead, the regulations themselves need to be implemented as digital code that has adequate oversight over the underlying systems, but with sufficient checks and balances that limit both abuse of power by inspectors or an attack on regulatory systems.

Containment

To simplify regulation, we could employ the adage of "divide and conquer". Overly complex and overly powerful systems can be broken down into modules. Modules are controlled by different parties. Modules communicate with each other in a well-defined way.

Almost every digital camera is a computer with full access to the network: it can do whatever, and whoever, can replace its software. While general-purpose computers are powerful, it makes more sense to limit the capabilities of certain devices. For example, a camera can record the video. The video can be transmitted to a specific device through a specific network protocol. Because of its design, a camera cannot do anything else.

Personal information should perhaps not be stored with companies at all. Instead, it should be collected and shared as needed by specialised personal data stores, for example separating the address from the contents of the message, as we're used to with post office envelopes.

Containment is a fundamental tool for creating robust and secure information and computation infrastructure.

Standardisation

Standards establish interoperability between organisations and devices. It was communication standards that established the World Wide Web and internet protocols. These standards have allowed thousands of companies and billions of devices to connect.

If we establish standards that will facilitate interoperability at the data and logic level between companies, individuals, organisations, governments, and regulators, the regulation will operate through and on top of such standards, creating a considerably more manageable system.

Standards will also reduce the cost of collaboration between smaller organisations, reigniting the digital economy that's currently trapped within the Big Tech.

Certification

Our devices and medicines need to be tested and certified before they can be sold. We already have certification for communication devices. We need certification for any kind of data collection and surveillance, for personal data access, and similar. To make certification tractable, we need to establish standards and containment strategies.

Bottom-Up Thinking

It is attractive to attempt to create an all-encompassing piece of regulation top-down. It rarely works. As with all natural things, new technologies start small, and if they work, they grow. When it becomes clear what works, standards emerge, and they can be followed by certification.

To proceed, we should focus on the most threatening types of AI and data failures and attacks. Removing one threat at a time is the way to move. At the time of writing this (end of 2019), the most worrying are:

- massive Big Tech data stores
- our reliance on Big Tech to provide news and search results
- our dependence on Big Tech to provide computation and storage
- uncontained IoT and mobile devices capable of surveillance and disruption
- the difficulty of verifying information

Ownership

AI can be seen to be somewhere between a machine and a wild animal. Like wild animals, AI carries a certain amount of risk and unpredictability. The owner should be responsible for keeping the AI in check and assuming the necessary precautions. The concepts above can help us categorize, regulate, and contain different AI systems based on the maximum damage they can cause.

Safety

AI is a new technology, so many claim skills, but they might not necessarily know what they're doing. There is abundant sloppiness in how data is collected, validated, processed, and stored. Much like food safety, we need data safety.

Moreover, the algorithms need to be methodologically sound, and we can check this the way inspectors verify food preparation and storage: AI model safety.

AI systems implementing essential functions for society need to be tested and certified by organizations that practice security thinking and stress-test an AI's performance in adversarial circumstances, for example analysing the contamination with unsafe data, or checking the AI outputs on a regular basis.

CONCLUSION

A successful approach to AI regulation will prioritize large risks over small partisan issues, and clear-and-present dangers of data leakage and device control over the impact of hypothetical technologies that are still decades away. Moreover, it will engage AI practitioners over outsiders.

I have attempted to illustrate the dangers of the digital world we have created over the past two decades. I have shown how Russia and China in many ways control AI dangers more effectively than the West. Finally, I have provided a set of patterns for analysis of regulatory proposals.

Europe has pioneered if not capitalized on several of the most transformative technologies of the past century: the GSM cell telephony standard now adopted across the world, and the Web standards. We now need to formulate and define the institutions, infrastructure, and legal practices around data. AI is dependent on data, so by regulating data, we will begin regulating AI. Instead of the Chinese model of embedded control, we should implement data infrastructure that lowers both risk and cost of security.

The data regulation should work both from bottom-up in terms of regulating devices, data, and platforms, as well as top-down in terms of managing government data. There is an urgency to implement a bottom-up approach to ensure security and robustness: there are significant societal risks. As for the top-down approach, the data controlled by the government can be stored and accessed in a better way. There is no best model yet, and the strength of the EU lies in the ability of each country to attempt a slightly different approach. One of the existing early examples is the Estonian X-Road system. After about a decade of such divergent work, the best-performing paradigms can converge and serve as broader standards.

