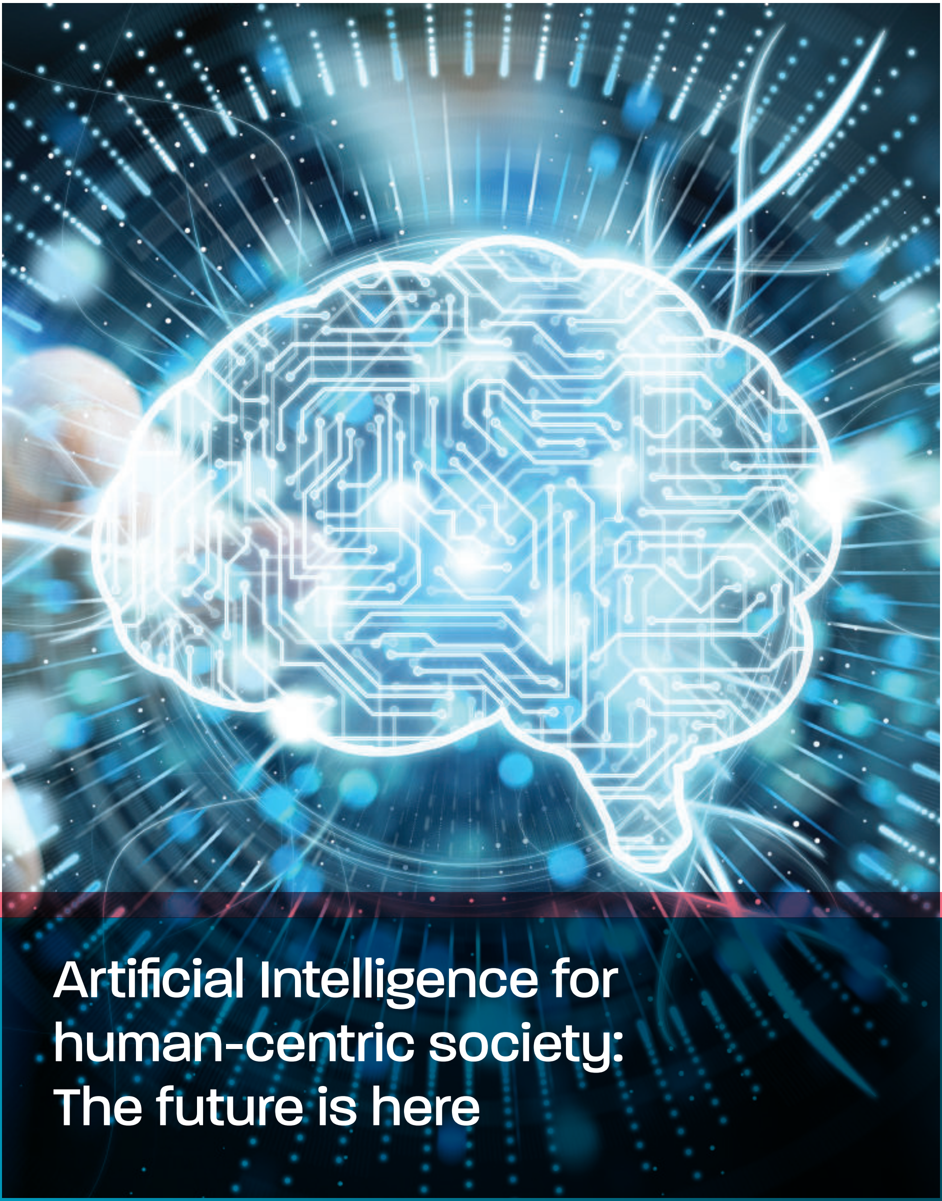
Towards emerging technologies and e-government

Hodžić, Sabina



Artificial Intelligence for human-centric society: The future is here

Edited by
Dr. Nina Tomaževič
Dr. Dejan Ravšelj
Dr. Aleksander Aristovnik

ZAVOD 14
zavod za sožitje in napredek
elf

ARTIFICIAL INTELLIGENCE FOR HUMAN-CENTRIC SOCIETY: THE FUTURE IS HERE

Edited by Nina Tomaževič, Dejan Ravšelj and Aleksander Aristovnik

European Liberal Forum EUPF,
Rue d'Idalie 11-13, boite 6, 1050 Brussels, Belgium
info@liberalforum.eu
www.liberalforum.eu

Zavod 14,
Ljubljanska cesta 16, 3000 Celje, Slovenija
info@zavod14.si
www.zavod14.si

Graphic Design by Spotnet d.o.o.,
depositphotos.com

ISBN: 978-2-39067-063-6 9782390670636

Published by the European Liberal Forum EUPF with the support of the Friedrich Naumann Foundation and Zavod 14. Co-funded by the European Parliament and ARIS (P5-0093). Neither the European Parliament nor the European Liberal Forum EUPF 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 necessarily reflect those of the European Parliament and / or the European Liberal Forum EUPF

ZAVOD 14
zavod za sožitje in napredek

elf

EUROPEAN LIBERAL FORUM (ELF)

The European Liberal Forum (ELF) is the official political foundation of the European Liberal Party, the ALDE Party. Together with 59 member organisations, across Europe we work to bring new ideas into the political debate, provide a platform for discussion, and empower citizens to make their voices heard. The ELF was founded in 2007 to strengthen the liberal and democrat movement in Europe. Our work is guided by liberal ideals and a belief in the principle of freedom. We stand for a future-oriented Europe that offers opportunities for every citizen. The ELF is engaged on all political levels, from local to European. We bring together a diverse network of national foundations, think tanks and other experts. At the same time, we are close to, yet independent from, the ALDE Party and other Liberal actors in Europe. In this role, our forum serves as a space for the open and informed exchange of views among a wide range of actors.

ZAVOD 14, ZAVOD ZA SOŽITJE IN NAPREDEK

Zavod 14, zavod za sožitje in napredek is a non-profit (ELF full member) organisation that has its headquarters in Celje (Slovenia). Zavod 14 promotes social liberal ideas (balancing between individual liberty and social justice) and protects liberal values (e.g., democracy, the rule of law, social development, good governance etc.) The mission of Zavod 14 is to support civil society, integrate and cooperate with the interested public, transmit the perspectives of interested stakeholders to state and other institutions, cooperation in the preparation and implementation of politics, and contribution in joining the civil society initiative into international integrations.

CONTENTS

INTRODUCTORY REMARKS

CONTENTS

INTRODUCTORY REMARKS

Society is faced with many societal, economic and technological issues that will be addressed with this project since these are long-term structural challenges such as global warming, natural resource depletion, rising inequalities, demographic trends, growing economic disparity, etc. The Covid-19 pandemic emerged as an additional challenge, revealing the digital divide and emphasising the importance of leveraging digitalisation and Artificial Intelligence (AI) even more. Accordingly, the main goal of the publication is to examine the current AI landscape and initiatives in the European Union (EU) and explore their role within the context of the human-centric society. It is crucial to understand the governance and regulation of AI, challenges and opportunities for the EU to leverage digital evolution and AI and identify good practices and scalable solutions to support future-oriented Europe and maintain a stable and healthy economy. Therefore, the publication has several aims. First, it aims to identify the challenges and opportunities related to the digital transformation and transition to the future of the EU through AI and its implications for the liberal society. Moreover, it aims to explore the state of the governance of AI and its regulation on the supranational and national levels around the EU, provide examples of AI applications, related opportunities and threats, and prospective improvements to the current digital and national policies intended to promote AI projects in the EU. Finally, it aims to provide the specialised recommendations that public authorities and certain types of institutions may use. These guidelines will serve as instruments for effective and efficient support of AI projects in the EU. Therefore, the publication is expected to give a precious tool for evidence-based policymaking and will be promoted through our liberal partners in the EU and beyond.

This first chapter discusses the emergence and challenges of the Metaverse, including defining it, concerns about its legal regulation, and the potential impact of AI, aiming to identify opportunities and threats in social, ethical, and legal terms while advocating for the development of legal frameworks to govern it within the EU and its member states. The second chapter presents the challenges of regulating AI, focusing on the difficulties in defining AI, the limitations of existing legal frameworks, and the disruptive nature of AI applications, ultimately arguing that conventional regulation may be inadequate for this rapidly evolving field. The third chapter explores the rapid advancements in information technology and AI, discussing their impact on society, the development of new AI models and technologies, and the regulatory approaches taken by different regions, particularly focusing on the EU stance on AI and data protection. The fourth chapter concerns the ethical challenges associated with AI in the context of the EU and aims to extract lessons that can be applied to AI governance and regulation in Africa as the continent embarks on its technological transformation during the Fourth Industrial Revolution. The fifth chapter is about two contrasting approaches to utilizing publicly held education data, focusing on the pros and cons of sharing the data with researchers and the private sector versus developing an education-specific AI foundation model directed by public authorities, emphasizing the benefits of the latter approach. The sixth chapter presents the role of emerging technologies, including AI and blockchain, in transforming the administration and e-government services in Croatia, emphasizing their potential benefits, challenges, and policy recommendations.

The seventh chapter presents the potential of Bulgaria as an AI hub and its reliance on EU policies in the context of AI technology adoption, highlighting the need for policies that ensure policy space of the EU member countries and proposing specific recommendations for harmonization, semiconductor research, workforce training, and state-sponsored investments to reduce spatial inequalities and support AI development. The eighth chapter discusses the challenges and opportunities related to digital transformation through AI in Poland, covering AI trends, the current state of AI policy and AI companies in Poland, the regulatory framework for AI at the Polish and EU levels, and formulates guidelines for governance and regulation in the field of AI. The ninth chapter is about the current status and trends of AI in Poland, comparing them with activities and strategies observed in the EU, and addresses key challenges and opportunities in AI policy in Poland, offering recommendations for various stakeholders. The tenth chapter highlights the role of AI in the digital transformation in the EU, exploring AI governance, regulation, and policies within the context of a human-centric society, including the emphasis on ethical AI, regulatory clarity, and its aim to be a global leader in trustworthy AI. The eleventh chapter examines the role of AI in decision-making, digital transformation, and the sustainable development of the EU, highlighting both its benefits and associated concerns such as manipulation, cyberattacks, privacy, and data protection.

The Editors

Chapter 1

AI and (in) the Metaverse: Interactions and legal implications

Gian Marco Bovenzi

1 Introduction

The word Metaverse is often mentioned in several domains of society. We talk about learning in the metaverse, playing in the metaverse, working in the metaverse, doing in the metaverse everything that we can do in real life, but with the use of augmented or virtual reality and the Internet. As highlighted below (par. 2), the metaverse is in fact being used for e-commerce and advertising purposes, as several firms and companies are currently selling and sponsoring their products in the metaverse; in the education sector, through virtual education and immersive experiences/ simulations; in professional/working contexts, developing new forms of (virtual) meetings; finally, for leisure purposes, such as socialising and networking, attending concerts, visiting virtual reproductions of cities, going to art expositions, and so on.

Nevertheless, there is probably as much talk about the metaverse as there is confusion concerning it;

namely, what it is, how it works, how it can work in the future, and according to those more sceptical even if whether will hold any sort of implications for tomorrow's society. What is sure is that the metaverse exists and, as such, raises concerns about its regulation from a legal standpoint – privacy, finance, torts, copyright and IP, and even crimes potentially being committed in the virtual world. Still, thus far legal regulation of the metaverse is lacking by either the EU (aside from the recent "good intentions to") and in the national legal frameworks of the member states.

Moreover, future potential applications of the metaverse await to be discovered. Today, hardware and tools supporting virtual immersion into the metaverse are still not fully developed and accessible to the general public, whereas Web 3.0 and 4.0 along with the yet not completely implemented infrastructure and decentralisation do not allow a total 'metaverse experience'. The role of artificial intelligence in and for the metaverse could nonetheless play a fundamental role in several domains, e.g., NLP, virtual assistants and bots and, generally speaking, to create environments, characters and objects in order provide tools for users to enable a more immersive and interactive world. In brief, AI is highly likely to translate today's metaverse into tomorrow's metaverse, exponentially increasing the number of its users (and uses) over time.

While the metaverse brings (and will bring) increasing opportunities for society and individuals, there are also growing threats which, without preventive regulation, might hold serious consequences for society. Therefore, after describing what the metaverse is, how it currently works, and what its uses in the future might be, the present paper aims to identify which opportunities and threats may be posed by use of the metaverse – as also implemented by AI –in social, ethical and legal terms. Finally, the paper provides a conclusion where it is suggested that the EU and member states should urgently address this issue by coming up with legal frameworks to regulate the metaverse and its possible uses.

2 What is the metaverse (and its expected uses)

There is no single, agreed-upon definition of the word metaverse. Although its etymology suggests the concept simply represents a space beyond (from the Greek 'meta') the universe (uni- 'verse'), a precise taxonomy of the metaverse's features is presently missing. In June 2022, the European Parliament Policy Brief issued "Metaverse. Opportunities, risks, and social implications", which defines the metaverse as a tri-dimensional, virtual and immersive world with the characteristics of realism, ubiquity, interoperability and scalability. In his essay "The Metaverse: what it is, where to find it, and who will build it", Matthew Ball lists additional characteristics of features of the economy and identity, as well as physical and digital dimensions.

From a practical standpoint, the metaverse is a digital platform that users may join by creating their own account, their online alter ego (called an avatar) and linking their wallet/portfolio in order to perform several everyday activities like gaming/playing and complete financial transactions and e-commerce for purposes of learning/teaching, but also leisure (like going to concerts, visiting virtual museums etc.). It is claimed that what distinguishes the metaverse from other social networks or platforms is its immersivity, that is, the sensation of having a real-life experience on an online platform.

In fact, metaverse platforms mostly include augmented reality (AR) in their architecture, where AR is a technology allowing users to observe the real world, but with virtual objects overlapping real objects. A few examples of how AR gives a perception of immersivity: the game Pokémon Go works superimposing a virtual Pokémon character in real-world locations: the gamer holds an AR technology (mobile phone, App) through which he sees the world, visualising the Pokémon in real-life locations, and then being able to caputre it. Also, IKEA developed an App that allows customers to superimpose scale models of their furniture in real rooms, enabling them to make the best choice for their own houses. AR is used not only in leisure, but also, for instance, in the U.S. Army, that developed the 'Tactical Augmented Reality' – eyeglasses and devices to be wore in order to increase awarenss on soldiers' localisation in a given place.

The concept differs from virtual reality (VR) that is also used by certain metaverse platforms: the latter is a technology capable of re-creating, through the interaction of software–hardware devices (headsets, goggles, tactile sensors),

> "There is no single, agreed-upon definition of the word metaverse.

a tri-dimensional, immersive and interactive environment capable of simulating a physical presence in a virtual landscape. As examples, in healthcare future doctors/surgeons wearing ad-hoc devices might practice on virtual bodies in a virtual surgery room; in tourism, guided tours of virtual cities might be made (a lot of cities worldwide have already their 'virtual twin'); socially, virtual cinemas, restaurants, or concerts can be visited; again in the military, virtual battlefields may be implemented in order to enhance soldiers' capabilities in potential real-life situations or conditions.

Understanding the difference between AR and VR is crucial given that is it important to stress that today most metaverse platforms use AR technologies, which means they have a lower level of immersivity than what is potentially given by VR platforms. The degree of immersivity a user experiences in the metaverse is fundamental: the greater the immersivity, the stronger the sensation of a real-life experience and, therefore, the bigger the potential legal issues associated with it. Further, artificial intelligence (AI) components are also used in the metaverse: ranging from software and hardware components in order to better exploit its potentiality, to the use of generative AI to carry out activities in the metaverse like content creation of literature, images, NFT etc. What is sure, is that along with technological developments in the future, there will be more and more examples of AR/VR technologies enabling users to feel even deeper and more immersive real-life environments – and thus, experiences.

Another vital aspect of the metaverse's full potential revolves around the concept of decentralisation. A decentralised architecture embraces the absence of a central node, namely, a central entity or hosting/Internet service provider controlling the network, and the subsequent data processing and

> "The degree of immersivity a user experiences in the metaverse is fundamental: the greater the immersivity, the stronger the sensation of a real-life experience.

storing by multiple nodes such as peer-to-peer networks and distributed ledgers like the blockchain system. Accordingly, a decentralised web implies the use of decentralised digital networks and technologies with the Internet, described as, "a system of interconnected, independent, privately owned computers that work together to provide private, secure, censorship-resistant access to information and services" (Aboukhadijeh, 2022) as well as a "series of technologies that replace or augment current communication protocols, networks, and services and distribute them in a way that is robust against single-actor control or censorship" (Griffey, 2022). Decentralisation is the key aspect of Web 3.0 and Web 4.0 (although not all Web 3.0 and Web 4.0 work in a fully decentralised way) since the full potential exploitation of these new forms of Web through a decentralised system enables considerable security, transparency, privacy, independence, accessibility and, thus, democracy of the Internet. In fact, the lack of central ownership of central entities implies full control of the activities made on the Internet by users. In applications of Web 3.0 such as blockchain, cryptography, distributed storage, privacy computing, and smart contracts, a given operation is entrusted by the power of the consensus mechanism that is a property of decentralisation (Chen et al., 2022). Instead, Web 4.0 is intended to fully blend the concepts of 'physical' and 'digital/virtual', including environments and landscapes, via the use of advances made in AI, the IoT, and extended reality (XR, including AR and VR) technologies (COM/2023 442/Final, official EU document). Ideally, decentralised Web 3.0 and Web 4.0 should then create an open, trust-less, permission-less Web where "users can accomplish content publishing, economic transactions, and other actions without going through a centralized platform. They employ DAO to manage their

The metaverse is likely to represent the future of the Internet also when considering the numerous activities carried out in such an environment.

digital identities, assets, and data by themselves, through the extended reality (XR) technology hardware and blockchain distributed storage technology together form the technical foundation of Web3.0" (Chen et al., 2022).

