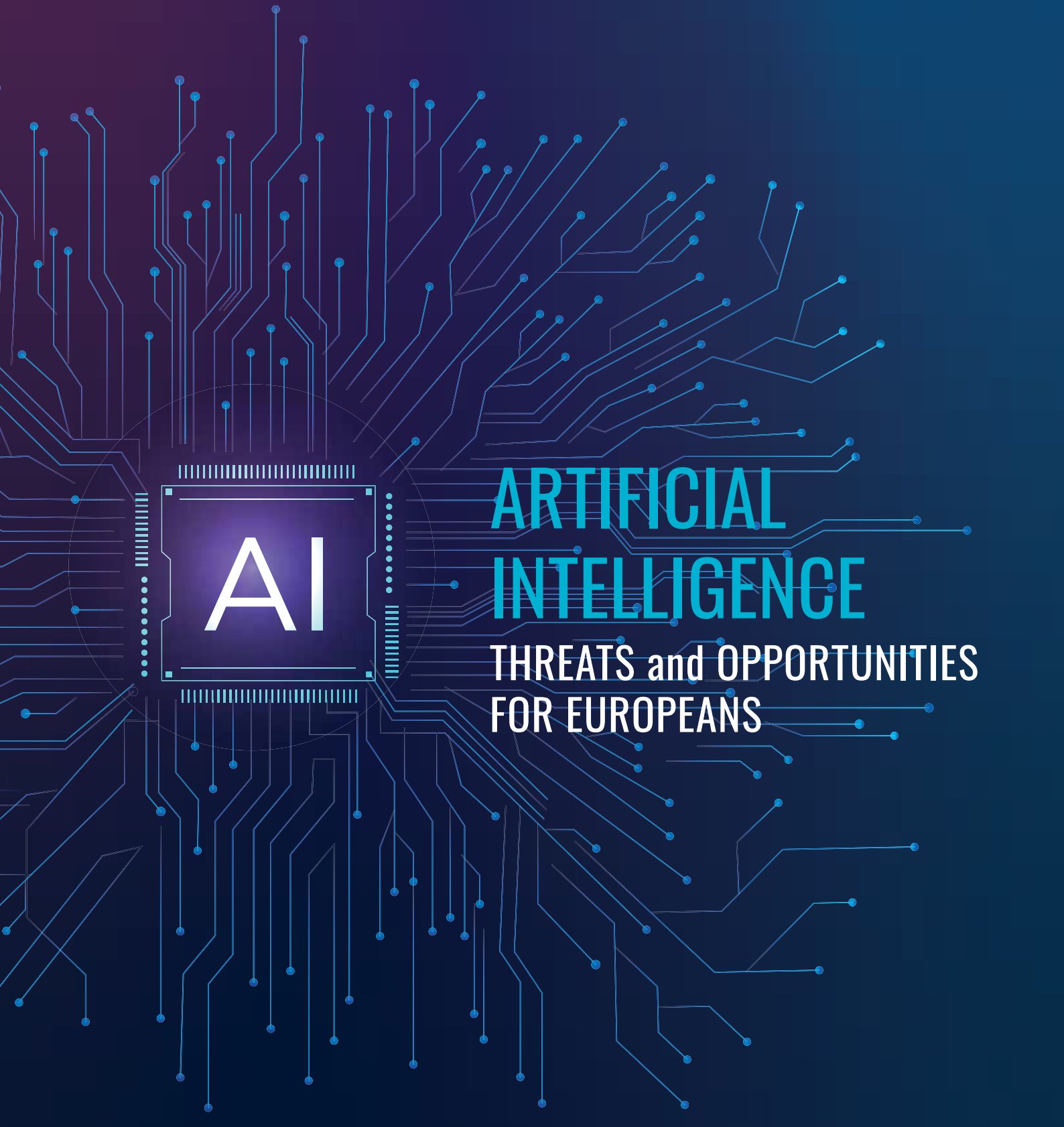
Wilfried
Martens Centre
for European Studies

Hanns
Seidel
Stiftung

СДРУЖЕНИЕ
ГРАЖДАНИ ЗА
ЕВРОПЕЙСКО
РАЗВИТИЕ НА
БЪЛГАРИЯ
ГЕРБ



ARTIFICIAL INTELLIGENCE

THREATS and OPPORTUNITIES FOR EUROPEANS

ARTIFICIAL INTELLIGENCE - THREATS AND OPPORTUNITIES FOR EUROPEANS

by Michal BONI, PhD
former Member of the European Parliament
and Senior Research Associate, Martens Centre

Artificial intelligence is being widely discussed in recent years, both globally and in the European Union. In this discourse, strong expression of both great opportunities of AI development as well as manifold threats for people was raised, especially with regard to the professions where robots and AI would replace humans. AI could take our private lives under permanent surveillance and we could eventually lose control over the technologies.