GLOBAL AI: THE ROLE OF THE EU
BY NINA PEJIČ

The current rapid development of artificial intelligence (AI) represents a powerful tool, not only for private companies that are driving progress in the field, but also for the economic, political and military advancement of individual countries, as well as supranational actors such as the European Union. Looking beyond gross domestic product (GDP) enhancement,[10] the potential applications of AI are extremely diverse, making it the key technology of our time. Countries that choose to opt in to this development aim to develop optimal ecosystems for the research, usage and advancement of AI technologies.[11] Until recently, their predominant approach was focused on providing funding opportunities, especially for research; now they are setting regulatory boundaries, providing direct incentives and needed infrastructure and, in some cases, immediate guidelines to industries.[12] Governments across the globe are now shifting up a gear and taking a more active role in enhancing their AI ecosystem to further their gains from the 'new technology revolution'.

The United States (US) has historically left technology companies unimpeded by oversight or regulation. This approach, combined with a very tech-appreciative consumer market, has been visibly successful: the world's five largest technology businesses - with a combined market capitalisation of over $3 trillion - are all US companies (Apple, Amazon, Facebook, Microsoft and Alphabet). At the research level, the United States remains highly invested in AI and other emerging technologies. The National Science Foundation (NSF) currently invests over $100 million each year in AI research.[213] However, AI development in the US is not primarily driven by state-funded opportunities, but stems from the largest annual net capital inflow in the world and the largest number of AI start-ups, with its AI start-up ecosystem receiving the most private equity and venture capital funding. Moreover, the US is the leader in the development of both traditional semiconductors and the computer chips that power AI systems. There are also softer elements, such as a highly developed mergers and acquisitions culture, access to business infrastructure (PR firms, legal counsel etc.), access to distribution and partnership opportunities, the higher cost of labour that attracts elite talent, experienced management and also a phenomenon of an established 'entrepreneurial culture ecosystem'.[14]

Nevertheless, where the US has established a strong lead in AI discovery, it is increasingly likely that China may dominate the implementation and industrialisation of AI. The strength of China's economy is a synergy between government policies and market forces. Due to the size of its market, China has advanced commercial capabilities in AI, but also a very coherent national strategy. The Chinese government introduced several strategies with regard to AI,[15] with the most concrete being the New Generation Artificial Intelligence Development Plan (2017), in which it outlined the Chinese path to becoming the leading global AI power by 2030. The Three-Year Action Plan for Promoting

10 The global market for AI is projected to reach $37 billion in revenue by 2025 and contribute up to $13 trillion to the global economy by 2030 as this technology transforms more and more sectors. See more at Jacques Bughin, Jeongmin Seong, James Manyika, Michael Chui, and Raoul Joshi, "Notes from the AI frontier: Modeling the impact of AI on the world economy," McKinsey Global Institute, September 2018

11 Stiftung Neue Verantwortung. 2019. Artificial Intelligence and Foreign Policy. Available at: https://www.stiftung-nv.de/en/project/artificial-intelligence-and-foreign-policy

12 FTI Consulting Inc. 2018. The Global Policy Response to AI. Available at: https://euagenda.eu/upload/publications/untitled-128126-ea.pdf

13 Forbes. 2019. Who Will Lead In The Age Of Artificial Intelligence? Available at: https://www.forbes.com/sites/danielaraya/2019/01/01/who-will-lead-in-the-age-of-artificial-intelligence/#6ab5a0236f95

14 Centre for Data Innovation. 2019. Who Is Winning the AI Race. Available at: https://www.datainnovation.org/2019/08/who-is-winning-the-ai-race-china-the-eu-or-the-united-states/

15 All of them are a continuation of the 13th Five Year Plan, Belt and Road plans for digital connectivity, and the industrial Made in China 2025 plan.

Development of a New Generation Artificial Intelligence Industry (2018–2020) introduces measurable indicators, objectives and goals to implement the overall development plan established in 2017. Several offices established under the Ministry of Science and Technologies are responsible for the implementation and coordination of AI-related projects and are intended to foster the Chinese AI ecosystem through several measures.[16] The latter are government-led subsidies and public contracts both at central and local level, loose regulations that facilitate an advantage when it comes to AI applications relying on big data sets, building physical infrastructure that (sometimes completely free-of-charge) supports a start-up ecosystem in large innovation centres. The national plan incentivising local provinces and municipalities is extremely effective due to inherent provincial competition in China.[17] From a cross-border perspective, many top tech companies in China also continue to acquire AI-related technology and "know-how" through notable investments abroad (e.g. the US's NVIDIA partnering with Alibaba and Huawei to build an AI city platform). Moreover, China's top tech companies teamed up with the government to set up China's AI 'national team' to assist with the country's bid to become the leading global AI innovator.