Although decentralised metaverse platforms are today fewer in number than centralised metaverses, a decentralised infrastructure of a platform would surely permit the enhanced use of all features of the metaverse and, accordingly, this is the goal metaverse developers have set for themselves in the (near) future. For instance, using a decentralised structure would secure a given transaction via smart contracts and blockchain, therefore becoming more trusted than a 'regular' transaction.

In summary: by being a digital, tri-dimensional, (relatively) immersive and potentially decentralised social platform developed through the combined and not necessarily contextual use of the IoT, AR, VR and AI, the metaverse is likely to represent the future of the Internet also when considering the numerous activities carried out in such an environment.

For instance, the metaverse can represent a new domain for e-commerce and the trade of goods and services both between users and between a user and a firm/company/online shop. There are already famous brands advertising their products in the metaverse (Prada, Gucci, Balenciaga, Adidas, to name just a few): here, a user can either buy a product from a brand using their digital wallet associated with their profile or using tokens, or even purchase virtual clothes or goods to dress up and fully personalise their avatar. In other cases, new forms of advertising are emerging: e.g., the automotive brand Skoda offers virtual test-drives in the metaverse so that a user might buy the subject car in real life. E-commerce and financial transactions in the metaverse can also be made with use of cryptocurrencies via the blockchain and/or smart contracts.

The metaverse may represent a future (yet also present) environment for education and schooling: users can attend online courses offered by institutions that provide virtual education, and several universities are already experimenting with this method. Learning in the metaverse provides students with a more immersive experience, not only learning from the books, but also in virtual classes – where interacting with people from around the globe is enhanced – or via the creation of virtual environments: for instance, RAI (an Italian TV network) offers the chance of learning Dante Alighieri's Divine Comedy in the metaverse, alongside Dante in the Inferno, Purgatory, and Paradise. Seemingly, the immersivity and virtual reproduction of real life might enable several professions to be practised: here one may think of medical treatments and practice that surgeons and doctors can experiment with in the metaverse during Med School and specialisation – and before practising on an actual person.

Where we witnessed the authentic explosion of new platforms like Zoom and Microsoft Team for carrying out working activities during the COVID-19 pandemic, the metaverse might represent a new working landscape as well. One might think about the online meetings via Zoom – something that still today constitutes an ordinary way of working – albeit in a virtual world: we would not sit in front of a screen watching our colleagues' faces, but our full-body avatars would be gathered around the same table, enhancing the immersivity and making the work meeting appear like a real-life in-person meeting. Several firms and companies are already present in the metaverse.

Aside from educational/professional/commercial activities and purposes, leisure, socialising and networking are certainly one of the main features of the metaverse. Just like in social networks, in the metaverse users might meet and greet not behind a screen but virtually, making it seem like an in-person meeting. There are already examples of singers performing concerts in the metaverse (e.g., Justin Bieber, Ariana Grande, Ozzy Osbourne); cities can be visited in the metaverse (Seoul is the first fully virtual metaverse city, although Singapore, Tokyo and New York are experimenting as digital twins, and London is a 3D metaverse city); art exhibitions are held in the metaverse as well, with various artists presenting their works and also creating fully-virtual pieces of art such as non-fungible tokens (NFTs): this is all made possible by the use of content creation enhanced by AI (as described in the next paragraph). If a person does not want to engage in any of these activities but to simply 'hang around', the metaverse also exists for socialising and chatting with other avatars/people.

The conspicuous number of activities potentially carried out in the metaverse accordingly raises several associated legal issues, added to by the potential growing use of artificial intelligence.

3 Use of AI in the metaverse: Opportunities, legal and social challenges

There are (or might be) several potential AI components of metaverse platforms, which would probably help to transform today's metaverse into the metaverse of tomorrow, meaning a much more immersive environment where real-life perception is even more tangible.

First, let us consider the way artificial intelligence might 'create' or shape the creation of avatars, avatar twins, or digital humans. AI might capture certain personal features of metaverse users, reflecting such features virtually and creating a human-like avatar with as many features as possible corresponding to a real person (e.g., hairstyle, face-traits, and so on).

Moreover, AI might play a role in natural language processing: chatbots, language translation, document analysis, predictive texts, or perhaps sentiment analysis. Enabling and enhancing communication in a virtual world might bring enormous benefits for accessibility in communication, also allowing certain categories of people to raise their voice when in real life they might encounter greater difficulties doing so (such as physical disabilities, or minorities not having freedom of expression).

Generative AI is another example of AI in the metaverse, being a type of artificial intelligence creating data as images (in the forms of pictures and videos), audio and music, text, 3D models, yet also pieces of art that might be considered (and sold) as NFTs. Generative AI might solve problems, answer questions, and enable a greater real-life experience in the metaverse given that users can create their own contents – protected by copyright as well, as shown below – in several domains.

Of course, AI might not only help as support in the metaverse, but as hardware for enjoying the full immersive experience. Headsets, goggles and other typologies of sensors stimulating and recreating the five human senses (already available on the market are sensors permitting olfactive and tactile – muscular memories, for instance – experiences, and it is no doubt a matter of time before the sense of taste will form part of these technologies).

The picture drawn thus far suggests that a lot of opportunities will emerge from use of the metaverse and the virtual worlds, particularly when it comes to the AI components building its infrastructure – both internal components (generative AI, NLP, audio-visuals) and external components (hardware such as googles and headsets). Further, whether enjoying the metaverse environment will become a 'daily thing' – for

The picture drawn thus far suggests that a lot of opportunities will emerge from use of the metaverse and the virtual worlds.

comparison, like with the 'boom' experienced by social networks around the year 2010 –it is highly likely that further opportunities will arise parallel to the development of technologies allowing a more immersive experiences. This scenario depends on the market flow the metaverse will be able to generate: the greater the number of users subscribed to metaverses; the greater the economic flows generated. Accordingly, the greater the investments made by private tech companies in the field, the greater the technological development expected. Of course, there is no 'Day X' for this to happen, one can only speculate how long it will take before (and if) that will happen. Still, it is certain that when it happens, its impact can be anticipated to be enormous. Not only opportunities, as described, but current and future social and legal

CONVERSION
The ultimate goal of your campaign. A conversion is whatever you decide it is – submitting a form,
LOADING...
MARKETING
MAIL
RESEARCH DATA
Research data are any physical and/or digital materials that are collected, observed, or created in research

challenges, perhaps ones that even limiting the opportunities, arise.

First, let us consider how artificial intelligence might create avatars or digital humans. While this is doubtlessly an opportunity for inclusivity and equality, on the other hand, it might bring with it the risk of generating deceiving and deceptive characters interacting with humans – the latter not knowing that they are actually speaking with an artificial replacement of a human being. The central example of this challenge is the use of deepfakes. A high risk arises from malicious use of deepfakes, that is, a replacement of a real human being using pictures, videos and audios created via AI software which, possessed with 'real' content, is able to modify or re-create in an extremely realist manner the features and/or movements of a face, body or voice. Specifically, deepfakes might represent the theft of identity: by stealing and using someone's image for different purposes a deepfake might 'think', 'speak', 'act' or 'be present somewhere' in a way that the real person has not genuinely done. Moreover, the most serious hypothesis of deepfake, the 'deep nude', depicts persons engaging in sexual behaviours or images, thereby raising serious pornographic issues – including pedo-pornography and the exploitation of child images. In politics as well, politicians might be depicted in deepfakes delivering speeches they never delivered, in turn influencing public opinion and increasing the risk of fake news.

OPENING FOLDER...
ANALYSING...
Data analysis is a primary component of data mining and Business Intelligence (BI) and is key to gaining the insight that drives business decisions. Organizations and enterprises analyze data from a multitude of sources using Big Data management solutions and customer experience management solutions that utilize data analysis to transform data into actionable insights.
9:47 AM

Second, the use of AI-NLP (natural language processing) in the metaverse and the virtual worlds could produce problems related to the dissemination of illicit and harmful content in terms of hate and discriminatory speech, libel or defamation without knowing who actually committed the offence since NLP is able to mask the real person. This must be regarded a serious concern for liberal society since hate speech-related offences can lead to discrimination for minorities – and not only them.

Third, generative AI (in the broader form of NLP) might not only allow deepfake and hate speech illegal activities, but also create intellectual property and copyright issues. The first question is who holds the copyright for the artificially-generated content: is it the person generating the algorithm, or user who first uses the content for a given purpose? Regulation is currently lacking in this respect. The second issue relating to IP and copyright is about user-generated content ('UGC') that entails several questions: when it comes to registering trademarks and patent for something generated in the metaverse (e.g., an NFT), how can that be protected and classified under the current trademark existing laws and classifications (including the 1957 Nice international classification)? How is it possible to keep track, identify and punish a user who has infringed upon someone else's copyright when such user has exploited AI technologies? What law on trade secrets and know-how applies as concerns the creation of AR and VR models, software and technologies? In addition, when an NFT is sold and bought the buyer surely becomes the owner of the source-code and metadata, but lacking a specific agreement, do they become owner of the copyright for that NFT as well? Finally, can data be considered as a product protected by IP and copyright law?

The fourth issue concerns commercial activities and the market, representing a crucial matter for liberal society, liberal economies and the free market, which is a cornerstone of the European Union. Let us start by saying that marketing activities might acquire a totally different aspect compared to traditional ones. In a virtual world, new types of marketing and advertisement can be invented (e.g., test drives of virtual cars), taking advantage of the immersivity and real-life sensation that AR or VR may provide a user with. And this is perfectly fair. But, on the other hand, generative AI might serve unfair purposes like creating false models of the products being advertised (that would not match the real-world products), publishing false reviews allegedly written by (non-actually existent) consumers, misleading and hidden advertising in a user's search results, or even the buy-and-sale of non-existent objects exploiting the use of bots – automatised computer programmes. In summary, the metaverse could pave the way for the delivering of offerings and provision of unique products that are impossible in the real world based on the highly immersive AR/VR environment. In turn, this might lead to manipulative advertisement techniques and, moreover, practices like the aggressive capturing of users' behaviours and personal data, bringing about unfair and misleading commercial practices (inducing the average consumer to make a transactional decision they would not have taken otherwise) eventually ending up in market imbalance.

Privacy and data protection is another sector potentially threatened by the use of unregulated AI in the metaverse, for several reasons. First, the type of data and information shared in the metaverse is not limited to 'traditional' data and information (personal and demographical, consumer preferences, opinions, market traceability), but also new and more invasive typologies of information like biometrical data, facial expressions, body movements, and emotional status or reactions. Given the hardware used to enjoy the full immersive metaverse experience – goggles, headsets, olfactive and tactile supports, this is potentially leading to such data and information exploitation and breaches for commercial, marketing yet also illegal and criminal purposes. The abovementioned data and information are not specifically protected by the current regulations and legal frameworks (as highlighted in the following paragraph): perhaps these categories of data might be included in sensitive or confidential information under the GDPR, although they do not fall under specific legal provisions. Further issues arise from how personal data are gathered, considering the possible lack of a uniform and detailed privacy policy for users, especially regarding AI tools (such as, for instance, ChatGPT), as well as the lack of uniform legal grounds regulating the collection and storage of personal data. Finally, when it comes to the concept of personal identity it is worth assessing whether avatars or digital twins should be considered as personal alter egos or simply digital representations of a user. For example, while looking for a particular person in search engines, search results show the person's subscription to social networks (Facebook, Instagram, LinkedIn): will search engines show metaverse subscriptions, if any, as well? If yes, how would this data be protected?

AI model training, and subsequent potential accountability for AI bias, is another sensitive matter. Model training is a specific phase when machine learning algorithms are 'taught' to function by minimising biases or the loss

of functions over time, and this process should be delicately shaped around the peculiar use of AI in the virtual worlds where the misuse of data and information, as well as potential biases, lurk around the corner. Accordingly, accountability for un-trained AI models should be regulated: while it is true that the EU's AI strategy includes solutions for AI accountability (aspects of which will be further deepened), it is likewise true that specific liability for AI in the metaverse is not provided at the moment, and this aspect should certainly not be underrated.

Finally, threats are possible – and highly likely – with respect to crimes and illegal conduct in the metaverse. The use of AI could facilitate the perpetration of an alarmingly higher number of criminal activities. We previously described hate speech and hate crime-related conduct where the use of UGC, NLP, deepfakes and generative AI might give birth to crimes without an actual perpetrator – since it is the AI actually committing the crime. This means that if the person behind the programme is not discovered, the crime would basically end up without someone being held liable for it. The same reasoning applies to the dissemination of pornographic and obscene contents, including child pornography and exploitation. Further, a new hypothesis of sexual harassment must not be excluded: given the immersive nature of the metaverse environment, notwithstanding that it is complex to think about a purely 'physical' assault or violence perpetrated with AI tools or software, the typology of harm experienced is more than mere 'psychological' harm because the immersivity and real-life experience provided by metaverse platforms make the violence more realistic. This hybrid form of sexual harm is currently unregulated and should become a priority for policymakers. Other issues concern the risk of the dissemination of terrorist or violent extremist-related content in the metaverse. As noted by the EU's Counter Terrorism Centre, terrorists or terrorist groups might acquire new typologies of actions to support their narratives: virtual propaganda, recruitment and communication, terrorist financing via cryptocurrencies or blockchain-based transactions, as well as new forms of 'military' trainings for terrorists. This whole 'underwater world' must be seen as a priority for policymakers since it represents an insidious threat to liberal society in a free European Union.

All of the threats and challenges identified thus far represent the biggest issues connected with the metaverse, especially within a metaverse built using AI components capable of making the experience extremely more immersive, realistic and, therefore, emotional. On top of them, one last aspect is to be highlighted – encompassing all the components highlighted – that is, the aspect of the Internet's decentralisation in the terms described above in section 2.

While on one side decentralisation appears as a form of Internet where central entities and tech giants are de-powered, hence resolving several critical issues surrounding data collection, storage and privacy, on the other side it raises one important issue to be dealt with: content moderation and removal in the case of the dissemination of illegal and harmful content online. This aspect could pose a serious challenge in decentralised metaverses (or platforms) that make decentralised architecture and AI components their grounding structure. In fact, while in a centralised network illegal and harmful content might be easily removed by a service provider or a central entity in general – that in fact are obliged to do so in the EU legal framework, as will be stressed below – decentralised architectures do not envisage this possibility. The peculiar peer-to-peer entrusted structure of decentralised operations in fact makes it impossible for central entities to take any action on distributed ledgers. Therefore, if illegal content is disseminated on a decentralised platform (e.g., hate speech, pornographic images, terrorist contents) there is a considerable risk that such content will continue to remain online. Ex-post solutions are no longer viable with decentralised platforms, and thus the only way to prevent illegal and harmful content from being published online lies in the adoption of preventive and ex-ante measures.

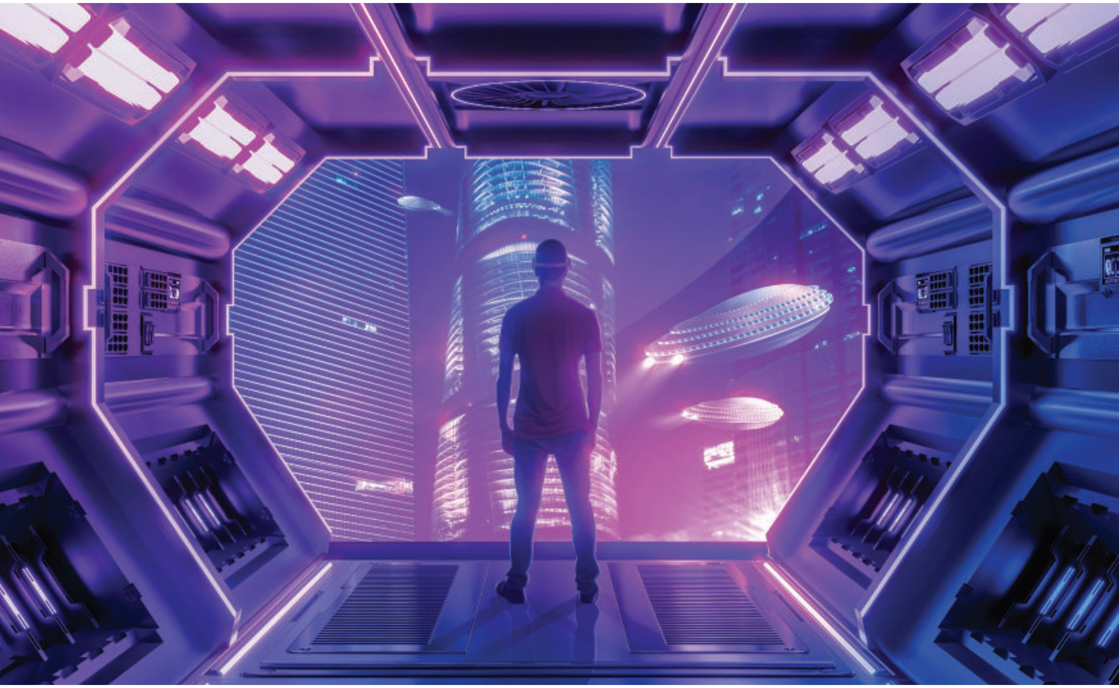
The latter issue is not yet regulated and policymakers should make efforts to address it. However, starting from the very beginning, let us explore the existing EU policy and legal framework that applies to AI in the metaverse.

As shown, the use of artificial intelligence in virtual worlds raises several problems. If the issues mentioned above remain uncovered, the implications for liberal society might be serious from several standpoints.

4 Current legal and policy framework in the EU and in Italy

As shown, the use of artificial intelligence in virtual worlds raises several problems. If the issues mentioned above remain uncovered, the implications for liberal society might be serious from several standpoints: economic (by threatening the free market and paving the way for abuses like dominant positions or unfair trade practices), social (privacy and personal data issues) and legal (liability, torts, criminal law). This means it should be a priority for liberal society to raise its voice to address these matters, potentially capable of bringing serious future problems – just like the un-regulated use of social networks in the past engendered serious political problems (hate speech and fake news, personal data used for orienting people's electoral choices, conspiracy theories etc.).