It is essential to overcome these threats and open the way for building the trustworthy AI.

In the time of pandemic, not only the awareness of possible AI opportunities has increased enormously, but also digitization accelerated. The latter exposed new challenges for people and societies, which are even more acute with the speeding AI development.

The years 2020 and 2021 have been exceptional for AI development in the European Union. The European Commission presented key documents: white paper on AI, the data strategy and the data governance act, and after consultations - the AI package: regulation on a European approach for Artificial intelligence, regulation on machinery products, coordinated plan on AI and communication on fostering a European approach to Artificial intelligence, all based on the concept of the risk pyramid and classification of four categories (from lower to higher risk, depending on their impact on fundamental rights). With that, the basis for AI development was established. The key challenge now is to build new European competitive advantages using AI development, especially in the context of progress on that front in the US, China and other countries hurrying up the AI investments.

There are four crucial dimensions for AI growth. The first one relates to the investments and business opportunities, especially with regard to the accessibility of SMEs to the useful functionalities of AI. The others concern: data infrastructure, regulatory framework with focus on risk problems and ecosystem of trust.

Investments

The basis for scientific and research efforts and effective commercialization of works on AI (which does not mean to be out of public purposes and common goods) will come predominantly from Horizon Europe programme, and many other European financial sources from the new budgetary perspective for 2021-2027. The key issue is to care about the equal level of development and open accessibility to AI services all over the EU and for all entities.

In that context, small business should be provided with special accessibility guarantees. Should SMEs pay for (some) AI services or get them for free in the public domain – remains to be decided. In addition, AI can have a pivotal role in boosting SMEs digitalization processes and become a new driver for their scaling up and development. It requires a redefined support for SMEs, with accessibility of high-quality network, data, computing possibilities and skilled and competent workforce.

Data

The only way to guarantee a high level of AI functionalities is to provide for data accessibility and data training. There are many types of data: from sensitive personal (under the GDPR and e-privacy frameworks), industrial and sectoral (agriculture, transport, health etc.) To complex data necessary

for the autonomous cars driving, for instance. Some of them are structured (ready to use), some require structuring processes. Data should be accessible but their use must be governed by a set of clear rules, such as fair principles (findability, accessibility, interoperability, reusability). In addition, the use of data for AI setup and AI functioning requires transparent principles and human based approach, which are crucial for data control.

Interestingly, what is currently discussed as data control challenge was previously a problem of the data ownership. Because of this mindset shift, the data control model can lead us to the sharing data patterns.

For many aspects of the data control models, the relation to fundamental rights is crucial. It unlocks the ethical dimension of the data and the AI. If we want ethical AI (not discriminatory and not based on biases), we need to implement this dimension in all AI based products and services. There is also an additional important aspect: all those solutions should avoid the establishing of the "black boxes", meaning completely non-understandable AI functionalities, fully hidden to users. To arrive there, we need to be able to explain to different types of users, the paradigm of the AI working schemes through the rule of explainability and interpretability. This is to overcome the information asymmetry (algorithmic sequences between users and AI or the algorithmic developers and deployers), a hazardous situation from the human, citizens and consumers perspective. All those aspects were raised in the proposal of the regulation – and some solutions have been presented, as requirements addressed to the model of the market entry.

Equally important for AI development are the data sharing schemes. They require adequate portability solutions between EU countries. This is key, for example, for the healthcare area, in order to ensure the flow of data and to establish European health data space. Such data space must be based on common political commitment because the health data and health systems are under the nation-

al competences. What is more, such portability and sharing data solutions are also needed for the use of the data from the third countries. Here, we need to establish clear rules of data transfer to and from the countries outside of the European Union.

We can already see a strong European effort to establish proper rules for data sharing services providers – data intermediaries with obligations of fairness, impartiality, secure storage etc. They will be essential for the conditions of data flows, retention and collection. All this requires an additional response: the European data cloud establishment.