The European Union (EU) has an opportunity to learn from both the advantages and disadvantages of the approaches of the currently leading AI powers. The EU should focus on observing the unintended consequences of both approaches, as well as adopting good practices. In the latter, it should focus on access to a larger pool of (both patient and venture) capital that would support seed stage ideas and later the scaling up process, as is the case in the US. With increased access to risky capital and its existing first-class talent and academic research excellence, it should not be difficult to create new opportunities for its young scientists at home. Public institutional support mentioned in the China example (subsidies) should be increased but developed smartly with strong monitoring processes (nevertheless without great bureaucratic obstacles). We can observe the negative consequence of the lack of control over support funding in China, where the AI sector recently suffered from an allocation of resources, particularly in relation to funding and subsidies, which has hindered their progress. Reports that some leading companies depend on government funding for 30 to 68 per cent of their profits reveals an issue with over-reliance on subsidies – a problem that has historically plagued a number of sectors promoted top-down by the Chinese government.

EU countries should also focus on spending larger parts of their individual GDPs on supporting innovation processes and developing 'self-confidence' in (national and European) AI strategies. European strategies do develop good approaches and ideas that are unique to Europe, but lack systematic indicators and benchmarks that would support the innovative development of AI ecosystem in the region. Looking at China again, the country developed a very specific set of measurable goals and objectives to be reached by different levels of governance – the state level, provincial level and city level, that all support innovative environment. The problem of European strategies is that they lack such a set of more concrete goals and measurable indicators of progress and outcomes (of course based on common European values and existing practices). Such measurable indicators should consist both of input (e.g. policy measures to strengthen the AI ecosystem) and output (e.g. achievement of results) indicators. The question of how to define achievement indicators and how to measure progress is also important, as it puts us in a position to better understand what a strong AI ecosystem means for Europe.

16 Sheenan, Matt. 2018. How China's Massive AI Plan actually works. Available at: https://macropolo.org/analysis/how-chinas-massive-ai-plan-actually-works/

17 For example, the Shanghai government issued its own implementation plan for new generation AI; Beijing announced a major new AI-focused industrial park to be constructed in Mentougou District; Guangzhou launched an International Institute of AI; and many other districts have promised funds for AI research.

HOW TO REGULATE AI?
TOWARDS TRUSTWORTHY ARTIFICIAL INTELLIGENCE

One of the advantages of the EU environment is the strong regulatory mind-set of its citizens and political leadership. Taking into account the US example, this mind-set is the very reason for the positive and trustful relationship of EU citizens towards AI development inside the EU area: while the US has truly made significant advances in AI, these, however, also encouraged negative publicity that influences the consensus on what the national strategy should be and, in the long run, can have consequences for the fragmentation of resources needed for AI development. Nevertheless, when it comes to taking the step of passing laws on the potentially negative impacts of technology, Congress is hesitant, pointing out the inherent insecurity of the regulation environment in the US. The extremely negative consequences of incorrect or absent regulation are also visible in China: in recent years, several domestic companies cheated the system in order to secure subsidies, which revealed that local authorities lack the industry-specific expertise necessary to regulate the market, similarly to the US.

Thus there are several areas of possible improvement which could define what precisely is the EU's '3rd way' in the global AI context. The 'AI revolution' in Europe is perceived as a wave coming from abroad that threatens its social, political and economic model. However, the EU should seek, not to restrain, but to govern AI development – guide it into the strategically and financially thought-through direction that both the US and China often lack.

EPILOGUE

Fear of AI is unnecessary - AI systems are currently quite simple tools, designed to perform (with "superhuman" efficiency) certain (narrow) tasks (e.g. examine data) – which is quite different from broader and deeper AI systems, which are commonly presented as science fiction. But AI does not come without serious risks and it does have a significant impact on our lives, so from a liberal perspective it must be ensured that the development and use of AI is in accordance with our values, human rights, ethical principles and "security mindset"; and that will most probably be achieved by a mixed set of legal rules in order not to hold back future progress: a general framework e.g. "Golden rules" (general rules of principle) regarding the development and use of algorithms (as written above or simplified 1. algorithms should not be developed solely or primarily to kill or harm humans; 2. People, not algorithms, are responsible agents. Algorithms are tools developed and programmed to achieve human goals; 3. The algorithms must be designed to ensure reliable operation and security; 4. It must always be clear who is legally responsible for the algorithm), "smart" regulation (the result of a search for the right balance between regulation, which establishes appropriate legal premises, and self-regulation of the AI industry), certification etc.