The current EU framework on artificial intelligence and the metaverse can be assessed in a three-level analysis: 1) acts that have been adopted and are already in force; 2) acts that call for future action; and 3) acts on their way to being adopted in the near future. In all of these, while it is important to recall that none specifically addresses the issues raised by the use of AI in the



metaverse, nevertheless all of the mentioned acts would be covered.

Within the first group, on one hand there are non-binding and directly enforceable documents such as: the Digital Decade Policy Programme 2030 (Decision (EU) 2022/2481) of the European Parliament and of the Council of 14 December 2022 setting out the framework for member states to cooperate on realising a democratic, accessible and sustainable digital transition; the European Declaration on Digital Rights and Principles recalling the need of uniform member state policies to accomplish the digital transition. On the other hand, there are legally binding documents such as: in the fields of privacy, data protection and personal identity, Regulation (EU) 2016/679 (the "GDPR"), Regulation (EU) 2022/868 (the 'Data Governance Act' – DGA), the upcoming E-Privacy Regulation, repealing Directive 2002/58/EC, and Regulation (EU) 2022/2065 (the 'Digital Services Act' – DSA) of the European Parliament and of the Council of 9 October 2022; in the fields of intellectual property law, trade secrets and copyright, Directive 2004/48/EC and Directive 2009/24/EC; Directive (EU) 2019/790 (the 'Copyright Directive'); Regulation (EU) 2017/1001 on trademarks; Directive (EU) 2016/943 on trade secrets and know-how; in the field of commercial practices and marketing, Directive 2005/29/EC concerning unfair business-to-consumer commercial practices in the internal market and Regulation (EC) no. 2006/2004, Directive (EU) 2019/2161 and, finally, Regulation (EU) 2022/1925 (the 'Digital Markets Act' – DMA).

The second group of acts, that call for future action, includes the communication "An EU initiative on Web 4.0 and virtual worlds: a head start in the next technological transition". This communication is 'fresh' (11 July 2023) and represents the first EU act specifically targeting the metaverse. In highlighting its potential and challenges, the document sets out the vision, strategy and proposed actions aiming to make a significant contribution to achievement of the Digital Decade objectives of technological leadership, sovereignty and competitiveness by 2030. The communication may be seen as a pillar for future policies in this sector, and perhaps must be viewed as pivotal.

The third group of acts includes the upcoming proposal of the European Parliament and of the Council of 23 February 2022 (the 'Data Act') and, primarily, the European AI Strategy. The European AI Strategy seeks to make the European Union a world-class hub for AI and ensure that AI is human-centric and trustworthy, while it includes the Communication on fostering a European approach to AI, a review of the Coordinated Plan on Artificial Intelligence (with EU member states), and a proposal for a regulation prescribing harmonised rules on AI ('AI Act') and a relevant impact assessment. Within its scope of building trustworthy AI that will establish a safe and innovation-friendly environment for users, developers and deployers, the Commission proposed three inter-related legal initiatives: 1) a European legal framework for AI to address fundamental rights and safety risks specific to AI systems; 2) a civil liability framework – adapting liability rules to the digital age and AI; and 3) a revision of sectoral safety legislation (e.g., Machinery Regulation, General Product Safety Directive).

While the civil liability framework and the revision of sectoral safety legislation have the fundamental goals of regulating certain aspects of the law ensuring citizens' security and safety generally, the European legal framework will be the most important. The proposed regulation specifically aims to tackle risks potentially created by the un-regulated use of artificial intelligence, including AI in the metaverse. As the Commission clearly states on its dedicated webpage (https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai), "although existing legislation provides some protection, it is insufficient to address the specific challenges AI systems may bring". The Commission stresses that the proposed rules will:

1. address risks specifically created by AI applications;
2. propose a list of high-risk applications;
3. set clear requirements for AI systems for high-risk applications;
4. define specific obligations for AI users and providers of high-risk applications;
5. propose a conformity assessment before an AI system is put into service or placed on the market;
6. propose enforcement after such an AI system is placed on the market; and
7. propose a governance structure on the European and national levels.

This future set of rules is likely to enhance the EU's capacities in the field of AI regulation and the mitigation of risks deriving from its different applications. Moreover, aside from substantial and technical assessments (list of high-risk applications, requirements for AI systems, conformity assessment), further obligations will be provided for both AI users and providers. Involving the latter in clear regulation embeds the policymakers' fundamental intention to keep up with the policies implemented thus far in assuring big tech companies' accountability, aligning with other in-force regulations in the field: let us consider the Digital Services Act that imposes several obligations on various online intermediary services (network infrastructures) depending on their role, size and impact, in terms of moderating and removing harmful and illegal content online;

or the Digital Markets Act that targets the 'gatekeepers' ("entities that manage strategic platforms and services directly linking consumers and enterprises") which must comply with several restrictions and obligations in order to guarantee citizens and consumers free access to the digital market, tackle market abuses and abuses of dominant position, as well as stimulate innovation and competition.

The accountability of Internet service providers is crucial in a centralised web as it represents the main legal instrument available to tackle online abuses of any kind made by users. Since preventive control is nearly impossible given the amount of content shared on the web, the only effective solution is encompassed by ex-post content moderation and removal. In addition, further restrictions provided by the Digital Market Act are able to prevent providers' abuse of dominant position or market distortions. On top of this, the future AI Strategy will likely set extra rules and obligations to ensure the safe and secure use of AI, and accordingly the framework on the EU level appears satisfactory – or at least it is staying pace with technological evolution.

On the national level (whose governance structure is among the goals of the AI Strategy), the country experience assessed in this chapter refers to Italy. First, it is necessary to highlight the institutional body in charge of proposing and enforcing policies the Department of Digital Transformation – under the Italian Presidency of the Council of Ministers – with the goal of promoting equality, ethics and justice, as part of a strategy of innovation and development concentrated on people and the planet. The Department's goal for a "Digital Italy 2026" is to be realised using funds of the Italian National Recovery and Resilience Plan (NRRP), under the umbrella of the Next Generation EU Fund. The funds allocated to digital policies amount to up to 27% of the whole NRRP, to be invested in the two core objectives of the plan: EUR 6.74 billion is allocated for digitalisation of the public administration and EUR 6.71 billion for the implementation of ultra-fast internet network throughout the country (https://innovazione.gov.it/italia-digitale-2026/). The five ways to make Italy one of the top EU countries in digitalisation are:

1. providing 70% of Italian citizens with a digital identity;
2. enhancing the digital competencies of at least 70% of Italian citizens;
3. adopting cloud systems for 75% of the public administration;
4. digitalising 80% of essential public services; and
5. ensuring ultrabroadband for 100% of Italian households.

On top of this, the Italian Strategic Programme on Artificial Intelligence 2022–2024 *(https://assets.innovazione.gov.it/1637777513-strategic-program-aiweb.pdf)* is focused on three areas of intervention:

1. strengthening and attracting the talents and competencies that will enable the AI-driven economy;
2. expanding funding for advanced research in AI; and
3. favouring the adoption of AI and its applications in both the public administration (PA) and Italian economy generally.

These areas are intended to be implemented through six objectives to boost Italy's strengths and mitigate its weaknesses:

1. advance frontier research in Italy;
2. reduce AI research fragmentation;
3. develop and adopt human-centred and trustworthy AI;
4. increase AI-based innovation and the development of AI technology;
5. develop AI-driven policies and services in the public sector; and
6. create, retain and attract AI talent in Italy.

Together with the objectives and priority sectors, the Strategy is intended to apply to 11 priority sectors (amongst which industry and manufacturing, education, environment and infrastructure, banking, national security, and information technology) and 24 specific policy initiatives (including strengthening AI skills in different sectors, funding and enhancing research, launching private-public AI research-innovation calls, making AI a pillar that supports enterprises' Transition 4.0, promoting the country as the go-to-market of AI technologies, creating integrated datasets for Open Data and Open AI Models, and other policies specifically targeting enterprises and governmental bodies.

While the Italian legal framework appears to be in line with the EU's strategy of prioritising AI-focused policies and regulations, any concrete results have still to be seen and there are several indicators of where attention should be paid. Primarily, Italy currently lacks a ministry to take care of policies associated with AI: although it is true that the Department of Digital Transformation is an institutional body, it is not a Ministry and thus, politically, its power to legally regulate is less. Even though a ministry did exist until October 2022 (Ministry of Technological Innovation and Digitalisation), it has been dissolved by the current government in charge.

5 Policy recommendations for policymakers

A recent case in Italy sheds light on the urgent need for adequate AI regulation. In March 2023, the Italian Supervisory Authority temporarily banned ChatGPT, the well-known AI software capable of emulating and imitating human conversation, from processing data in Italy in breach of privacy laws. Specifically, the limitation order states that "no information is provided to users and data subjects whose data are collected by Open AI; more importantly, there appears to be no legal basis underpinning the massive collection and processing of personal data in order to 'train' the algorithms on which the platform relies" and that "the lack of whatever age verification mechanism exposes children to receiving responses that are absolutely inappropriate to their age and awareness, even though the service is allegedly addressed to users aged above 13 according to OpenAI's terms of service" (https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9870847#english).

This shows how the potential for non-regulated use of AI to pose a serious threat to individual rights and, specifically, generative AI – within the Metaverse alike, as shown in the previous sections.

What has been stressed and analysed in this chapter leads to several recommendations for policymakers, namely, the following specific appeals:

a) although the recent communication "An EU initiative on Web 4.0 and virtual worlds: a head start in the next technological transition" and the study on the metaverse issued in June 2023 and prepared by the European Parliament's Policy Department for Citizens' Rights and Constitutional Affairs on the request of the JURI Committee demonstrate the EU's starting willingness to commit itself to regulating the virtual worlds and the metaverse, little has actually been done so far. While it is true that several acts presently in force might be applicable to metaverse-related challenges, too many issues remain unsolved. European policymakers should hence focus on issuing legally binding laws for the member states regulating the metaverse and AI applications in the virtual worlds, specifically in the fields of copyright and IP law, privacy and data protection, marketing, and cybercrime;

b) noting what has been stressed concerning the obligation for Internet and hosting service providers to moderate illegal and harmful online content, it is crucial to find a similar solution when applied to decentralised networks where content removal by central entities has proven to be practically impossible. Although decentralised webs and apps are not yet as used by the common user/consumer, it is highly likely that along with technological development this trend will change, meaning that solutions have to be found urgently;

c) the potential implementation of AI components in the metaverse remains relatively undiscovered. We know that generative AI or NLP software could be used in the metaverse and the virtual worlds, but the extension of its uses are still unclear. Therefore, while assessing the metaverse as a whole and/or calling for action and regulations in the field policymakers should consider how AI may interact with the metaverse, and how such interaction might become relevant for tomorrow's society;

d) generally speaking, the last recommendation for policymakers is simply not to underestimate the metaverse's impact on society. Over the last 2 years, the metaverse's market impact has been swinging and related investments oscillating. This has led many to be sceptical and underrate the metaverse's actual impact on society, or to consider it as a new 'social network'. While this will surely become a fact, perhaps it is a fact already now. In the future, the development of AI and the technological progress will cause use of the metaverse and the virtual worlds by common Internet users to grow, which is when the major challenges described in this chapter will become more evident. This makes it crucial that policymakers constantly view these issues as a priority; and

e) specifically as concerns Italian policymakers, a revision of the national AI strategy is required. The current policies do not appear to be as effective for regulating AI in the long term, also noting the lack of a fully empowered ministry able to issue binding policies in the field. Moreover, regulation of the metaverse is also fully absent in political terms since no institutional formal communication expresses a willingness to regulate this field.

6 Conclusions

This chapter outlined what the metaverse is, its present and future uses, its interaction with AI, together with the associated threats and opportunities, and existing regulations (on both European and national (Italian) levels). The metaverse raises numerous opportunities for society and is potentially capable of shaping tomorrow's use of social interactions online. However, its associated challenges in several fields of law are being considered, as are the current legal gaps to be filled.

For these reasons, policymakers must view the recommendations as essential so as to try to prevent those issues that will become tomorrow's challenges – even though we might not know it yet: being proactive and looking at the future with the lessons of past experiences is fundamental, just like, as shown, all the issues associated with the non-regulated use of social networks.

The recommendations listed above might serve as a useful basis for EU and national policymakers to protect liberal values and civil rights in the field of artificial intelligence in the metaverse and the virtual worlds. Summing up the chapter's findings, it is essential that policymakers focus their attention on the field and come up with legal frameworks regulating use of the metaverse and AI in its several domains.

REFERENCES

- Aboukhadijeh, F. What is the Decentralized Web? 25 experts break it down. Syracuse University. Retrieved on 13 July 2023 from: https://onlinegrad.syracuse.edu/blog/what-is-the-decentralized-web/#:~:text=%E2%80%9CThe%20term%20'Decentralized%20Web',%2Dactor%20control%20or%20censorship.%E2%80%9D
- Anderson, J., Rainie, L. (2022). The Metaverse in 2040, Pew Research Center. Retrieved on 30 June 2022 from: https://www.mckinsey.com/~/media/mckinsey/business%20functions/marketing%20and%20sales/our%20insights/value%20creation%20in%20the%20metaverse/Value-creation-in-the-metaverse.pdf
- Ball, M. (2020). The Metaverse: What It Is, Where to Find it, and Who Will Build It. 13 January 2020. Retrieved on 3 July 2023 from: https://www.matthewball.vc/all/themetaverse
- Bodo, L., Trauthig, I.K. (2022). Emergent Technologies and Extremists: The DWeb as a New Internet Reality? Global Network on Extremism and Technology, ICSR, King's College, London. Retrieved on 10 July 2023 from: https://gnet-research.org/wp-content/uploads/2022/07/GNET-Report-Emergent-Technologies-Extremists-Web.pdf
- Bond, T., Stephens, K. (2022). Why IP lawyers need to pay attention to the EU's draft Data Act. Retrieved on 20 July 2023 from: https://www.twobirds.com/en/insights/2022/uk/why-ip-lawyers-need-to-pay-attention-to-the-eus-draft-data-act
- Carbone, M. R. (2022). Digital Markets Act, cosa dice la nuova legge: ecco l'impatto sui mercati digitali. 3 November 2022. Retrieved on 7 July 2023 from: https://www.cybersecurity360.it/news/digital-markets-act-cosa-dice-la-nuova-legge-ecco-limpatto-sui-mercati-digitali/
- Condemi, J. (2022). Digital Markets Act: cos'è e cosa prevede. 6 August 2022. Retrieved on 7 July 2023 from: https://www.agendadigitale.eu/mercati-digitali/digital-markets-act-cose-e-cosa-prevede/
- Chen, C., Zhang, L., Li, Y., Liao, T., Zhao, S., Zheng, Z. ... Wu, J. (2022). When digital economy meets web 3.0: Applications and challenges. IEEE Open Journal of the Computer Society
- Chen, D. (2022). The Metaverse is Here... But is the Hardware Ready? on 14 March 2022. Retrieved on 3 July 2023 from: https://www.spiceworks.com/tech/hardware/guest-article/the-metaverse-is-here-but-is-the-hardware-ready
- Cheong, B. C. (2022). Avatars in the metaverse: potential legal issues and remedies. International Cybersecurity Law Review, 1-28.
- Dwivedi, Y. K., Hughes, L., Wang, Y., Alalwan, A.A., Ahn, S.J., Balakrishnan, J., Wirtz J., et al. (2023). Metaverse marketing: How the metaverse will shape the future of consumer research and practice. Psychology & Marketing, 40(4), 750-776.

- Fernandez, C. B., Hui P. (2022). Life, the Metaverse and everything: An overview of privacy, ethics, and governance in Metaverse. in 2022 IEEE 42nd International Conference on Distributed Computing Systems Workshops (ICDCSW) (pp. 272-277), IEEE, July 2022
- Gadekallu, T. R., Huynh-The, T., Wang, W., Yenduri, G., Ranaweera, P., Pham, Q.W. ... Liyanage, M. (2022). Blockchain for the metaverse: A review, arXiv preprint arXiv:2203.09738.
- Grant, J. I. (2021). Removing the risks from a decentralised internet. The Strategic, Australian Strategic Policy Institute. 30 July 2021, retrieved 15 July 2023 from: https://www.aspistrategist.org.au/removing-the-risks-from-a-decentralised-internet/
- Haber, E., The Criminal Metaverse. 99 IND. L.J. (forthcoming 2024)
- Hooda, P. (2019). Comparison–Centralized, Decentralized and Distributed Systems. Retrieved on 5 July 2023 from: https://www.geeksforgeeks.org/comparison-centralized-decentralized-and-distributed-systems/#article-meta-div
- Jha, S. (2023). Web 3.0 Explained. A Comprehensive Guide, May 8 2023. Retrieved on 13 July 2023 from: https://www.simplilearn.com/tutorials/blockchain-tutorial/what-is-web-3-0
- Kasiyanto, S., Kilinc, M.R. (2022). The Legal Conundrums of the Metaverse. Journal of Central Banking Law and Institutions, 1(2), 299-322
- MacDonald, R. (2022). What Is the Decentralized Web (Web 3.0)? August 5 2022. Retrieved on 12 July 2023 from: https://www.1kosmos.com/blockchain/decentralized-web/
- Morgese, G. (2022). Moderazione e rimozione dei contenuti illegali online nel diritto dell'UE. 12 January 2022. Federalismi.it – Rivista di diritto pubblico italiano, comparato, europeo, ISSN 1826-3534
- Ramos, A. (2022). The metaverse, NFTs and IP rights: to regulate or not to regulate? WIPO Magazine. Retrieved on 12 July 2023 from: https://www.wipo.int/wipo_magazine/en/2022/02/article_0002.html
- Trifunovi⊠, D. (2021). Cybersecurity–virtual space as an area for covert terrorist activities of radical islamists. Teme-⊠asopis za Dru⊠tvene Nauke, 45(1), 95-109
- Turillazzi, A., Taddeo, M., Floridi, L., Casolari, F. (2023). The digital services act: an analysis of its ethical, legal, and social implications. Law, Innovation and Technology, 15(1), 83-106.
- Wahl, T. (2022). Rules on Removing Terrorist Content Online Now Applicable, EUCrim. 22 June 2022. Retrieved on 7 July 2023 from: https://eucrim.eu/news/rules-on-removing-terrorist-content-online-now-applicable/
- Werbach, K. (2018). The blockchain and the new architecture of trust. Cambridge. Massachusetts: The MIT Press. ISBN 978-0-262-03893-5. OCLC 1029064460.