Generally, it all means that for AI development the most advanced, transparent and technically confident data infrastructure is necessary.

Risks

Why are we discussing about all those rules for better and transparent functioning of the AI?

Because we do not know exactly what kind of unintended consequences, the AI development can bring. If we want to avoid threats, we need to understand more and to have an adequate regulatory framework focused on the risks. To arrive there, we need to first analyze the potential of harm in material and non-material dimension (i.e. Mental harm of AI incidents) and establish the risk approach in the legal and institutional framework. The risk assessment is indispensable, but the efficiency of this solution depends on the impact assessment applied timely in the preparatory works on AI functionalities.

In addition, we have to also distinguish what are the risky and non-risky applications of AI, as it is suggested in the AI regulation proposal.

In the debate around the AI white paper, the necessity of ex-ante model implemented by companies and AI developers was raised (which may require more time before starting the business). Who should perform the ex-ante analysis and conformity assessments is widely disputed. Business side prefers companies to perform a self-assessment. The consumers and civic organisations propose human rights impact assessment (HRIA) or ethical technology assessment (ETA) prepared by external reviewers.

In the legislative proposal the requirements of ex-ante self-assessment (using various tools) and clear schemes for conformity assessment were presented. The discussion should be continued as we need to find a balanced model, combining the respect of humans and consumers rights and innovation needs. This is the way to the trustworthy AI.

Trust

Trust should be built on the transparent principles uniting fundamental and consumers' rights.

All users ought to be properly informed about the significant roles of AI (possibilities, opportunities, consequences, side effects, benefits) and rules, which govern them: from data accessibility and governance, explainability schemes to the consumers' procedures as clear redress mechanisms guaranteed in law.

The trustworthy AI requires an ecosystem. This ecosystem requires several key components:

- a transparent assessment model based on risk analysis related to the fundamental and consumers rights,
- a requirement of explainability and interpretability of the functioning technical and practical mechanisms (avoiding the asymmetry of information),
- a holistic and democratic oversight of AI development guaranteed by collaborative work of all partners (academia, business, civic and consumers organisations, policyholders).

Conclusions

Working in the EU on the AI development, we are not at the starting but at the turning point. Public and professional debates and decision-making processes are highly advanced.

There are five crucial steps to deliver advanced AI.

Firstly, we need to empower all partners to work together on AI applications (secure, safe, under the ethical umbrella).

Secondly, we should not limit all discussed solutions to the European Union borders.

Thirdly, it is critical to seize the opportunity of establishing standards and norms for AI development, which will be global. Obviously, it means that EU can play a norm setter role globally.

Fourthly, it is important to find the most adequate way for AI investments, joining the private and public efforts (also using the european financial sources).

Finally, we have to develop the AI digital literacy for the European society to really comprehend the opportunities and benefits the AI can bring to all of us.

WHY BIG DATA AND ARTIFICIAL INTELLIGENCE WILL CHANGE EVERYTHING

By Dirk Helbing (ETH Zurich / TU Delft / Complexity Science Hub Vienna)

We are living in a time of Big Data and Artificial Intelligence. As a consequence, we are seeing unprecedented opportunities for business, science, health, and governance. In fact, there are few business fields that have not been changed or reinvented in the wake of the digital revolution. However, every area of society seems to be undergoing disruptive changes as well. "Move fast, break things" was once the motto of Facebook, and probably large parts of the Silicon Valley as well. However, did we take enough time to reflect what we were doing, and what would be the undesired side effects? Have we really built the future we would like to live in, or have we lost control? The following paragraphs will summarize some of the issues, challenges, and threats of digital societies, which appear to require more public debate as well as political and legal action.

Big Data

The world of today is under constant surveillance. Our entire planet, not just the economy and natural resources, but also all our lives are being measured, recorded and evaluated. Since the revelations of Edward Snowden, we know of mass surveillance as a fact, but it is still difficult to imagine the sheer amount of data amassed about the world. There are several reasons for this, e.g. "to save the world", "for security reasons", or "knowledge is power".

Surveillance Capitalism

"Data is the new oil" is a further reason, which is the driving force of "Surveillance Capitalism". It pays off to know a lot about everyone, and in a time of personalized products and services, people and their data are now to a considerable extent the product by with companies earn money. We are not only being watched in public, but also at work, at home, in our living and sleeping rooms. There is barely any secret about us, if sensors of smartphones and the Internet of Things are being used to judge us. Statements such as "Privacy is dead, get used to it", "You are already a walking sensor platform", and "We know, where you have been, and we more or less know what you are thinking about" are characterizing this new reality well.