Once we have the general framework we can tackle regulation a little more specifically; first we find some flexible definitions (such as in the case of nanotechnology regulation), then we perform a risk assessment and focus on the responsibility for the result (based on due diligence in the development and use of the algorithm); and so we adopt some international guidelines or, in the case of the EU, at least a regional directive, so that Member States can more easily transpose regulation into national law (because not every country can regulate it in its own way). It is somehow clear that some exact "classical" legal rules, or instructions are not always the right approach to solving these issues. It may be necessary to introduce some general compulsory insurance for financial protection against physical and other injuries resulting from a malfunction of the algorithm. The question of transparency remains, because even if (despite trade secrets) you have full insight into the algorithm code itself, it is likely that you would still not know where the problem lies.

People do something simply because they want it; we go through the list of arguments for and against, and nevertheless we can (intentionally) make a "bad" decision. Algorithms are not good or bad, fair or unfair, just or unjust - although a person can put these values into the design of the algorithm and also algorithms themselves can come to discriminatory results because of using (human made) historical biased data – in the forefront it is the responsibility of the human operator in the background; but should he assume also the legal responsibility if the algorithm works poorly, dishonestly and unfairly? The problem with complex algorithms is that they are not always predictable. The code itself can be buggy (mainly because developers are under a lot of time pressure these days and all too often they run out of time for quality control). We certainly already have cases where things went really wrong (Flash Crash in 2010, a United States trillion-dollar stock market crash, or when the Knight Capital Group lost $440 million in 2012 etc.). Should we just penalise poor results without actually regulating algorithms?

In reality, it may not be as difficult as it might seem at first glance. Namely, at least we usually always have a person in the background, a "supervisor" or someone who commissioned, developed and "let" an algorithm into the world. In

addition, we may have "classic" documentation ("memos", "hot documents") that contains critical information about the case and may tell us what the real goals in the background were (such as "destroy the competition"); we may not understand the operation of the algorithm, but we understand those. As, in the end, we also have, after all, the results; regardless of which algorithm was used. We also have some judgments where the companies had to take responsibility for their algorithms (e.g. the German court required Google to remove offensive autocomplete results) and at least indirect legal regulation of the algorithms (e.g. legislation for autonomous vehicles).

It is important to inspect the entire data value chain because a lot of AI systems are built from data sets collected from around the world and they are the core of functionality of those systems. It is not enough to inspect just the functionality of algorithms.

There is a clear and important need for transparency of how algorithms work – to understand what technologies are used (by governments, big companies etc.) and how they affect us, but the key idea is that when designing systems such as self-driving cars there is a key lack of standards or procedures, how to take ethical and moral standards into account at the time of development (which is a key difference to "traditional" engineering). Because these AI systems should in the end take decisions on their own – they are the ones who, in a fragment of a second, choose an action. It is up to engineers developing these systems to encode and enshrine into these systems the logic that guides their actions.

In the end it all comes down to the ethical principles of the society that we decide to live in; it is complicated and it is an emergent property of society, because these rules are not typically something that any society builds just by sitting down and writing them on any specific day, this is a process that takes many generations and is non-universal in a global sense; because from society to society these principles may differ a lot – but these are the only end base for any such decision making, which can then only be distilled through an engineering process into these AI systems. It would also be ethical for AI to remain solely for the benefit of mankind and not to be used to create an advantage over another society.

AI or algorithm regulation itself is basically an experiment where we will not have a "single and definitive solution". It is a complex inaccurate process, which is ongoing, and only when we face concrete problems will we be able to (more precisely) regulate them. But along this process one of the major roles and global values of the EU – also in a modern and technological advanced society – is to defend and secure our fundamental human rights, our democracy and the rule of law – because it looks like nobody else will.

The EU should also continue promoting open science and open data but, on the other hand, it should make sure that its research data is fairly used by other countries and companies and that they also share their data in return.