Chapter 2

What makes AI regulation so difficult?

Hin-Yan Liu

1 Introduction

The modern field today known as Artificial Intelligence (AI) is widely regarded as having been launched during the Dartmouth Summer Research Project in the summer of 1956. In the ensuing decades, cycling through booms and busts, the milestones of what is to be considered as AI have faded to become mere computation as earlier goals have been achieved. The latest boom cycle has involved combining big data with deep learning techniques, driving frenzied interest in AI that has arguably culminated in the large language models represented by GPT-4 released earlier this year. Remaining agnostic to the claim that the spark of an Artificial General Intelligence may lay within GPT-4 (Bubeck et al., 2023), the hype and hysteria currently surrounding AI and its applications is abundantly clear.

AI applications permeate our world today: as a general-purpose technology like electricity (Lipsey et al., 2005), it can be usefully deployed for a vast range of human activity such that it is almost impossible to conceive of anything to which AI cannot be applied. The implication for law, regulation, and governance is that regulating AI is not possible.

Let me qualify this: in order to regulate AI in anything approaching the conventional manner, we must aim to either regulate the underlying technology, or aim to regulate how that technology is applied to the world. To regulate the underlying technology, we must be able to understand and describe what it is; that is, we need to be able to define, categorise, and communicate in precise language what it constitutes. Attempts to define AI have been notoriously elusive, and are usually tautological (e.g., "creating machines that perform functions that require intelligence when performed by people" (Kurzweil, 1992)) or anchored by reference to other ill-defined concepts (e.g., "making machines intelligent, [where] intelligence is that quality that enables an entity to function appropriately and with foresight in its environment" (Nilsson, 2010, p. xiii)). An aspect of why defining AI is so difficult is that there a bewildering array of definitions for "intelligence" (Legg & Hutter, 2007), suggesting that there is no real consensus on what this might be or comprise. To make matters more complicated, recourse to the concept of intelligence is itself only one of many possible metaphors we can deploy to understand AI. As each metaphor or analogy drawn for AI foregrounds certain characteristics and capabilities over others, each holds significant ramifications for AI regulation, a point that I will raise in detail later.

To illustrate the difficulty of defining AI in law, consider Article 3(1) of the draft EU AI Act which states that "artificial intelligence system" means:

... software that is developed with [specific] techniques and approaches [listed in Annex 1] and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with.

This approach does away with the metaphor of 'intelligence' altogether, thereby sidestepping many definitional pitfalls. Yet, in doing so, it foregrounds the contemporary basket of techniques that are currently driving AI forward (thus making the regulation brittle to technical changes), and renders neutral the other aspects of the process. Boiled down to the essentials, AI for European Law could simply be an input-output device with a few technical caveats. While this is as apt a definition as any might be for AI applications, there is little here that might indicate what the regulatory challenges and opportunities downstream may be since these rely at least partly on the metaphors and analogies that are deployed. The neutrality of the definition also neutralises our understanding and imagination as to why we might need to regulate in the first place.

A different way to understand the definitional impasse is to appreciate that this manifests the crux of the AI regulatory problem. On this approach, rather than attempting to resolve the definitional conundrum, staying with the trouble turns out to be more instructive. What we can learn is, at a minimum, that AI is and can be many things, that these things can change, and thus their impact on the world would likely also be startlingly diverse and dynamic. At a minimum, AI regulation must recognise and respond to this diversity and dynamism.

At a minimum, AI regulation must recognise and respond to this diversity and dynamism.

Thus, another strategy for regulating AI would be to regulate the use of AI applications, that is, how AI is deployed in the world. This approach avoids the definitional pitfalls by shoring up more general legal regulatory approaches to whichever activity area is concerned. This approach has been epitomised as the "Law of the Horse" by Frank Easterbrook in the context of cyberlaw, where he advanced the notion that "that the best way to learn the law applicable to specialized endeavors is to study general rules" (Easterbrook, 1996, p. 207). The assumption underpinning this view, however, is that the law is clear, competent, comprehensive, and complete: that is, capable of adequately and appropriately regulating all human activity. On this view, the application of new and emerging technologies does not present anything truly novel, and as such all we need to do is to ensure the proper functioning of the existing general legal rules to resolve any emerging issues.

While the Law of the Horse approach appears prudent, there are many missed opportunities to enhance our understanding of law and regulation by staying with the phenomenon and exploring its implications (Lessig, 1999). The requirement to accommodate new challenges into the existing legal system can overstretch legal categories and strain legal processes and, since legal accommodation tends to take place through metaphor, analogy, and interpretation, how a particular AI application and its challenges are accommodated becomes critically dependent upon how it comes to be understood in the law, as we will see later. This process of accommodation thus exposes the weaknesses of the legal system because how AI applications are treated is essentially arbitrary (insofar as a range of possible treatments becomes reduced to one particular legal view), and makes legal regulation brittle (in the sense that excluding alternative framings means that entire ways of understanding what AI applications are and which consequences this may hold are largely ignored).

One way of framing why regulating AI is so difficult is precisely that it is treated as an ordinary legal problem. If one were to take AI and its applications on one hand, and the law on the other, and smash them together, many discrete legal problems would fall out. Just to take a few familiar examples: in the field of intellectual property law, who owns a work that is created by an AI application? Under international humanitarian and human rights law, who is responsible for unlawful deaths and destruction? Under liability laws, if an AI application drives a vehicle, who is responsible for the accidents it might cause (H.-Y. Liu, 2016b. Furthermore, which legal regulatory challenges become obfuscated by focusing on the questions of responsibility and liability for accidents? See H.-Y. Liu 2018.)?

This familiar type of legal problem arising from AI applications being deployed in existing human activities is not inherently technological, but stems from the incompleteness, inadequacy, and indeterminacy of legal doctrine. Essentially, what is at stake is that AI applications are perceived to possess functional autonomy (in the sense of being able to operate in a non-deterministic manner) while the legal categories of agent and object remain mutually-exclusive (H.-Y. Liu, 2016a). This core problem also manifests itself in questions of rights and responsibilities with other functionally autonomous entities not imbued with full legal personhood, such as minors, animals, the incapacitated, and those with disabilities. If the incompatibility between functional autonomy and legal personhood is at the root of the regulatory challenge for AI, it should be clear that the law is not clear, competent, comprehensive, and complete.

Beyond this, I argue that AI and its applications are altering the range of affordances (capabilities and limitations), thereby presenting a new portfolio of possibilities and problems (H.-Y. Liu et al., 2020). Questions of legal regulation primarily address the problems that arise, but even so the new and expanded problem portfolio presented by AI and its applications is at best only partially covered by existing law. Hence, at the very least, AI and its applications are legally disruptive insofar as they reveal latent ambiguities within legal doctrine, or when they enable or facilitate new types of human activity that have hitherto been un- or under-regulated. (Multi-layered issues also arise for AI governance beyond the scope of this paper, see H.-Y. Liu & Maas, 2021.)

For the remainder of this chapter, I argue that AI provides further challenges for contemporary regulation and governance for three reasons: Change; Perceptibility, Awareness, and Understanding; and the implications of the Collingridge Dilemma. Change involves speed, scale, and type, but might be divided into incremental and disruptive change for the purposes of explaining why AI regulation is so difficult. An aspect of disruptive change that makes AI regulation difficult involves how many of the effects of AI applications are subtle or small, and therefore difficult to perceive, even as their effects accumulate in the same direction or converge towards the same endpoint. Exacerbating these effects, AI applications are often deployed in ways that mediate our interactions and experiences of the world, thereby affecting our awareness of their effects. And given the technical features of, and sophisticated narratives surrounding, AI and its applications understanding its processes and effects is also both limited and compromised. Furthermore, contemporary AI regulation and governance focus on problem-solving and thus take on a firefighting approach that reveals a lack of understanding of the worlds that we would like to build, and live in, with AI. These all represent a different type of problem for AI regulation and governance that make these endeavours difficult. To wrap this chapter up, I present the Collingridge dilemma for the social control of technology, which posits that "When change is easy, the need for it cannot be foreseen; when the need change is apparent, change has become expensive, difficult, and time consuming" (Collingridge, 1980, p. 11). While the Collingridge dilemma itself poses a formidable difficulty for AI regulation, I argue that we are simultaneously on both sides of the dilemma, and that the dilemma reveals implicit goals for the social control of technology that might not be achievable with transformational and disruptive technologies like AI.

2 Change as a source of difficulty for AI regulation

There are many ways to frame the regulatory challenges brought by AI and its applications. One popular way has been to focus on the technology itself: what is AI, and what can AI do? This approach situates the regulatory and governance questions internally within the technical aspects and characteristics of the technology, adopting the position that it is the features and contours of the technology itself that represent the actual challenges for legal regulation. Another approach, at the other end of the same continuum, is to ignore the technical aspects and look merely at the societal consequences that flow from AI and its applications. On this view, the technical aspects of the technology are uninteresting and irrelevant to the regulatory endeavour, therefore making it unnecessary to know or understand what the technology is and what it can do. Indeed, the focus on regulatory and governance questions means that the emphasis is placed downstream on the consequences rather than the source. For example, AI applications can be treated as being akin to magic: Arthur C. Clark, after all, famously quipped that "any sufficiently advanced technology is indistinguishable from magic".

As different as these positions might appear, the spectrum upon which they sit presupposes that something has meaningfully changed. Looking at the regulatory and governance challenges spurred on by AI through the lens of change might appear both superficial and obvious, but staying with this perspective may lead to important regulatory insights. For example, if the impetus for AI regulation becomes instead 'change regulation' we would able to cut through a lot of these AI-specific jargon-dense debates above. We would not need to overcome the definitional hurdles to regulate AI, nor describe the technical characteristics and parameters which might change as technology develops, nor would we be completely agnostic to the consequences that arise from the deployment and use of AI applications. Furthermore, the regulatory lessons learnt might be generalisable to other new and emerging technologies, saving us from re-inventing the wheel.

With respect to 'change' as the lynchpin for regulatory endeavours, AI and its applications present special challenges due to the speed, scale, and type of change they introduce (H.-Y. Liu, 2022). Still, before moving on to that, it is important to also draw attention to the baseline against which change might be measured (essentially 'change from what?'). Again, while this appears to be a very simple point, it holds implications for thinking about AI regulation. Since change cannot take place in a vacuum but requires a point (or more accurately a trajectory) of reference, the attention shifts to the baseline set of presumptions that undergird law and regulation. What emerges then is a picture of the legal and regulatory perspective of the world, and therefore how (technologically-driven) change might alter the trajectory and lead to regulatory challenges. A different way of putting this is to say that regulatory challenges arise in the distance between regulatory expectations on one hand, and actual (technologically-driven) realities on the other.

This means it is necessary to understand how this distance emerges, and why new and emerging technologies like AI increase this distance. At the outset, it is important to recognise that law,

With respect to 'change' as the lynchpin for regulatory endeavours, AI and its applications present special challenges due to the speed, scale, and type of change they introduce.

regulation, and governance essentially treat the world as being relatively stable: that is, past, present, and future are linear and that the rate of change is more or less flat. Yesterday is the same as today, and tomorrow will be the same as today. We can treat this as the baseline of expectations when it comes to law, regulation, and governance.

One view relating to the speed of change is the well-known "pacing problem" (Marchant et al., 2011). This views the distance between regulatory expectations and actual reality as being along the same continuum. The distance within the pacing problem is then essentially a race between technology and regulation. The mentioned problem posits that technology advances rapidly, while law, regulation, and governance are slower moving and therefore eternally playing catch-up. This is not helped by the fact that legal processes in particular often require manifest harm before being able to initiate any response: a retrospective orientation that effectively institutionalises the pacing problem in that legal regulation can only ever be responding to previous problems (Dershowitz, 2005). Even without this inbuilt structural disadvantage, the sheer speed of technological advances would leave an increasing distance that would be a source of legal disruption. Attempts to close this gap through anticipatory governance have been historically rare, and involved legal scholarship jumping the gun to think about the regulation of technological advances that appeared imminent at the time, but which subsequently took far longer to materialise or which have yet to be developed today (Picker, 2007. Unfortunately, these examples are provided to suggest an imprudent approach that has since fallen out of favour).

The distance within the pacing problem is then essentially a race between technology and regulation.

While the pacing problem is a prominent explanation of the growing gap between the rapid technological advances and sluggish legal regulatory responses, this remains relatively simplistic and it would be surprising if this were a one-dimensional phenomenon. Instead, we can plot linear, low to no rate of change, legal regulatory expectations against the non-linear and/or higher rate of change technological development in order to see the growing gap arise as a result.

One possibility that emerges is that, at first, technological capabilities may fall short of legal regulatory expectations and appear underwhelming. The effect is that that technology is written off as being a disappointment and falls off the regulatory radar. What is important about this is that there may be latent inadequacies and ambiguities within legal doctrine, but these remain hidden and irrelevant because the scope of affordances does not allow for the legal order to be sufficiently strained and tested. But as long as the rate of change is steeper for technological development than it is for legal regulatory expectations, there will at some stage be a cross-over point when technological capabilities outstrip legal expectations. The prospect for legal disruption arises after this point where surprise, and possibly chaos, reign. Whereas before, when legal regulation was more than sufficient for underwhelming technological capabilities, not only is the situation inverted, but the distance grows continuously since these are divergent trajectories. What is important to note here is that we do not need to buy the claims of exponential technological development as espoused by technological evangelists; mere advance will do as long as the rate of technological change is greater than that anticipated by law and regulation. To be even more precise, it is not even that we require technological development as such to occur: rather, all we need are new applications, and these can arise from new ways of exploiting and deploying existing technologies (Brynjolfsson & McAfee, 2014).

By considering the speed of technological change against the backdrop of legal regulatory expectations, we can begin to see that there is more to change than merely its speed. A different way to understand this is would be that, if it were only speed that was at stake, all we would need to do to close the pacing problem would be either to slow the rate of technological advances, or to speed up the regulatory processes. Despite the calls for a moratorium on AI development, or the development of more agile and adaptive regulatory processes, it should be clear that simply closing the gap between technology and law would be insufficient.

Part of the reason for this is that, at some point, mere quantitative increases in speed would translate into qualitative differences in the real or perceived effect downstream. Take mobile communications for example: 3G provided the necessary core network speeds to enable Internet connectivity and web browsing, 4G enabled buffer-free video streaming and provided the foundation for connected Internet of Things devices and services, and 5G for immersive augmented, virtual, and mixed reality, autonomous vehicles, and more. While these feel like quite very different things, very different types of capabilities in a qualitative sense, the lynchpin from a technical perspective is a simple quantitative increase of mobile connection speeds.

This conversion from mere quantitative technological improvement to perceived qualitative changes in the ensuing capabilities that the technology affords is one way of understanding the scale of change. In other words, regulation that merely responded to an increase in mobile connection speeds would comprehend the source of technologically-driven change, but be entirely blind to the societal developments built upon this seemingly banal change. This way of understanding the scale of change foregrounds the qualitative changes that are undergirded by quantitative improvements in the underlying technology.

In the case of AI, however, the scale of change can also take on an additional meaning. As a general-purpose technology, AI can be applied to virtually any domain of human activity. The result is that any technological advance in AI cascades into a vast range of possible applications in the real world. This sheer breadth of applications leads to a different type of scale in terms of change that regulation needs to grapple with, and in itself is potentially overwhelming. To get a sense obtain of this, we can recall the earlier examples that I gave where AI applications are collided with legal doctrine to produce discrete legal problems such as questions of ownership for AI creations, or liability questions for autonomous weapons systems and autonomous vehicles. The scale of change for law and regulation is readily seen since the application of AI in each discrete area of human activity will raise legal questions in the corresponding area of legal regulation. It may be overstating the case, but the scale of change brought about by AI and its applications could require a wholesale revision of virtually every legal area. Thus, the sheer breadth of AI applications can represent a paralysing range of changes which, taken together, suggests large-scale change.

When we take the speed and scale of change together, we can start to see differences in the type of change for the purposes of law and regulation. One way of thinking about this is to consider that "something has begun to change in change" (H.-Y. Liu & Maas, 2021). This meta-perspective of changing change converges with the idea of "turbo change" advanced by Daniel Deudney:

The features of turbo change – rapid rate, large magnitude, high complexity, significant novelty, and disruptiveness – make foresight of directions, assessment of consequences, and explanations of patterns very difficult (Deudney, 2018).

We are then able to put this together in the "change-stability matrix" to yield four different types of change: Changing change (Turbulence); Changing stability (Phase transition); Stable change (Incremental innovation); and Stable stability (Stagnation) (H.-Y. Liu, 2022). This view of the different types of change that impact law, regulation, and governance opens up different strategies for responding to AI-driven legal disruption. The change-stability matrix also reveals the limitations inherent to the orthodox legal approach, which implicitly assume Stability as the governing influence leading to either incremental innovation and stagnation.

The upshot is that 'change' as the governing influence leading to turbulence or phase transition remains beyond the contemplation of legal regulatory thinking. In other words, the linearity of legal expectations, discussed above, precludes the ability to consider both turbulent and phase transition types of change. The fact that legal regulation is incapable of contemplating different types of change like these renders the entire edifice of legal thinking and reasoning susceptible to systemic shock and legal disruption. In slogan form, legal regulation can contemplate incremental change, but is largely incapable of understanding or responding to disruptive types of change.

A different way of describing turbulent and phase transition types of change is to think of those that are producing a fundamentally different world (A particularly vivid, albeit fictional, example of a 'Changing-change' turbulent world is that of the Trisolarans in, C. Liu, 2014). This would be legal disruption writ large, a different type of revolution wherein the legal system is reimagined or simply set aside and replaced by competing forms of regulation, for example potentially technologically-driven non-normative forms of regulation such as technological management (Brownsword, 2019, 2022; H.-Y. Liu, 2022).