Digital Crystal Ball

Given so much data, it has become possible for secret services, the military, and big IT companies to build something alike a digital "Crystal Ball", which allows to see what is happening in almost every corner of the world. Using predictive analytics, it is even possible to look into the future to some extent. The probably best-known company offering services of this kind is Palantir, but it is by far not the only digital Crystal Ball that exists today.

Profiling

Data about us is now being collected by thousands of IT companies, which we typically do not even know by name. The estimated data volume stored about every one of us every day exceeds several Gigabytes (corresponding to several photographs). This allows some companies to create pretty detailed profiles about us, containing data about our home, income, consumption patterns, social network, psychological traits, health, and more.

Digital Doubles

In some cases, personal profiles are so detailed that they are considered to be digital doubles or digital twins. Their behavior in response to certain information and environmental stimuli may be simulated in a computer, i.e. brought to virtual life (so-called "avatars").

World Simulation
Going a step further, there are even attempts to simulate possible futures of the world, using about 8 billion digital doubles. One of the platforms used for this is called Sentient World.

"Benevolent Dictator"
The purpose of such world simulations is to determine a proposedly "optimal" future of the world, and to implement it then. This approach often comes under the title "benevolent dictatorship".

Targeting and Behavioral Manipulation
The implementation of a particular scenario of the world requires that every person plays its given role to avoid a deviation from "the plan". Therefore, such attempts to steer the future of humanity are based on behavioral manipulation. This is achieved, for example, by targeting, i.e. influencing the attention, opinions, thinking, emotions, decisions and behaviors of people through personalized information.

Citizen Score and Behavioral Control
The Chinese citizen score known under the name "social credit score" goes a step further by applying behavior-based incentives and punishments, using mass surveillance.

Digital Policing
Digital policing, however, is also increasingly being used in Western democracies, e.g. under the name of "predictive policing". This is particularly disturbing in view of extremely high error rates and the systematic discrimination of certain groups of people. Generally, however, it seems that the statement "code is law" reflects our new reality quite well. Illegitimate "social engineers", whose names are typically not known by the public (nor by scientists or politicians), are now shaping our future.

Cashless Society
An alternative way of controlling behavior by algorithms is through digital money. This concept is known under the label "cashless society". Using blockchain technology, for example, it would not only be possible to detect money-related crime and corruption, but also to prevent certain kinds of purchases and bookings, e.g. airplane flights (say, to enforce a reduction of CO2 production).

Reading and Controlling Minds
The digital revolution does not stop here. In the meantime, there are new emerging neuro-technologies, so-called non-surgical human-machine interfaces, which will allow AI-systems to interfere with human thinking directly. Such technologies are based, for example, on nano-technologies such as nanobots.

Neurocapitalism
In view of this technological breakthrough, some people expect Surveillance Capitalism to be followed up by Neurocapitalism. In this new form of economy, companies would not only more or less know what people think, feel, and do – they could even control it.

Transhumanism
The development of neurotechnologies is part of the agenda of transhumanism. According to it, humans will upgrade themselves with technical devices and become cyborgs. Eventually, humans will even become indistinguishable from intelligent machines. This process is called human machine convergence. Transhumans would upload their memories and personality to the cloud (see the discussion of digital doubles above), while normal, biological humans as we know them today may go extinct. Brain activity mapping/brain indexing is, in fact, expected to be a huge future market.

Singularity

According to transhumanists believing in "the singularity" (artificial superintelligence), humans will be succeeded by superior machines. Moreover, one day, a superintelligent system may control world affairs, machines, and the behavior of individuals like a "digital God".

Algorithm-Based Dying and Killing

Transhumanists further seem to believe that humans are something like biological robots. In contrast to robots, however, normal humans are considered to be weak. They get tired and ill, and they are blamed for overconsuming the resources of planet Earth. Therefore, some people think that humanity should be controlled, and the overpopulation problem "solved" by computer-based eugenics or euthanasia. It is shocking that, in some places, Corona triage decisions are already taken by AI-based system.