POLICY RECOMMENDATIONS

- **Down-to-earth approach**
  The real problems of AI in today's society are a lot more down-to-earth than news coverage might sometimes imply. Wild speculation only hurts discussion on how to address these. Related big problems, such as self-driving cars, automatic mass surveillance systems, or autonomous military weapons, need a lot of thought and debate, but it needs to be level-headed and fact-based. Better public understanding of artificial intelligence would not only enable better dialogue regarding these big problems but would also alleviate a lot of them by informing peoples' actions.
- **Strive to technological literacy**
  The principles of AI algorithms used today can be understood using high-school-level mathematics. The payoff would arguably be greater than just mitigating existing problems. These methods can be used by individuals as tools to enhance their lives and work. We should strive towards equipping everyone, especially children and young adults, with the skills for being wary of and handling something so impactful.
- **Trustworthy Artificial Intelligence**
  Trustworthy AI is one of the declared goals of European Commission. An important component of Trustworthy AI is the ability of the system to explain its decisions. The need for explanation increases with the importance of decisions, and also with uncertainty about the correctness of machine decisions. If there is a considerable possibility that a machine's decisions are incorrect or debatable, then the explanation is needed all the more. Understanding a machine's decisions and intentions is also essential in cases where a human and a machine cooperate in solving a task. Methods for automated explanation generation are being developed in the area of AI now popularly called Explainable Artificial Intelligence (XAI). Given the importance of explanation in building Trustworthy AI, more research in XAI should be supported.
- **The importance of ethics in the AI development**
  Today we have AI that can also make decisions and also influence our decisions. Therefore, the issue of ethical AI comes to the fore. When AI makes important decisions, it is important that they are the "right decisions". Therefore respect for ethical standards and the maintenance of human control must be accomplished in the future of an AI-enriched society. It would also be ethical for AI to remain solely for the benefit of mankind and not to be used to create an advantage over another society.
- **Importance of human rights protection in AI**
  It must be ensured that the development and use of AI is in accordance with our liberal values, human rights, ethical principles and "security mindset". The regulation of development and use of AI will most probably be achieved by a mixed set of legal rules in order not to hold back future progress: a general framework (general rules of principle) regarding the development and use of algorithms, "smart" regulation, certification etc. Exact "classical" legal rules, or instructions are not always the right approach to solving these issues. Some flexible definitions should be found, a risk assessment and focus on the responsibility for the result should be performed (based on due diligence in the development and use of the algorithm); afterwards the EU should adopt some regional guidelines, so that Member States can more easily transpose regulation into national law. It may be necessary to introduce some general compulsory insurance for financial protection against physical and other injuries resulting from a malfunction of the algorithm.

- **Protection of rights to privacy and the protection of personal data**
  Many of us find it disappointing that it is virtually impossible to use smart technologies without compromising our online privacy more than is necessary. The EU legislator faces the challenge of calibrating the legal environment in such a way that the digital technologies are developed and applied, but at the same time the exercise of fundamental human rights is not precluded.
- **The AI Regulation we need is about Checks and Balances**
  We should establish the separation of data, control, and access: along with audit trails and approval mechanisms that operate at a societal level. It is important that the development of regulations is led by staff who are intimately familiar with the technology; there are also a few cognitive tools that can help regulators evaluate regulatory proposals: Security Thinking, Enforceability, Cost Consciousness, Matching Modes, Containment, Standardisation, Certification, Bottom-Up Thinking, Ownership and Safety. A successful approach to AI regulation will prioritize large risks over small partisan issues, and clear-and-present dangers of data leakage and device control over the impact of hypothetical technologies that are still decades away.
- **EU should strategically govern AI development**
  Governments across the globe are now shifting up a gear and taking a more active role in enhancing their AI. Where the US has established a strong lead in AI discovery, it is increasingly likely that China may dominate the implementation and industrialisation of AI. The EU should focus on observing the unintended consequences of both approaches, as well as adopting good practices. It should focus on access to a larger pool of (both patient and venture) capital that would support seed stage ideas and later the scaling up process. EU countries should also focus on spending larger parts of their individual GDPs on supporting innovation processes and developing 'self-confidence' AI strategies – which currently lack a set of more concrete goals and measurable indicators of progress and outcomes. The EU should seek, not to restrain, but to govern AI development – guide it into the strategically and financially thought-through direction that both the US and China often lack.
- **Fair use of know-how**
  The EU should continue promoting open science and open data but, on the other hand, it should make sure that its research data is fairly used by other countries and companies and that they also share their data in return.

AUTHOR
ALJAŽ KOŠMERLJ, IVAN BRATKO, JANEK MUSEK, KASIA SÖDERLUND, ALEKS JAKULIN, NINA PEJIČ
EDITOR
GREGOR PLANTARIČ

PUBLISHED DECEMBER 2019

Co-funded by the European Parliament. Neither the European Parliament nor the European Liberal Forum asbl are responsible for the content of this publication, or for any use that may be made of it. The views expressed herein are those of the author(s) alone. These views do not necessary reflect those of the European Parliament and/or the European Liberal Forum asbl.