Taking change as the crux of the regulatory challenge posed by AI and its applications reveals multifaceted and deep-rooted issues that are missed by more familiar framings of the AI regulation and governance debate. If this were not difficult enough, we have to consider change within the broader array of global challenges that we are facing at the moment, most notably climate change and ecological devastation, heightened geopolitical tension, and ongoing advances in biotechnology and quantum computing, to name just a few. The point is that beyond AI and its applications, there are rapid and accelerating changes in a bewildering range of areas whose logics intersect and interact with AI and its developmental trajectory. To use the Manhattan Project as an analogy, the trajectory of nuclear fission technology would likely have taken a very different trajectory had it not been developed in the context of the Second World War and the technological arms race to produce a weapon that would end all wars (Bird & Sherwin, 2005). That some of the earliest legal discussions of AI entailed military applications and that questions of autonomous weapons systems (Bhuta et al., 2016) and the governance of military AI remain thorny questions (Maas, 2019a, 2019b) reveals a similar logic driving at least sectors of AI development and deployment. The broader geopolitical configuration characterised by anarchy and competition (Waltz, 1979) might usefully frame AI as a strategic technology that leads to its own set of perils and problems.

A different way of broaching the question would be to ask what might AI and its applications look like if not deployed within our contemporary constellation of economic and political logics that pursue private profitability (Zuboff, 2019) or political control (Dai, 2020)? In this line of thought, change exacerbates the differences between actual and potential trajectories of development, and enhances its own logics. Hence, AI that is developed in a competitive and conflictual context is likely to produce characteristics which reflect this developmental path, thereby producing applications that are predisposed to producing profit or providing political control. As a thought experiment to illustrate this idea, one can imagine that the path of big data and algorithmic processing could have taken a significantly different turn had it not converged with, or been captured by, the logics of capitalism (Cohen, 2019; Zuboff, 2019). From this vantage point, it becomes clear that any subsequent regulatory endeavour would be confined to merely tinkering and finessing the system but that the general trajectory has a locked-in path-dependent quality to it that has already been predetermined by the current economic and political context.

3 Perceptibility, awareness, and understanding underpinning the difficulty with AI regulation

Questions of AI regulation and governance therefore have a fractal quality, insofar as there are repeating patterns across different scales (Gleick, 1997; Johnson, 2002). Yet, only some of these scales are open for regulatory responses and governance debates as we have just seen. Perhaps it could not be otherwise without precipitating an economic paradigm shift or a true political revolution.

In keeping with this fractal perspective, perceptibility, awareness, and understanding in regulatory and governance terms only take place on particular levels and are largely isolated from others. This appears to be an obvious point: when we think about regulating AI, we are not generally imagining an entire upheaval of the social, political, and economic system. Instead, we are concerned with ameliorating some of the worst excesses that AI applications have produced. We can think about this as an exercise in alignment, to bring AI applications on par with human performance within human systems.

Take for example military applications of AI, and whether the use of autonomous weapons systems (AWS) can be lawfully deployed according to international humanitarian law (IHL). One camp has emphatically asserted that AWS can never satisfy the principles of distinction and proportionality in the use of force that underpin IHL (Human Rights Watch, 2012), while the opposing camp argues that there can be situations where AWS can be lawfully deployed and indeed perform better than human beings in combat (Anderson & Waxman, 2013; Schmitt, 2013). What is interesting about these debates is that they do not fundamentally concern legal questions, but are instead questions of technical capability: if AWS satisfy the legal requirements, they can then be lawfully used, but if these fall short they cannot be used (Heyns, 2016; H.-Y. Liu, 2016a).

Compartmentalising the introduction of AWS within the legal framework provided for by IHL excludes several significant challenges to law, regulation, and governance. These include: the concept of responsibility raised by the liminal position of AWS between the legal categories of agent and object (H.-Y. Liu, 2016a); the meaning of human dignity (Heyns, 2016); how models and metaphors enable or exclude regulatory thinking (H.-Y. Liu et al., 2019); and macrostrategic considerations pertaining to arms control (Maas, 2019a, 2019b).

These examples illustrate how, when deploying the lenses provided by law, regulation, and governance, only certain challenges can be identified. This is perhaps why AI ethics is discussed more than AI law and regulation, and why multi-level AI governance in general is exceedingly difficult to grapple with (H.-Y. Liu & Maas, 2021). This means that one type of difficulty pertaining to AI regulation and governance is that, tautologically, AI and its applications are examined under the rubric of regulation and governance such that only certain types of questions arise.

To unpack this further, and see why these are problems that are especially prominent with AI and its applications, we should delve further into each aspect of perceptibility, awareness, and understanding. The point here is that AI and its applications are altering the range and scope of affordances at a much more fundamental level (H.-Y. Liu et al., 2020), and that what we perceive to be legal regulatory issues can be understood as signs and symptoms of the underlying causes. Taking this view, given that legal and regulatory responses can only treat the symptoms while leaving the root causes unaffected, these responses can at best be symptom management strategies. This would mean that we should remain vigilant and adaptively responsive, since a cause can manifest different symptoms of differing severity. This suggests we need radically different legal regulatory styles than those provided for by the orthodox doctrinal view.

Another way of unpacking the difficulties for law, regulation, and governance introduced by AI and its applications is to look specifically at the challenges to perceptibility, awareness, and understanding (H.-Y. Liu, 2022). One can look at the arc of AI applications with respect to decision-making to illustrate this point. Early discussions concerned questions of fairness, accountability, and transparency where algorithms were deployed in decision-making processes, a process culminating in the European Parliament forbidding automated decision-making processes (Article 22 of Regulation (EU) 2016/679 (General Data Protection Regulation)). Thus, the readily perceptible issue was a familiar one: administrative and human rights law-related protections surrounding decision-making.

The point here is that the perceptible problems are by far the easiest to address: not only because they can be recognised as problems, but also since the process of recognition tends to then compartmentalise and contextualise that particular problem within a known discipline and discourse. So in the example above, issues related to algorithmic decision-making are transformed into a particular, but familiar, question related to decision-making. Similar legal questions would arise if the algorithm were swapped out for an oracle, for example, and indeed one of the ways of seeing the issue more clearly would be to ask which considerations and protections would have to be in place if decision-making were really done through ritual and oracle instead of through algorithms and computation.

When it comes to algorithmic processes and decision-making, the subject–object relationship presupposed by administrative law is only one possible framing. Another more pernicious possibility is that, rather than AI applications making decisions about us, AI applications instead interfere with our decision-making processes (Susser, 2019; Susser, Roessler, & Nissembaum, 2019; Susser, Roessler, & Nissenbaum, 2019). What is interesting about this frame is that it simultaneously captures an important part of the phenomenon (we can recognise the truth and applicability of this claim in the real world), while excluding legal consideration and response (it is very difficult to articulate precisely the legal problems that follow). Since the decision is not made about, and then imposed upon, us, the familiar hierarchical legal relationships are not directly applicable and so much of administrative and human rights law protections fall away.

One of the reasons why AI interference with our decision-making processes is imperceptible to the law is that legal doctrine recognises and protects the agency of the agent (legal person). Where a decision made about a person affects the agent from 'the outside', interference with our decision-making processes affects the agent from 'the inside'. In seeking to protect the agency of the agent, legal doctrine would paradoxically give effect to the agent (with or without interference in their decision-making processes).

Perceptibility becomes an even greater challenge with some AI applications, especially those related to virtual, augmented, and mixed realities (H.-Y. Liu & Sobocki, 2022). Since these applications mediate, in a very direct and real sense, one's sensory inputs it is possible to claim that these applications craft and create us (H.-Y. Liu, 2022; Seth, 2021). Obviously, if our sensory inputs are mediated by technological artefacts and processes, our ability to perceive their effects becomes neutralised. This is a deeper claim to algorithmic opacity than is often stated, and much more perilous since it disarms our very ability to perceive the problem.

One of the differences between perceptibility and awareness is that the latter involves a benchmark: awareness usually means that something is foregrounded against something else. Daniel Susser, in pointing to the invisibility of technology, emphasised that "once we are habituated to technologies we stop looking at them and instead look through them to the information and activities we use them to facilitate" (Susser, 2019, p. 1. Emphasis in the original). This is akin to the research milestones in AI that, once achieved, have been relegated to mere computation as research and development march ever onward. Familiarity leads to habituation, which dulls its salience (salience was also proposed as the focal point for law and regulation with regard to technological innovation, Balkin, 2015). As our world becomes increasingly infused with AI and its applications, we will become ever less aware of its impact and will instead see our world through the (distorted) lens of AI applications.

As our world becomes increasingly infused with AI and its applications, we will become ever less aware of its impact and will instead see our world through the (distorted) lens of AI applications.

Furthermore, AI and its applications raise a different issue of awareness, that of the poverty and insufficiency of the social values and legal concepts we rely upon for legal and regulatory responses. The traditional way of looking at legal regulatory responses has been to clarify the law to accommodate any new phenomenon within legally-cognisant terms. The underlying presumption is, as I argued above, that the law is clear, competent, comprehensive, and complete. From this approach, we become aware only of certain issues through the challenges posed by AI and its applications to existing legal doctrine, and these fade as legal regulation absorbs, accommodates, or adapts to these changes. Whereas before, awareness of the problem faded because of familiarity with the technology, here awareness fades since legal ambiguities and uncertainties have been resolved.

Yet, one of the legally disruptive aspects of AI and its applications is precisely that it enables a novel vantage point with which to examine legal doctrine. Using the examples above relating to autonomy and responsibility for AI applications such as autonomous weapons systems and autonomous vehicles, the problems surface not because of technological innovation but due to latent legal inadequacies and doctrinal uncertainties (H.-Y. Liu, 2016a, 2016b, 2018, 2019). Regulating AI and its applications is difficult because awareness splits the potential problem-space: those problems that are recognised are responded to and we acclimatise to their effects; those problems that remain outside of our awareness not only linger, but fester. But because we cannot recognise these deeper systemic and structural sources of problems, they continue to create further problems just beyond our awareness. Ways that we are able to redress this effect are to adopt different frames of reference, deploy alternative models and metaphors, and adopt problem-finding approaches. Effectively, we need as many different perspectives and different paradigms on the phenomenon as possible in order to expand and deepen our awareness of the problems brought by AI and its applications.

Another major difficulty with respect to AI regulation is understanding. This involves both understanding what AI is, and what it can do, but more importantly understanding what it is that we want an AI infused world to be like. The first is relatively straightforward, fuelled by both the aura of technical sophistication around ‘artificial intelligence’ and the surrounding popular narratives of devastation and dystopia. In theory, these can both be moderated by education and civic participation since these are practical problems.

More difficult is understanding the potential and problems posed by AI and its applications writ large. In part, this is because technology drives changes in our underlying system of values (Danaher, 2021; Danaher & Sætra, 2023). So what we have valued and sought to protect in the past is not necessarily a guide as to what will be valuable and in need of protection in the future. Take for example the right to privacy, which originated as a right against physical interference and intrusion, through technologically-driven questions of wiretapping and GPS tracking, to ubiquitous surveillance online. In each case, technology has changed the envelope of affordances and the right to privacy has had to adapt, but the underlying interest in privacy has also morphed due to what might be reasonably expected, and indeed, possible.

Since one suitable metaphor for AI applications is as an optimisation machine, it would do well for us to know the values we wish AI to optimise. But, the technologically-driven changes to our values system subvert this very possibility. We seem to know what we do not want an AI-infused world to be like, but we do not seem to be able to agree on how a 'good' world with AI would be like (H.-Y. Liu & Maas, 2021). In this sense, understanding the problems requires an understanding of the possibilities. It might turn out that, for example, asserting the protection of our existing interests and rights would be oblique, meaningless, or valueless in the face of both the problems AI poses and the possibilities it might usher forth. Not only might our contingently derived rights (Dershowitz, 2005) be antiquated, but it could also turn out that our hopes and desires for an AI infused world are underwhelming or under-ambitious. Failing to understand, and to explore, both the problem and possibility space hinders our ability to govern AI to prevent or minimise harm and to extract the benefits and build a truly 'good' world.

4 The Collingridge Dilemma as a way of explaining AI regulation is difficult

It is worth setting out the original dilemma on the social control of technology again: "When change is easy, the need for it cannot be foreseen; when the need change is apparent, change has become expensive, difficult, and time consuming" (Collingridge, 1980, p. 11). One of the examples that David Collingridge gives concerns the introduction of the automobile, and is worth setting out in full below:

The British Royal Commission on the Motor Car of 1908 saw the most serious problem of this infant technology to be the dust thrown up from untarred roads. With hindsight we smile, but only with hindsight. Dust was a recognised problem at the time, and so one which could be tackled. The much more serious social consequences of the motor car with which we are now all too familiar could not then have been predicted with any certainty. *Controls were soon placed on the problem of dust, but controls to avoid the later unwanted social consequences were impossible because these consequences could not be foreseen with sufficient confidence.* (Collingridge, 1980, pp. 16–17. Emphasis added)

This is a stark lesson for AI regulation: if an expert committee is unable predict the problems with something so seemingly simple as the motor car, what hope do we have for a truly transformative general-purpose technology?

Unpacking the Collingridge dilemma suggests that it is simultaneously too simple and overly complex as a theoretical framework for understanding why AI regulation is so difficult. It is too simple because it views regulation as a one-shot, mutually-exclusive, activity. The motor car example is, of course, a caricature not least because neither the dust nor the responses to dust feature prominently in the contemporary regulation of vehicles. One could even make the case that the regulation of dust being thrown up from untarred roads has been so successful that it is no longer a problem that we recognise today and, as a result, contemplating those early problems as problems appears ridiculous (similar examples can be found in Taleb, 2008). In this vein, we should be aware of the iatrogenic effects of prior regulatory responses, which would have an influence over the types of problems that ensue.

More to the point, it is not as if we only had a single opportunity to get the regulation of the nascent motor car right just as its use was becoming more widespread. As we have seen above, legal regulation may be limited to responding to problems that emerge from technological development, and may do so at a slower pace, but there is an iterative feedback cycle of responding to problems as they appear. Looking at the other side of the dilemma, responding to the manifest problems introduced by the motor car is not only expensive, difficult, and time-consuming: I would argue that it is simply impossible. Its introduction fundamentally reshaped our bodies and minds, our physical environment, our societies, our values, our economic and political priorities, and other fundamental aspects of contemporary civilisation. It is a point perhaps best made with reference to Douglas Adams, who has his character, Ford Prefect say that he thought that cars were the dominant life form on earth (Adams, 1979. Not to mention, of course, that the Vogons were slated to demolish Earth to make way for a hyperspace bypass.). In a nutshell, the motor vehicle has refashioned our world. The Collingridge dilemma, by segregating a 'before' from an 'after' of technological deployment, overlooks that transformational technologies radically transform the world in ways that preclude social or regulatory control. One lesson we can learn for AI regulation might be that regulation or control were never possibilities in the first place: rather, the question has become one of how to live in an AI infused world? We might also ask what a 'good world' would be like with AI? And how we might want to relate with AI applications, and to relate with each other mediated through AI applications (Balkin, 2015)?

If we take the Collingridge dilemma at face value we might learn more generalised lessons regarding the social control of technology. I would like to separate out three implicit features of the dilemma: first, that it is a mutually-exclusive dilemma; second, that there is only one dilemma at a time; and third, that there is only one technological innovation at a time. I would argue that the regulatory picture is significantly more complex than David Collingridge envisaged.

The dilemma posits two antagonistic positions pertaining to the 'information' problem and the ‘power’ problem, but in reality it is obviously not that we are completely in the dark about which sorts of problems might arise, nor are we completely impotent when responding to manifest problems as these arise. Rather than a dilemma, perhaps this is better conceived of as a spectrum, and one upon which we oscillate back and forth in homeostatic fashion. As a new technology is introduced, our insight and understanding regarding the potential problem and opportunity space is limited, and even minor divergences between predictions and the emerging reality will lead to a vast gulf, as David Collingridge described. If dust thrown off roads is a problem, and legal regulation responds to that problem, its salience will decrease. At the same time, other problems would emerge, ranging from pollution, to architectural exclusion (Schindler, 2014), which as they appear might also trigger responses (Dershowitz, 2005). Much less than a dilemma where we inexorably move from the ‘information’ problem to the ‘power’ problem, a better analogy might be the thermostat where the position we occupy traces a sine wave function. This tells us that we can never ‘solve’ technological challenges: instead, we should seek a stable yet dynamic equilibrium between innovation and control.

Second, given that AI can be modelled as a general-purpose technology and that its applications can be integrated into almost any form of human activity, it presents a vast range of potential social impact. Unlike the motor car, which is a relatively discrete technological application, AI and its applications will spawn Collingridge dilemmas with each use case. While looking at a Collingridge dilemma for AI in the abstract might lead to some generalisable insights, the dilemma only really works for technological applications. In the original example, it is the motor car as an application and not the internal combustion engine as the underlying technology, that was given. The obvious implication for AI regulation

One lesson we can learn for AI regulation might be that regulation or control were never possibilities in the first place.

is that we would be confronted with a huge array of largely unrelated Collingridge dilemmas, each initiated by the introduction of AI applications in a given type of human activity. To further complicate this assessment, we can find ourselves in different positions on different spectrums, since for example, knowledge and familiarity with AI applications in weapons systems might not shed light on the ramifications raised by generative AI applications. This appears to be a disadvantage because we might not be able to apply the lessons that are learnt from different domains, but it might also prove to be an advantage since we would be able to test out a wide range of regulatory strategies across different sectors. If there is one lesson to be learnt here, however, it might be that a single unified AI regulatory strategy might prove unwise because it would reduce all this complexity and put us into an actual Collingridge dilemma.