Technological Totalitarianism and Digital Fascism

Altogether, there seems to be a real danger that societies around the world could end up in one kind of technological totalitarianism or another. Some people also see the emergence of a form of digital fascism, characterized by features such as the following: mass surveillance, profiling and targeting, unethical experiments with humans, censorship and propaganda, mind control or behavioral manipulation, social engineering, forced conformity, digital policing, centralized control, different valuation of people, messing with human rights, humiliation of minorities, digitally based eugenics and/or euthanasia.

An Alternative Future

Even though the above development is often characterized as an alternative-less, technology- and data-driven future, visions of alternative digital futures have been worked out. Some of the cornerstones of this new framework are:

- Value-sensitive design (i.e. a particular design of digital platforms such that constitutional, human, social and cultural values are supported),
- "peace rooms" rather than war rooms (being more transparent, interdisciplinary, multi-perspective, ethical, and participatory),
- a digital upgrade of democracies (such that collective intelligence is boosted),
- a platform for informational self-determination (allowing one to control personal data and digital doubles by means of a digital assistant, i.e. a local, personal AI),
- City Olympics (a friendly competition among cities and regions to mobilize the power of civil society and address the world's problems such as sustainability challenges),
- democratic capitalism (making money a property and instrument of The People and introducing a recurrent investment premium to facilitate "crowd funding for all"),
- a socio-ecological finance system called "Finance 4.0" (promoting the co-evolution of a circular and sharing economy.)

It is still up to us to decide in which digital future we want to live, but we should act quickly…

ETHICAL CHALLENGES OF AI RISKS

by **Jana Mišić, PhD candidate, University of Twente and Rathenau Institute, The Hague**

Not all risks are equal. Artificial intelligence systems pose a myriad of ethical questions in their use that ought to be central to the risk analysis and risk classification systems of AI in the European approach.

Avoiding AI risk, uplifting people

The European Commission is working hard on the European approach to trustworthy Artificial intelligence (AI). The crown of these efforts, the so-called AI Act, introduced a sophisticated risk-based framework around four risk categories. The risk-based pyramid is a layered enforcement mechanism that builds up from minimal risk of automated systems to limited risk, then high risk, and finally explicitly prohibited, unacceptable risk category of AI applications. To enforce such categorisation, a range of mechanisms will be at disposal: non-binding soft law impact assessments, codes of conduct, external audits and strict compliance throughout the system life-cycle. The main strength of the European framework lies in embedding humanist norms, values and fundamental rights into how we think about risk. Such explicit commitment should be applauded.

Regardless, thinking about risk remains a risky business. This is partly due to the difficulty of determining when something becomes too risky. Or classifying an AI employment software as low or high-risk. The severity level differs between what the system is used for (the first screening of candidates or autonomous reliance on the system without the human in the loop) or the context of use (the police or a retail store). Risk comes due to many causes. A technical variable would be the quality of the system, the integrity of the designer, transparency and safety. Social variables are also important. Do some social groups feel more at risk than others? Do specific AI applications disproportionately affect the more vulnerable such as the elderly? Or do we think that particular uses of AI (such as to confirm our identity or determine our employability) should always be done solely by humans?

Determining low or high risk of AI is already a challenge today. But, because of its complexity and numerous possible uses, AI systems also carry a great deal of risk in their potential future applications. False results, repurposing of the system from one context to another, patent secrecy, technological dependence and the lack of transparency are only a few dangerous aspects of AI.

The challenge ahead is the ability to understand AI risks better. Although it seems self-evident what potential AI risks exist and that we need to assess them properly, the question of applying a moral compass to risks can be problematic. Whether something is too risky or not often leads to stalemates, and technological risk is no exception. Therefore, how can we think about acceptable and unacceptable risks from a moral point of view? One way would be to deepen our understanding of risk ethics and to distinguish important ethical aspects to be considered for the ethical risk-benefit balance.

What is ethics of technological risk?

Technologies always carry a certain amount of risk. This is particularly true for emerging technologies, like AI, that are still in the research and development phase. But what is technological risk exactly? Some technologies can lead to pollution, accidents, physical damage, or mental health issues. Controlling risks means controlling the potential of a damage (a hazard) from occurring to the best extent possible. Once the potential hazard and the probability of it happening are known, we can

calculate the risk of that hazard occurring. For example, we can calculate the risk of being falsely identified by an AI facial recognition system.

Technological risk is often defined as a probability of an unwanted effect that is a consequence of a specific technology. If the likelihood of damage is unknown, we must deal with technological uncertainty. If the effects are also unknown, we are also dealing with technological ignorance. Uncertainty and ignorance are then central aspects when determining AI risks. Still, both factors are purely statistical information. Knowing that we have a five per cent chance of wrongfully being rejected for a job due to an automated error system, are we right to accept the risk? Some people would say not. It becomes clear then that statistics are not enough for a proper risk assessment. People's feelings, the greater good, and our desired technological future are all non-statistical factors of risks that we care about.

Ethics of risk or introducing moral philosophy can help policymakers. To be ethical, assessing risk can never be a fully quantifiable task. From a moral point of view, even if there is only a five per cent chance of harm happening, each human has the right not to be exposed to such risks due to the actions of others. In practice, however, our technological development inadvertently impacts humans, the environment, quality of life and the expected future. If the European approach to AI, based on the risk pyramid, is to stand the test of time, it should not be blind to the guidance coming from risk ethics.

Make ethical analysis a part of risk analysis

Risk ethics questions under which conditions it is morally acceptable to expose someone to risk. What types of risk are ethically acceptable? Not only statistical but also individual emotions and societal perception play a role in answering these questions. Most policy is based on the principles of consequences, rationality and maximum utility. The goal is to achieve maximum benefit and minimal risk for all. However, these principles can best be followed in the case of mature technologies where uncertainty and ignorance do not play a role. For example, the probability of and the effects of engine failure in an airplane are reasonably known, and we can talk about rational, calculated risks. The same cannot yet be said for AI systems. This was quite visible once algorithms were used in the social security system to predict welfare scams. Vulnerable groups of citizens were falsely accused of embezzlement. Significant harm has happened before welfare AI is categorised as an unacceptable risk. Risk ethics shows that the cost-benefit analysis is too simplistic. Important ethical aspects to be considered for the moral risk-benefit balance are informed consent, available alternatives, and value trade-offs for a fair distribution of risks and benefits.

Informed consent
Informed consent shields human autonomy and freedom. It protects these central values and states that for risk to be ethically acceptable, the human at risk should be well informed about the potential harm and freely and autonomously consent to be exposed to it. Informed consent is well established in ethical medical practices, protecting individuals in trials. It is a staple of ethical medical innovation. Applying it to AI risks, however, while necessary, is particularly difficult because technological risks are often collective and impact even those who never consented. For example, informed consent of each citizen affected by facial recognition analytics in public spaces has never happened.

Available alternatives
Some risks are inevitable due to the lack of alternatives. When assessing risks, one should not only focus on one technology but look for the best available technology. In line with the facial recognition example above, we could consider whether deploying real-time facial emotion recognition to keep public spaces safe is the best available option if the software makes frequent misidentifications and mistakes.

Value trade-offs

To deliberate ethical AI and when risk exposure is allowed, it is necessary to consider values such as freedom, autonomy, justice, responsibility, fairness, democracy, and others. As now today, technologies over time impact what we value and lead to techno-moral value change. Privacy remains a central value for most of us. Yet, we seem increasingly inclined to give up privacy for comfort, or precise location services, uploading our lives online and even trusting AI-based mental health apps. Similarly, not all AI impacts people equally and achieving a fair distribution of risk is always a challenge. If policymakers need to assess the ethical risks of AI systems, users' and societal value trade-offs should be discussion points.

As a society, we are at a crossroads. The European approach to AI and the low-high risk pyramid promise a safer, more responsible technology landscape. Fundamental human rights, trust and the shared future, are in the spotlight more than a decade ago. However, the proposed policy frameworks and debates continue categorising risk most often as a purely rational, objective measure of potential harm. Individuals' emotions, opinions, collective values, and social acceptability are too often left out of policy discussions. Risk ethics can help guide the talks beyond the cost-benefit viewpoint. For starters, it allows us to shift the narrative from risk as severity and probability to risk as risk-taking and risk-imposing. Then, we could begin to have a more holistic approach to risk.

Disclaimer

This is a joint publication of the Wilfried Martens Centre for European Studies and the Hanns Seidel Foundation. This publication receives funding from the European Parliament. The Wilfried Martens Centre for European Studies, the Hanns Seidel Foundation and the European Parliament assume no responsibility for facts or opinions expressed in this publication or any subsequent use of the information contained therein. Sole responsibility lies on the author of the publication. The processing of the publication was concluded in November 2022.