Third, the Collingridge dilemma is overly reductionist by treating technological innovation in isolation. In slogan form: just regulating AI is too easy. The combinatory and interactive effects between technological applications alone yield dramatic societal transformation (Brynjolfsson & McAfee, 2014). For example, what happens if good old-fashioned AI can be run off of quantum computers? Or if generative AI applications recursively improve themselves? Couple these technological advances with our contemporary economic logics and geopolitical realities, as well as our global challenges, and we can see that AI and its applications are perhaps unique in being both the source and potential solution to our present predicament. Technological unemployment provides a good example: regulating AI as either a technology or an application would largely miss the societal ramifications of widespread automation made possible by AI introduced into our present economic model, reverberating through to the question of what it means to be human (Danaher, 2019). In short, the point is that regulating AI qua AI will be overly narrow.

What the Collingridge dilemma does well is to draw attention to the tension underlying the regulatory endeavour, and to highlight the difficulty of timing interventions well. What we should be mindful of is that the dilemma is just one way of modelling the regulatory challenge and that, like all dilemmas, the logical tension is produced by the parameters and the framing. Rather than worry about the lack of information or the lack of power to alter the path of technological progress, we need to design regulatory systems that seek homeostatic equilibrium. To do that, there must be constant monitoring and adjustment (just like in a thermostat) to minimise the severity of the oscillations because it is the amplitude differentials that lead to societal disruption. Furthermore, since it is the societal impact that is of concern, we should look at how technological capabilities are actually applied and how their interaction with social, economic, and political factors might generate challenges for AI regulation and governance.

5 Concluding thoughts

It may be trite to state that regulating AI is difficult. And it may be controversial to claim that we will not get AI regulation right. Taken at face value, this might come across as being depressing and discouraging, that we are stumbling towards dystopia and have little power or control. But rather than inducing paralysis, or continuing with business-as-usual forms of denial, there may be important lessons to be learnt from staying away from the trouble (Haraway, 2016).

In examining why AI law, regulation, and governance are difficult, my aim in this chapter has been to move past well-trodden law and policy debates and reflexive regulatory responses. While such thinking may be necessary, it remains insufficient for the challenges and opportunities that AI and its applications introduce. Often, this has involved deploying a new perspective or frame, substituting the metaphor or analogy, or stepping back to focus on the complexity, interactions, and the emergence of behaviours and outcomes.

But perhaps the greatest difficulty remains that we have neither a clear concept and vision of what a 'good' world with AI would be like, nor the desire to attempt to build towards such a world. Instead, AI regulation and governance are firmly retrospective, responding to the last worst controversy. In doing so, we overlookthe fact that AI does not have mere generative abilities, but possesses truly constructive potential. The upshot is that AI will play a definitive role in the future worlds we inhabit.

If we accept that AI is a truly transformative technology, the goal of AI regulation and governance should not be one that is restricted to conserving and preserving today's world – that would be in explicit contradiction to the transformational power of the technology and would ignore the changing change taking place across the board. Rather, AI regulation and governance should seek out the features and characteristics of what a 'good' world would be like to live in with AI and its applications. This is a very different endeavour than the problem-solving responses that have dominated this space today.

REFERENCES

- Adams, D. (1979). The Hitchhiker's Guide to the Galaxy. Pan Books.
- Anderson, K., & Waxman, M. (2013). Law and Ethics for Autonomous Weapon Systems: Why a Ban Won't Work and How the Laws of War Can (Jean Perkins Task Force on National Security and Law). Hoover Institution, Stanford University.
- Balkin, J. M. (2015). The Path of Robotics Law. California Law Review Circuit, 6, 45–60.
- Bhuta, N., Beck, S., Geiβ, R., Liu, H.-Y., & Kreβ, C. (Eds.). (2016). Autonomous Weapons Systems: Law, Ethics, Policy. Cambridge University Press.
- Bird, K., & Sherwin, M. J. (2005). American Prometheus: The Triumph and Tragedy of J. Robert Oppenheimer. Knopf Doubleday Publishing Group.
- Brownsword, R. (2019). Law Disrupted, Law Re-Imagined, Law Re-Invented. Technology and Regulation, 10–30.
- Brownsword, R. (2022). Law, authority, and respect: Three waves of technological disruption. Law, Innovation and Technology, 14(1), 5–40.
- Brynjolfsson, E., & McAfee, A. (2014). The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies. W. W. Norton & Company.
- Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y. T., Li, Y., Lundberg, S., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4 (arXiv:2303.12712). arXiv. https://doi.org/10.48550/arXiv.2303.12712
- Cohen, J. E. (2019). Between Truth and Power: The Legal Constructions of Informational Capitalism. Oxford University Press.
- Collingridge, D. (1980). The social control of technology. Frances Pinter.
- Dai, X. (2020). Toward A Reputation State: A Comprehensive View of China's Social Credit System Project. In O. Everling (Ed.), Social Credit Rating: Reputation und Vertrauen Beurteilen (pp. 139–163). Springer.
- Danaher, J. (2019). Automation and Utopia: Human Flourishing in a World without Work. Harvard University Press.
- Danaher, J. (2021). Axiological Futurism: The Systematic Study of the Future of Human Values. Futures, 132, 102780.
- Danaher, J., & Sætra, H. S. (2023). Mechanisms of Techno-Moral Change: A Taxonomy and Overview. Ethical Theory and Moral Practice.
- Dershowitz, A. (2005). Rights from Wrongs: A Secular Theory of the Origins of Rights. Basic Books.

- Deudney, D. (2018). Turbo Change: Accelerating Technological Disruption, Planetary Geopolitics, and Architectonic Metaphors. International Studies Review, 20, 223–231.
- Easterbrook, F. (1996). Cyberspace and the Law of the Horse. University of Chicago Legal Forum, 1996, 207.
- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).
- Gleick, J. (1997). Chaos: Making a New Science. Vintage.
- Haraway, D. J. (2016). Staying with the Trouble: Making Kin in the Chthulucene. Duke University Press.
- Heyns, C. (2016). Autonomous weapons systems: Living a dignified life and dying a dignified death. In N. Bhuta, S. Beck, R. Geiß, H.-Y. Liu, & C. Kreß (Eds.), Autonomous Weapons systems (pp. 3–20). Cambridge University Press.
- Human Rights Watch. (2012). Losing Humanity: The Case Against Killer Robots. Human Rights Watch and International Human Rights Clinic.
- Johnson, S. (2002). Emergence: The Connected Lives of Ants, Brains, Cities, and Software. Simon and Schuster.
- Kurzweil, R. (1992). Age of Intelligent Machines. MIT Press.
- Legg, S., & Hutter, M. (2007). A Collection of Definitions of Intelligence. arXiv:0706.3639 [Cs]. http://arxiv.org/abs/0706.3639
- Lessig, L. (1999). The Law of the Horse: What Cyber Law Might Teach. Harvard Law Review, 113, 501.
- Lipsey, R. G., Carlaw, K. I., & Bekar, C. T. (2005). Economic Transformations: General Purpose Technologies and Long-Term Economic Growth. Oxford University Press.
- Liu, C. (2014). The Three-Body Problem. Tor Publishing Group.
- Liu, H.-Y. (2016a). Refining Responsibility: Differentiating Two Types of Responsibility Issues Raised by Autonomous Weapons Systems. In N. Bhuta, S. Beck, R. Geiβ, H.-Y. Liu, & C. Kreß (Eds.), Autonomous Weapons Systems—Law, Ethics Policy (pp. 325–344). Cambridge University Press.
- Liu, H.-Y. (2016b). Structural Discrimination and Autonomous Vehicles: Immunity Devices, Trump Cards and Crash Optimisation. In J. Seibt, M. Nørskov, & S. Schack Andersen (Eds.), What Social Robots Can and Should Do (pp. 164–173). IOS Press.

- Liu, H.-Y. (2018). Three Types of Structural Discrimination Introduced by Autonomous Vehicles. UC Davis Law Review Online, 51, 149–180.
- Liu, H.-Y. (2019). From the Autonomy Framework towards Networks and Systems Approaches for 'Autonomous' Weapons Systems. Journal of International Humanitarian Legal Studies, 10(1), 89–110.
- Liu, H.-Y. (2022). Rule-following robots? Transitional legal disruption through regulatee design and engineering. Law, Innovation and Technology, 14(1), 41–70. https://doi.org/10.1080/17579961.2022.2047518
- Liu, H.-Y., Maas, M., Danaher, J., Scarcella, L., Lexer, M., & Van Rompaey, L. (2020). Artificial Intelligence and Legal Disruption: A New Model for Analysis. Law, Innovation and Technology, 12(2), 205–258.
- Liu, H.-Y., & Maas, M. M. (2021). 'Solving for X?': Towards a problem-finding framework that grounds long-term governance strategies for artificial intelligence. Futures, 126, 102672.
- Liu, H.-Y., & Sobocki, V. (2022). Influence, Immersion, Intensity, Integration, Interaction: Five Frames for the Future of AI Law and Policy. In B. Custers & E. Fosch-Villaronga (Eds.), Law and Artificial Intelligence (pp. 541–560). Asser Press.
- Liu, H.-Y., Van Rompaey, L., & Maas, M. M. (2019). Beyond Killer Robots: Networked Artificial Intelligence Systems Disrupting the Battlefield? Journal of International Humanitarian Legal Studies, 10(1), 77–88.
- Maas, M. M. (2019a). How viable is international arms control for military artificial intelligence? Three lessons from nuclear weapons. Contemporary Security Policy.
- Maas, M. M. (2019b). Innovation-Proof Governance for Military AI? How I Learned to Stop Worrying and Love the Bot. Journal of International Humanitarian Legal Studies, 10(1), 129–157.
- Marchant, G., Allenby, B., & Herkert, J. (Eds.). (2011). The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem. Springer.
- Nilsson, N. J. (2010). The Quest for Artificial Intelligence: A History of Ideas and Achievements. Cambridge University Press.
- Picker, C. B. (2007). A View from 40,000 Feet: International Law and the Invisible Hand of Technology. Cardozo Law Review, 23(1), 149–219.
- Schindler, S. (2014). Architectural Exclusion: Discrimination and Segregation Through Physical Design of the Built Environment. Yale Law Journal, 124(6), 1836–2201.
- Schmitt, M. (2013). Autonomous Weapons Systems and International Humanitarian Law: A Reply to the Critics. Harvard National Security Journal Features.

- Seth, A. (2021). Being You: A New Science of Consciousness. Faber and Faber.
- Susser, D. (2019). Invisible Influence: Artificial Intelligence and the Ethics of Adaptive Choice Architectures. Artificial Intelligence, Ethics and Society. http://www.aies-conference.com/wp-content/papers/main/AIES-19_paper_54.pdf
- Susser, D., Roessler, B., & Nissembaum, H. (2019). Online Manipulation: Hidden Influences in a Digital World. Georgetown Law Technology Review, 4, 1–45.
- Susser, D., Roessler, B., & Nissenbaum, H. (2019). Technology, autonomy, and manipulation. Internet Policy Review, 8(2), 1–22.
- Taleb, N. N. (2008). The Black Swan. Penguin.
- Waltz, K. N. (1979). Theory of International Politics. Addison-Wesley Publishing Company.
- Zuboff, S. (2019). The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power. Profile Books.

Chapter 3

CISO, DPO, AIHO? Navigating the EU's AI regulatory efforts in pursuit of data protection and information security compliance

Tihomir Katulić

1 Introduction

After decades of theoretical discussion and scientific development, advancements in machine learning have finally found widespread commercial and governmental use facilitated by the continuous and rapid development of information technology. In the last year alone, dozens of commercial products as well as open-source technologies have been made available to the global public, fostering the development of sophisticated language-, image- and video-processing products and services applicable to many commercial uses (Acharya).

While large language models (foundation models) such as OpenAI ChatGPT or Google Bard, as well as publicly available models like Meta Llama are currently in the focus of experts and users, there have been (and probably will continue to be) various other approaches to developing AI which, alone or combined, may bring about general-purpose AI. For example, reinforcement learning is a sort of machine learning whereby an AI agent learns to make decisions by acting in the environment to accomplish a goal (Silver et al. 2019). The agent is encouraged to make better judgments over time by being rewarded or penalised (rewarded negatively) for its activities. In the last few years, significant advancements have been made in reinforcement learning as seen in DeepMind's AlphaGo's victory over Go human players, mirroring the success of IBM Deep Blue over then reigning chess world champion Kasparov in 1997. Reinforcement learning holds the potential to lead to the creation of AI systems able to learn from and adapt to their surroundings, making them useful for a variety of general-purpose jobs. Some AI scientists are experimenting with still different techniques such as transfer learning (applying the information obtained in one problem to another that is related but not the same, which is especially useful when data are limited or expensive to collect), capsulated (neural) networks, federated learning or different hybrid approaches. At this point, it is probably safe to conclude that should general-purpose AI arise, it will be the product of several existing and future development approaches, not any single one currently being investigated (Engelbrecht 2023).

Artificial intelligence (AI) also presents unique data protection and cybersecurity challenges, owing to its complexity, autonomy, and data-intensive nature, particularly given the need for massive data collection, the relationship between data quality and AI outputs, and issues with transparency and explainability. The inclination of AI systems to hoard data contradicts data minimization principles of data protection legislation. A fundamental difficulty is balancing the requirement for huge datasets with the desire to reduce data collection and retention. Furthermore, AI systems can be the subject of cyber-attacks such as data poisoning (changing training data to distort AI judgments), model stealing, and adversarial assaults (falsifying AI conclusions). It is vital to have strong cybersecurity to protect AI systems.

Advances in the previous two decades inaugurated pervasive information technologies such as broadband Internet, mobile connectivity, smartphones, mobile applications, social media platforms, cloud computing, augmented and virtual reality, and many others. Big Data is based on the premise of the availability of the collection and analysis of vast amounts of data, providing various actors, from the business community to governments and public institutions with unprecedented insights enabling better decision-making and targeted services. Any of these technologies by themselves, as well as all of them together, have influenced the pace of information society development and led to the development of even more innovative products and services, transforming society in previously unimaginable ways. Each of these technologies has significantly influenced the creation and expansion of information society services, and their continuing development and convergence has inspired even more innovative applications and services. As these technologies continue to evolve, they promise to further shape the information society in ways we can only begin to imagine, affecting the job market, the availability and quality of medical and education opportunities, and even the political process.

It is no surprise that the pace of the information revolution is accelerating steadily. Each subsequent phase of the change in how our civilisation collects, analyses and exchanges data has been shorter than the previous one (McGrath, 2013). Whereas it once took decades or years for an invention, product or service to reach the threshold of 50 million users, today it takes months or even days – as the recent case of Meta Threads reveals, having acquired 100 million users in less than 1 week. The progress of the information revolution is now on the doorstep of its final step – the advent of general-purpose AI, comparable and soon vastly outperforming human cognition, transforming our civilisation and billions of years of biological evolution into a technological one, with profound opportunities and potentially grave pitfalls.

> Each subsequent phase of the change in how our civilisation collects, analyses and exchanges data has been shorter than the previous one.

The shift to a post-industrial information society has spotlighted the importance of data as a resource whose processing serves as the foundation for new, cutting-edge information society services. The common European digital market is now host to many locally and internationally developed and deployed platforms and services, ranging from delivering information society services such as e-commerce, audiovisual content hosting, social networking and entertainment to smart city and Internet-of-Things applications generating an enormous amount of economically exploitable data. Large-scale data processing, especially in the case of the processing of personal data, while being a boon for information society products and services, simultaneously causes and entails many potential risks to the rights and freedoms of individuals.

The Charter of Fundamental Rights of the European Union recognises and simultaneously ensures the internationally highest level of explicit recognition of these rights, such as the right to privacy and right to the protection of personal data, as well as freedom of information, political activity, freedom from discrimination etc. These rights are today particularly endangered by intrusive surveillance technologies fostered by advances in information technology, largely in the ability to collect, process and store data seemingly without limits. These technical advances have already given rise to deeply disturbing use cases, from the Great Firewall and social credit systems of China to the insulated Internets of North Korea or Saudi Arabia (Roberts, 2018).

In the recent past, the European Union and its member states passed several significant laws to safeguard the acknowledged fundamental rights of individuals and govern the responsibilities of service providers to ensure safe and secure data processing. Currently, several legislative proposals are in the making, most notably the Artificial Intelligence Regulation, tasked with creating a social and business environment suitable to the development and deployment of new AI-based services, while also assuring the highest level of protection of fundamental rights of individuals – an approach standing in stark contrast with that taken by other large economic and political players.

In April 2023, the Chinese Cyberspace Administration published a draft document entitled "Measures for Generative Artificial Intelligence Services". Intended to control generative artificial intelligence products like ChatGPT, the proposed measures will contain guidelines that generative AI services must adhere to, including the kind of material these products are permitted to produce. In addition, the draft measures underscore concerns that the Chinese government holds regarding the use of generative AI, including transparency, algorithmic bias and prejudice, information distortion and abuse, and content regulation (Wu, 2023).

The United States, in comparison, presently seems more focused on promoting the benefits of developing and researching AI technologies

> The Charter of Fundamental Rights of the European Union recognises and simultaneously ensures the internationally highest level of explicit recognition of these rights.

rather than regulating against possible risks. The absence of any centrepiece statutory initiative to regulate artificial intelligence in the USA, one comparable to the EU's AI Act, often causes observers to either incorrectly believe that the USA has not taken any significant action on AI or to point to specific initiatives like the recent Blueprint for an AI Bill of Rights or AI Risk Management Framework as being representative of the overall US strategy. The current legislative framework acknowledges AI-relevant ethical principles (such as "bias", "privacy" and "explainability") without being specific about how they should be applied in the AI context. It views AI issues as ethical concerns in existing law ("civil rights") or agreeable high-level values ("trustworthy" systems, "responsible" use). In the end, this gives US government agencies both restrictions and flexibility. Without legislation that provides additional powers, agencies are forced to interpret their existing powers in order to control how AI is developed and used in industry (Pouget, 2023). However, they can maintain some discretion in determining how to handle this by being less prescriptive about how concepts pertinent to AI should be normatively implemented. This pragmatic approach, well understood in comparative data protection law, certainly holds merit from the perspective of liberalising the development and deployment of AI, favouring already invested Big Data stakeholders which are not exactly welcoming of the idea of stringent oversight and European-style administrative fines (Voigt, 2017).

Returning now to EU experiences with data protection, ever since it was introduced the General Data Protection Regulation has been responsible for a marked influence on businesses around the world, particularly online platform corporations with their headquarters in the USA, encouraging even scholarly research into privacy competition effects (Cooper, 2022). Article 3 of the GDPR states that, regardless of where an enterprise is situated, the GDPR places strict rules on data protection on all organisations that handle the personal data of EU citizens. Increased compliance costs, by way of corporations being compelled to establish thorough procedures to protect personal data, are some of the reasons that US-based Internet platform companies tried to undermine or lobby against the GDPR. While these costs entailed adding new IT systems, changing policies, training employees, appointing data protection officers and other relevant tasks that were and still are perceived by many in Big Data as an unnecessary administrative burden, at the same time the thriving EU NGO sector welcomed these provisions as a means for reigning in what was increasingly recognised as a callous, insatiable thirst for personal data.

In addition to offering new rights for individuals such the right to data portability, it explicitly affirmed existing rights like the right to knowledge, access and deletion. Many data controllers have found that it can be technically difficult and operationally complex to comply with these rights. In the event of non-compliance, the GDPR imposes strong fines of up to 4% of a company's annual global revenue or EUR 20 million (whichever is greater). These penalties might run into billions of dollars for major Internet businesses. For their business models, many Internet companies collected and analysed user data, especially for targeted advertising. The strict consent requirements and processing restrictions imposed by the GDPR have made it more difficult for some businesses to use data in the manners they were used to.

A key provision in several of these texts is the designation of a compliance expert that serves as a contact between various stakeholders – organisations, individuals, and regulatory bodies. The introduction of the General Data Protection Regulation (GDPR) in May 2016 had a considerable impact on the job market for legal services in the European Union (EU). This was chiefly due to the requirement for certain organisations to appoint a Data Protection Officer (DPO). The DPO as regulated by the GDPR has a mandate to advise, inform, monitor compliance, cooperate and consult with authorities and handle data subject requests (Lambert, 2016). The new AI Regulation proposal also contains certain references to the human oversight of AI systems. One goal of this paper is to explore what competencies and tasks await these experts and ways to regulate their position based on the experiences of half a decade of GDPR application. As the new legislative framework is developed and adopted, it will function alongside and complement the data protection framework already in place.

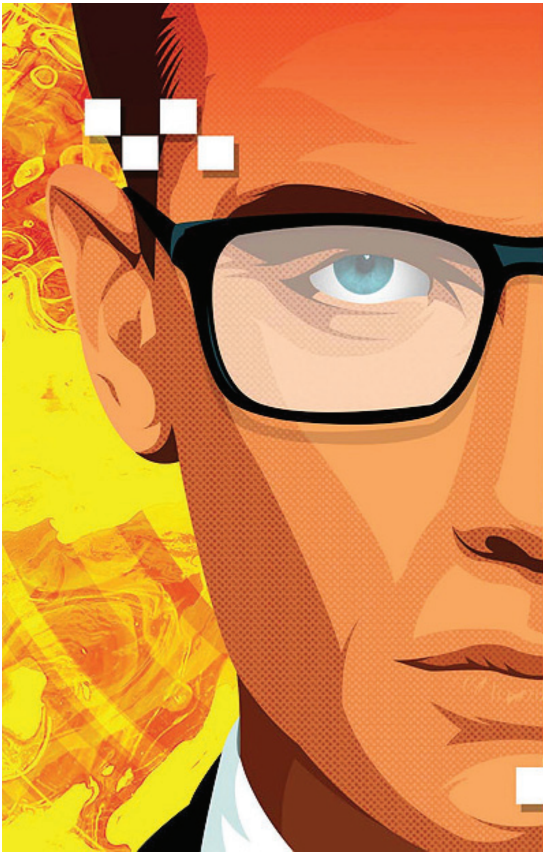
2 Position, competencies and experience with data protection officers and information security advisers

The proposed EU Artificial Intelligence Regulation (EU AI Act) currently being adopted by the EU Parliament and the Council is a complex new regulatory framework that addresses the development, deployment and use of machine-learning-based products and services in the EU's single market.

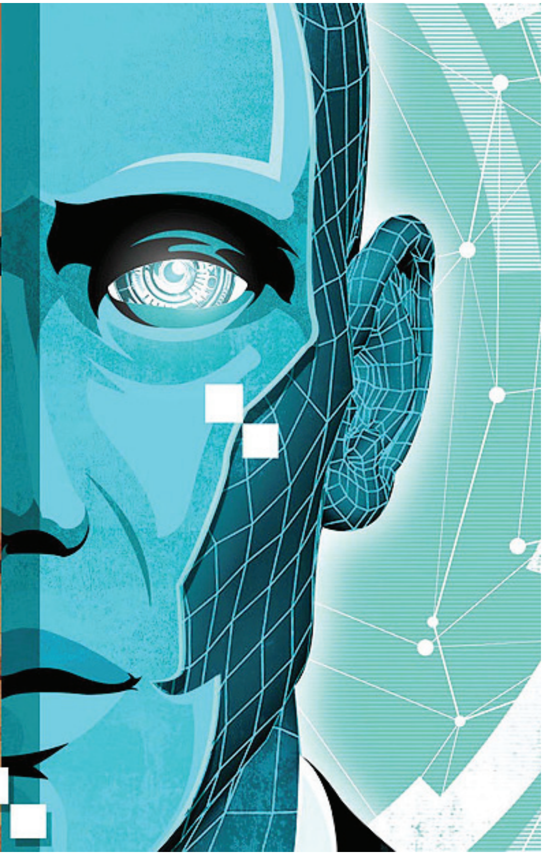
The complexity and obligations it will impose on AI system developers, distributors, deployers and users means it has faced strong opposition from stakeholders that fear additional regulatory burdens, limitations and restrictions concerning when and how AI systems can be used, large proposed administrative fines, and the usual period of interpretation and uncertainty that follows when norms of such complexity start to be applied.

Because of its complicated data processing capabilities and automated decision-making processes, artificial intelligence (AI) has the potential to endanger many of the General Data Protection Regulation personal data processing principles. Due to the complexity of AI algorithms, data controllers may find it difficult to explain how personal data is processed, thus jeopardizing transparency. Furthermore, if AI systems are educated on biased data, they may make conclusions that are unfair or discriminatory, breaking the establish standards of lawfulness and fairness. AI systems, particularly those that use machine learning, may repurpose data for training or other purposes that go beyond the scope of the original legal basis, possibly breaking the GDPR purpose limitation principle. Large amounts of data are frequently required by AI systems in order to train and develop their algorithms which may result in the collection of more data than is required for the specified purpose, contrary to the data minimization principle. Of course, when AI is based on erroneous, obsolete, or biased data, it can occasionally generate errors, particularly in decision-making processes. AI's reliance on large amounts of data for training and continual learning may result in longer data retention periods, which may clash with the GDPR obligation to keep data for no longer than is required for the purposes for which it is processed. Because of their complexity, AI systems may be vulnerable to security flaws, increasing the danger of unauthorized access, data breaches, or data misuse, threatening the integrity and confidentiality of personal data.

Finally, GDPR holds data controllers accountable for complying to its principles. However, because AI decision-making processes are frequently opaque, it can be difficult to show compliance or determine responsibility for decisions made by AI systems. To solve these issues, firms using AI have to implement strong data governance, ensure openness in AI processes, conduct regular audits, and keep clear documentation to confirm GDPR compliance. Furthermore, ethical AI design and early consideration of data protection considerations (privacy by design) are crucial.



The proposed AIA will impose various additional legal restrictions on AI systems, such as data governance rules, transparency standards, and conformance evaluations. These restrictions could put a heavy



administrative and financial strain on businesses, particularly smaller or newer ones.

Since the AI proposal clearly forbids the deployment of manipulative or exploitative AI activities as well as real-time remote biometric identification systems in public places, these and similar limitations will certainly be viewed in the Big Data industry as being too rigid, anti-competitive or as an outright hindrance for innovation. In a manner similar to the GDPR, the Artificial Intelligence Regulation prescribes significant fines in the event of non-compliance, this time up to 6% of total annual worldwide turnover for certain violations (Schuet, 2023). These possible fines can involve a sizeable financial risk, even if the EU's GDPR track record shows that it took over 5 years for EU data protection authorities to issue fines totalling over EUR 1 billion, issued to Big Data companies like Facebook/Meta with a proven track record of mostly ignoring data protection developments in the EU.

Without any serious intention to delve into the semantics of AI vs. machine learning terms (sentient general-purpose AI and AI cognition seem to remain a distant prospect for current advancements in information technology), the provisions of this new law will govern the adoption and oversight of the use of this technology in the next decade. With this proposal, EU legislators have several objectives in mind. One is to ensure that fundamental rights and values are respected during the development and deployment of AI systems. The second is to level the playing field for AI businesses operating within the EU. A number of provisions in the proposed AI Act are intended to assure that AI service providers do not endanger the established fundamental rights of individuals in the EU. Before these systems are employed, all AI systems must pass a risk assessment and, as part of this process, steps need to be taken to reduce any potential hazards that were found. In addition, consumers need to know about how AI systems will handle their personal data. The vendors of AI services must allow customers to contest the choices made by AI systems.

AI regulation efforts have aimed to solve many potential obstacles and problems. One of these is the pace of technological change. Simply put, technology is evolving faster than regulation, which makes it challenging for lawmakers to develop norms to apply to a rapidly changing technological landscape. Previously, member state lawmakers faced this problem in diverse areas like electronic communications, e-commerce and electronic signature regulations, data protection, intellectual property, cybercrime and information security – attempting to regulate existing technology in a directly related, technology-specific way typically vastly underestimates the pace of development, rendering such regulation obsolete before it even comes into power. A more abstract, technologically neutral approach is better, yet it also comes with substantial downsides – the abstract nature of such norms requires significant interpretation, mechanisms ensuring cohesive understanding and application, as most recently demonstrated by the efforts of the European Data Protection Board and member state data protection authorities regarding GDPR enforcement, and could pose problems for AI regulation as well (Pukhainen, 2021, Lerch, 2023).

Another obstacle to effective AI regulation is the inherent complexity and inscrutability of AI systems, at least from the perspective of the current technology and AI development approaches. The difficulties in understanding how machine learning technologies access data, analyse and reach conclusions greatly challenge lawmakers to develop adequate norms, and oversight and supervisory authorities to assess whether the product or service is behaving in line with such norms. This difficulty while dealing with data processing operations was already apparent in the application of key GDPR provisions, such as the data protection impact assessment, and will only become worse with even more opaque AI operations.

The GDPR extensively protects transparency as a fundamental principle of data protection. A transparent processing operation informs the data subject, the individual whose data is being processed, about the extent of such processing in a clear, simple and easy-to-understand manner, as well as about the identity of the data controller performing the processing. Transparency is considerably endangered by the current machine learning practices. A machine learning process is often described as a black box – its internal functioning, the exact how and why it produces the results it creates, is often difficult to understand even by experts, which makes it fiendishly difficult for authorities to scrutinise how and why a decision was made in order to ascertain whether a fundamental right has been breached. This becomes an even bigger problem when a system has evolved (been trained) on a dataset that may include biased data, carrying this bias into the decision-making stage. As machine learning products and services become ever more autonomous, taking over more and more responsibility from humans, especially in uses connected to transport, medical services, financial services or security and law enforcement, the question of liability for when things go wrong becomes increasingly salient. The complexity of these systems makes it very hard, if not impossible, for ex post analysis should adequate measures ensuring transparency and decision logging not be embedded into these systems during the development state. This mirrors the reasons explaining why privacy by design and by default provisions were considered and ultimately mandated by Article 25 of the GDPR.

Further, the global nature of this technology also brings a challenge. Regardless of the place of deployment, these technologies, like all Internet-enabled technologies, will exert an influence over individuals globally, requiring a significant amount of international consensus, namely, something that even well-established areas of fundamental rights regulation like privacy and data protection have yet to achieve.

There are also important economic ramifications of regulating AI. Should regulation be too restrictive, capital and entrepreneurship will find a way to develop and deploy in a more favourable legal forum, a country or a bloc with laxer rules and obligations, creating both economic and legal repercussions.

In comparative law, the position of Data Protection Officer (DPO) has been recognised as one of the key data protection compliance institutes since the 1980s. However, the General Data Protection Regulation (GDPR) adopted by the EU in 2016 introduced a fundamental change in the position, competencies and duties of a DPO, underlining the significance of this function. Especially when dealing with sensitive personal data, the DPO has become a fundamental part of the legal obligation data controllers must meet in order to effectively oversee data protection activities and

> In comparative law, the position of Data Protection Officer (DPO) has been recognised as one of the key data protection compliance institutes since the 1980s.

ensure compliance with the applicable data protection rules and regulations. Abandoning the previously often misused quantitative criteria for designating DPOs for a qualitative one, with the GDPR European lawmakers opted to require public authorities (and other data controllers satisfying the new qualitative criteria) to designate a DPO. Where some national laws required the data controllers to designate a DPO after meeting quantitative criteria like the total number of employees exceeding a certain number (e.g., designating a DPO was mandatory for data controllers with 20 employees or more under the Croatian Personal Data Protection Act of 2009), the criteria of the new Regulation consider the nature of the data processing activities, the status of the data controller (is it a public authority?) and potential risk.

The DPO plays a vital role in data protection compliance by ensuring that data controllers or processors understand the data protection requirements and perform their processing operations in line with the regulated principles of data processing and obligations mandated by the applicable regulation, from the GDPR to the sector-specific laws member states enact. One of the most important design decisions while developing the modern data protection framework was to regulate the DPO position as an informative and advisory role somewhat akin to the position of financial auditor – ensuring their independence, access to top-level management, and providing the officer with the resources required to help with compliance efforts as an adviser, not as a direct participant, due to the obvious risks of conflicts of interest. Another was to absolve and protect the DPO from liability for data breaches caused by the behaviour of the organisation – the DPO is solely responsible for adequately performing their duties. Finally, the provisions of Article 37 regulate the required competencies for a DPO; namely, understanding and knowledge concerning European data protection law and practice. When designating a DPO, the organisation must choose an individual with a proven understanding of the applicable European legal framework – both the GDPR and other applicable laws governing processing requirements in the organisation's field of activity, but also with practical skills in ensuring that these requirements are observed in the day-to-day operations of the organisation that they are advising.

The GDPR principles of confidentiality and integrity, as well as ultimately the principle of accountability, require data controller organisations to conduct specialised risk assessments regarding their processing activities. One task of the DPO is to assist with these activities by participating in data protection impact assessments, procedures created to identify and help mitigate the risk of data breaches in data processing activities. There are many situations where these procedures are mandatory and for the DPO to be able to contribute to them in a meaningful way they must be able to understand both the technical and legal risks involved. This is also apparent and required while planning and implementing procedures inside organisations such as information security or privacy policies designed to prevent or respond to data breaches.

The DPO is also tasked with cooperating and communicating with supervisory authorities and data subjects – individuals whose data is being processed. Occasionally, when the supervisory body comes to inspect the organisation's data protection practices or is responding to a complaint the DPO is the contact point for the supervisory body and can in various ways influence the inspection's outcome. Similarly, by serving as the contact point for data subjects the DPO is essential for protecting individuals' rights. As DPOs respond to requests made by data subjects, they are instrumental in enabling individuals to enjoy their rights, such as rights to information, access, rectification, erasure, and data portability regarding their personal data.

More succinctly stated, a DPO requires legal, business and information technology knowledge and skills, especially with respect to information security as a foundational principle of data controller accountability. These skills range from understanding the legal obligations to practical knowledge of safe data processing practices, recognition of the risks and ways to mitigate them, understanding and application of established industry standards, self-regulation as well as following the development of binding and non-binding guidelines, opinions and by-laws of national and EU supervisory bodies, such as the European Data Protection Board or the European Data Protection Supervisor. While this has repeatedly proven to be a tall order for the contemporary education system and one of the toughest and most divergent skillsets in demand in the present market, and not only in Europe, experience with compliance efforts shows that DPOs are invaluable. A similar role is already present in the proposed AI Regulation, although at the time of writing this paper it is underregulated.

The practice and application of the GDPR is in essence a textbook example of the observed

Globally, many countries from various legal systems and traditions have adopted or are adopting the new generation of data protection laws.

Brussels effect. Globally, many countries from various legal systems and traditions have adopted or are adopting the new generation of data protection laws. In order to satisfy the requirements of both their local laws and those of foreign markets in which they offer goods and services, such as the EU, organisations in these countries have also regulated similar positions. Many of them struggle to fill the DPO jobs, which was a concern while the EU was working on adoption of the GDPR.

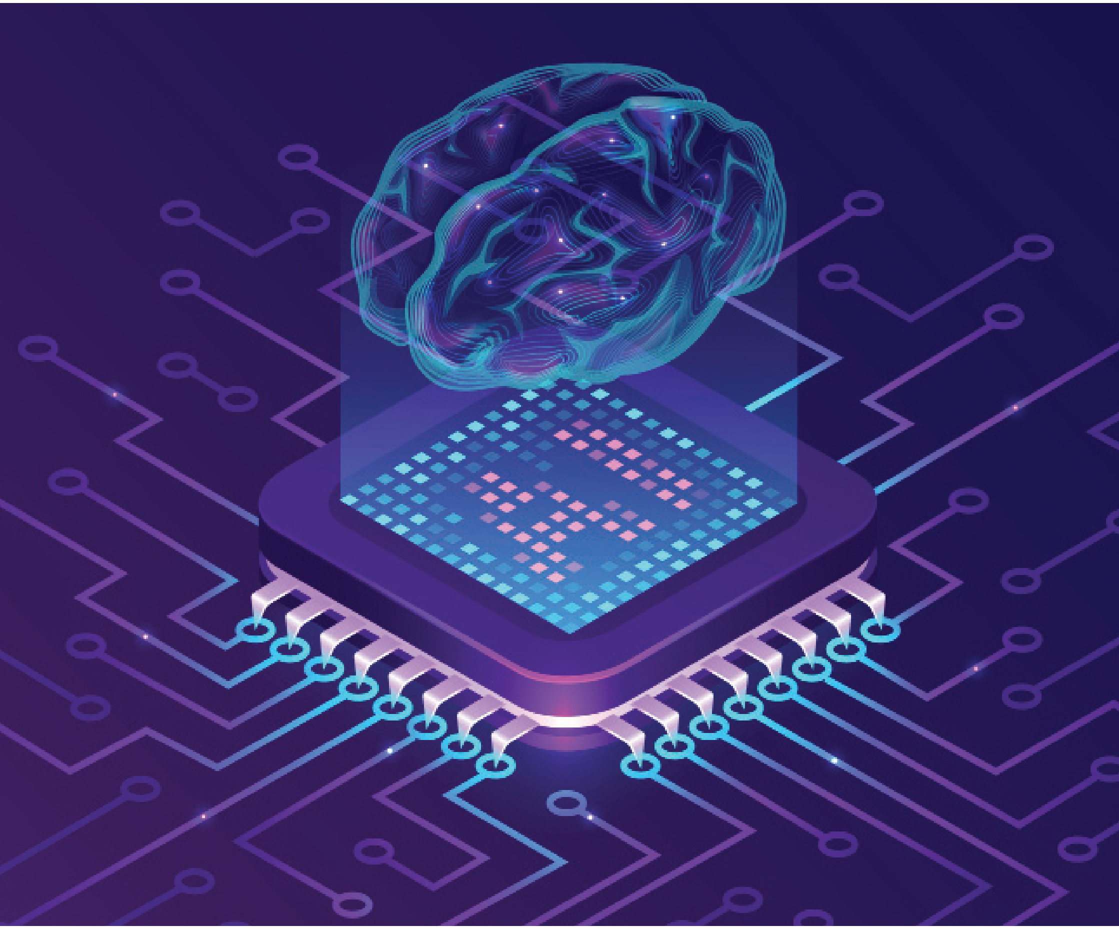
Detractors of the EU approach to data protection regulation, especially in the USA but often enough in various industry circles in the EU, also focus on the administrative burden of compliance. Still, the Regulation has affected the creation of hundreds of thousands of new jobs, notably in the information security industry. Several studies from industry associations like the IAPP and the European Commission itself suggested that the Regulation would create between 50,000 to 75,000 new DPO jobs globally (IAPP, 2017).

In reality, the total number of new DPO positions opened is probably several times higher, largely due to the inclusion of the criteria mandating public authorities to designate DPOs, but also thanks to the lively supervision efforts of European data protection authorities while issuing numerous administrative fines, the media attention following high-profile cases against leading Internet platforms, and the influence of the European Court of Justice's data-protection-related decisions. Some studies claim the actual number of designated DPOs in the EU exceeds half a million (Brook, 2019).

Similar to the GDPR DPOs, some member states had previously enacted national legislation concerning various information security requirements. Some of these laws, such as the 2007 Information Security Act (ISA) of Croatia, solely focused on infosec requirements for public authorities while others imposed obligations on any organisation operating (critical) infrastructure (Katulic, 2017).

With the recent adoption of the EU Network and Information Security (NIS) and NIS2 Directives, the EU has recognised the cybersecurity risks for key sectors of public infrastructure of the member states, and proceeded to mandate that service operators apply adequate technical and organisational measures. The NIS/NIS2 framework, however, does not contain a mechanism similar to the GDPR DPO, or CISO (chief information security officer) as acknowledged by established information security industry practices – or existing member state laws, such as the aforementioned ISA. While the position of information security adviser is not thoroughly developed by this or similar norms, it does serve as an example of introducing an additional layer of responsibility when employing critical information systems.

While the USA currently does not regulate DPO or CISO positions through the generalised systematic approach present in the EU's GDPR or member states' information security regulation, there are examples of similar provisions in



state laws or national security standards. A US law called the Health Insurance Portability and Accountability Act (HIPAA) establishes privacy guidelines to safeguard individuals' medical records and other health information given to insurance companies, physicians, hospitals, and other healthcare providers. Although the HIPAA does not expressly call for the appointment of a Chief Information Security Officer (CISO) or Data Protection Officer (DPO), it does set similar obligations for the covered organisations, e.g., to designate a privacy official to be in charge of creating and implementing the HIPAA policies and procedures in line with the law (Moore, 2019).

The US National Institute of Standards and Technology (NIST) has released a list of recommendations for information security officers. Similar provisions can be found in industry standards such as PCI DSS or the ISO 27001 family. The roles, responsibilities and duties of information security officers are clarified by these rules. Organisations that gather personal information from the residents of California are also required to have a DPO by the California Consumer Privacy Act (CCPA). The DPO is in charge of making sure the company complies with the CCPA's privacy regulations. Given that more than a dozen US states are currently in the process of developing data protection regulation, one may safely expect that at least some of these laws will include provisions on similar compliance mechanisms (NIST 2020).

In general, while CISOs perform many functions in a modern organisation, a few of these are vital for maintaining adequate information security and ensuring cyber resilience in modern business organisations. One of them is the expertise and experience to identify and mitigate security risks – these information security professionals have deep understanding of potential information security threats and vulnerabilities and developed skills and acquired experience to choose, implement and monitor effective security controls, ranging from developing and implementing security policies and procedures for managing access to sensitive data, protection from malware and other cyber-attacks through to preparing procedures to timely and adequately respond to security incidents.

In larger organisations, CISOs are critical for establishing and maintaining security staff, conduct training and overseeing the handling of security incidents. To find and address security weaknesses, CISOs can assist with security audits and assessments. These inspections and audits can help to identify weak points with security and confirm that security guidelines are being followed. Even the most ambitious compliance programmes equipped to handle diverse threats with deep understanding of potential threat landscape sometimes ultimately fail if the organisation does not invest in permanent training and education, or continuous resourcing. Organisations invest too often in these activities spurred by imminent supervisory activity or recent attacks on themselves or another similar organisation merely to abandon these efforts at the first opportunity – information security practices are essential for the modern organisation to function, not a stop-gap measure to undertake once things have already started to go wrong (Karanja, 2020).

3 Challenges and opportunities in AI policy – establishing a framework for efficient human oversight

The 5 years since the GDPR has been in operation have taught all stakeholders concerned a range of valuable lessons about modern regulation and oversight in information technology law. Some issues that have cropped up over this 5-year period did not only pertain to data protection enforcement – those who practise in information technology law (IT-related aspects of intellectual property, cybercrime and cybersecurity, electronic commerce regulation, electronic media and Internet governance etc.) have seen issues like weak and under-resourced regulatory bodies, normative complexity, and risks to fundamental rights emerge over and over again.

On the surface, the lessons these legislative experiences hold and teach for future AI regulation attempts are straightforward and clear – and already at least notionally present in the currently proposed AI Regulation. Independent internal oversight – the human in the loop – is a valuable mechanism for ensuring compliance and preventing potentially devastating data breaches. The proposed AI Regulation Act of the EU, as presently adopted by the EU Parliament and on its way to the European Council for the next step in the adoption process, includes certain human oversight provisions (e.g., Article 14).

While these provisions are not sufficiently developed in the proposal, and certainly lack the clarity and extent of the DPO provisions of the GDPR, they are obviously meant to help ensure the safe development, deployment, and use of artificial intelligence services and products. At the moment, the proposal only requires human oversight for AI systems classified as high-risk, such as critical infrastructure systems, AI-assisted biometric identification systems, AI in the criminal justice system, health, finance, education or employment areas etc. The Regulation proposal requires AI systems to be designed in a way that allows humans to understand and explain how they make decisions enabling humans to challenge AI decisions that they believe are unfair or discriminatory.

How can humans safeguard and be safeguarded against machine-learning-induced algorithmic harm? The proposed AI Regulation builds on the existing European legal framework, which contains analogous provisions when regulating data protection, competition and platform functions.

> Another obstacle to effective AI regulation is the inherent complexity and inscrutability of AI systems.

These include the Data Protection Impact Assessment (DPIA) as part of the GDPR, the Digital Services Act's (DSA) risk assessment procedure, and the Conformity Assessment (CA) anticipated by the proposed AI Regulation (Calvi, 2023).

Having recognised the limitations of these procedures, some experts call for a more detailed interdisciplinary investigation, an Algorithmic Impact Assessment (AIA), which may eventually become a useful tool for assessing the safety of an AI product. Assessments like the AIA could help with obligatory monitoring and re-examination during the life cycle of AI systems, even after their application, encouraging the accountability of the developers, deployers and users of the AI system as well as public scrutiny (Calvi 2023, Hamon, 2022).

Another provision reminiscent of the GDPR is the proposed establishment of a new EU-level cooperation and oversight body – the European Artificial Intelligence Board (EAIB), with a similar position and powers as the European Data Protection Board (EDPB) established under the GDPR. The similarities are obvious – each body has been created to uphold and enforce major pieces of legislation related to the digital economy and society, and the individual's fundamental rights. The European Artificial Intelligence Board, proposed under the EU's Artificial Intelligence Act, oversees the AI regulatory landscape. The EDPB ensures the consistent application of data protection rules across the EU. The European Artificial Intelligence Board's primary responsibilities and competences include aiding the Commission in creating guidelines, specifications and other pertinent components related to application of the AI Act, offering opinions and advice to the Commission on any issue concerned with implementation of the AI Act, and enabling the uniform enforcement of the AI Act in all member states, by facilitating the exchange of information and best practices. Both the European AI Board and the EDPB have advisory functions and endeavour to assure that their respective legal frameworks are applied consistently throughout all EU member states. Both offer recommendations on best practices, express viewpoints to the European Commission, and promote coordination between national agencies (EDPB, 2021).

Still, certain differences can be found among these bodies as well. Although the proposed AI Act does not directly include the AI Board's involvement in the arbitration of disputes between national data protection bodies, the EDPB has proven to be very effective in this role, especially as regards the administrative fines procedures. Since it is in charge of all facets of data protection, the EDPB has a larger range of tasks than the EAIB yet, given that the EAIB is a more recent entity and as AI becomes more prevalent, it is expected that the EAIB will take on a more significant role in the future.

Human oversight can identify any biases or flaws that an AI system might overlook or unintentionally introduce, decreasing the likelihood of unfavourable or erroneous results. Because they have to understand the AI's decision-making process, human overseers can push for the creation of AI systems that are more transparent and comprehensible, which will enable better explanation. The decisions made by AI systems can be held accountable when humans are involved in the supervision process. By ensuring ethical and legal accountability, this can raise public confidence in AI systems and make sure that AI systems follow accepted moral guidelines and social standards.

Humans in the loop should be able to intervene if the AI's behaviour deviates from these norms, preventing the potential misuse or abuse of the technology, and can monitor for adversarial attacks or attempts to manipulate the AI system, enhancing the system's overall security and integrity. Finally, human overseers can provide the AI system with real-time feedback, guiding its learning process and helping to improve its performance over time. These are all arguments in favour of the notion that while regulating AI human oversight is vital for maintaining control over AI systems, ensuring their ethical and safe operation, improving their explainability, and guiding their learning and evolution.

Tasks entailing the human monitoring of high-risk AI systems are present in the proposed AI Regulation. Some of these are outlined in Article 16 and include the need for human oversight to be performed by a group of specialists qualified to comprehend the AI system and any potential threats. If the human oversight team thinks the AI system is making a mistake or is about to make one, they must be able to step in and correct the mistake while the human oversight team's choices to interfere with the AI system's operation must be justifiable and explained. To be able to fulfil these tasks, organisations developing, deploying or using high-risk AI systems may require help to design and implement an adequate human oversight process, train the oversight team on the specifics of the AI system in use and the particular potential risks of its use, provide guidance to the oversight team on how exactly to intervene with the system's performance and how to document, explain and justify their decisions.

Like with the case of establishing the competencies needed of DPOs before the commencement of the GDPR's application, a foreseeable issue with choosing the right experts for meeting the AI Regulation obligation will entail understanding the requirements of the human oversight function.

Notably, these experts will need to possess competency in various, previously usually not very related or connected areas of expertise, such as expertise in the understanding and function of AI systems, experience with establishing the function of human oversight and a working understanding of the legal and ethical principles involved with operating an AI system.

4 Policy recommendations for the EU, the member states, and others

In an increasingly multipolar world, the future of global AI governance does not rest squarely on the shoulders of EU and US regulators. While major efforts are currently being undertaken mostly in Europe, the USA has a distinct approach to technology regulation that serves its interest primarily by facilitating the unfettered development, deployment and commercialisation of information technology, as demonstrated in the last 50 years. After Brexit, the UK government promised a different way forward, more in line with the practices of other Commonwealth nations. Participating in the common market, however, requires application of the EU acquis communautaire and the case of data protection clearly shows that leaving the EU does not mean abandoning the GDPR if the UK is to keep its access to the common market.

Three years as well as three reform efforts later, the UK's GDPR is still a facsimile of the EU law with no realistic chance of substantially changing it in the near future. In April 2023, the UK government was cautioned that the proposed bill runs the risk of weakening consumer protections and making it more difficult to hold businesses accountable for their data practices. It was also warned of the enormous costs to UK businesses should data adequacy be lost, the difficulty of having to adapt to new rules so soon after the introduction of the GDPR, and also of the potential implications for individual rights.

In an increasingly multipolar world, the future of global AI governance does not rest squarely on the shoulders of EU and US regulators.

The criticisms are similar to those made by civil society organisations, which encouraged the government to "scrap this bill and begin again" (26 organisations, including the Open Rights Group, Privacy International, and Index on Censorship). The signatories asserted that the bill would give the government more unilateral authority, reduce citizens' rights to redress, and lead to inadequate oversight of data processing. They claimed these changes would disproportionately affect women, immigrants, racialised groups, and the LGBTQ community (ORG, 2023).

These approaches vary substantially. While the EU approaches the issue of AI regulation through specific legislation, this legislation is built on the already broad foundations set by previous efforts like the General Data Protection Regulation, Digital Services Act, Digital Markets Act, Digital Governance Act, Network and Information Security Directive etc., the USA again takes a more pragmatic, hands-off approach by choosing not to introduce specific regulation – hardly surprising given the palpable resistance there to data protection regulation or Internet platform regulation in general. The two legal regimes are already so far apart in their approach to regulating the effect information technology has on fundamental rights that common ground and alignment on an issue such as this seems a thing of the distant past, as repeated rulings of the European Court of Justice have shown in the cases Schrems, Schrems II and will probably continue with Schrems III in the very near future (NOYB, 2023).

Even if AI Regulation has not yet been adopted as EU law, there are already reactions to the approach the EU has taken, along with perceived 'improvements' that the EU could consider to "aid future cooperation" (Brookings, 2023). Depending on the perspective of the analysis, classic critique of the EU's systematic approach to regulation applies in this case as well. This situation may soon start to conjure up memories of the now almost completely forgotten E-Privacy Regulation, the substitute of the ePrivacy Directive (2002/58/EC, as revised by 2009/136/EC), commonly referred to as the "Cookie Directive". Although the GDPR and the proposed regulation have been developed to be consistent with each other, intense lobbying and disagreements between member states in the Council of the European Union about key aspects of the Regulation have prevented its adoption. A key point of contention was how to strike a balance between individuals' right to privacy and commercial interests, particularly those of companies that rely on online advertising. Some stakeholders preferred stronger measures required to preserve data protection rights and privacy, while Big Data lobbied against any harsher rules on tracking cookies and electronic communications as they could harm businesses (Mazurek, 2019).

A similar debate could again unfold here. More 'flexible sectoral implementation' of the AI rules, as some US-based analysts suggest, could in practice mean greater opportunities to circumvent the cogent EU rules – indeed, the EDPS and the EDPB have repeatedly commented on the proposed AI Regulation, as well as other proposed EU regulations (Digital Governance Act, Data Act etc.), detecting and criticising attempts to circumvent the GDPR standards.

Further, similar statements, such as "flexibly tailored ... to specific applications" and "manage harmonization so member state regulators do not implement high-risk requirements differently", may be at odds with each other if interpreted from an EU perspective based on the experience of applying the GDPR (Engler, 2023).

As the AI Regulation proposes the EAIB, it will probably not, certainly not unanimously and without considerable opposition from its participating members like the EDPS, encourage development that would undermine the half a decade of struggle to impose adequate understanding of the data protection principles on the member states as well as other EU institutions. Cohesive understanding of new rules, yes – circumvention of the established data protection (and other) standards, a resounding no.

Here are a few ideas that could relatively quickly and easily be achieved to positively impact the efficiency of the regulatory framework as well as to further the standards of the protection of individual rights, while not endangering the current balance between regulation and entrepreneurial freedom. Revisiting the DPO provisions, the information security adviser and human oversight in AI regulation and similar mechanisms is a relatively simple and straightforward approach with little political risk and can be accomplished through member state instead of European legislation. In practice, this would mean the further clarification of DPO responsibilities in national GDPR implementation laws as well as forthcoming legislation that will accompany the NIS2 transposition and future AI Regulation. Member states could, as some already have, introduce (non)obligatory certification schemes for DPOs, with the possibility to enact similar provisions for information security advisers and AI human oversight experts (CNIL, 2020).

While the public is still not privy to the results of this year's EDPB enforcement action – a pan-European obligatory DPO poll conducted by national data protection authorities initiated by the EDPB in May 2023, its focus on DPOs and especially their experiences concerning reporting directly to organisation management, understanding and the avoidance of conflicts of interest and performing mandated DPO tasks from previous similar research, it is safe to conclude that the results will point to areas where considerable improvement is possible (EDPB 2023). For example, member states might consider strengthening the provisions of their national implementation laws protecting DPOs from dismissal or retribution for performing their tasks which, by design, frequently put them at odds with management of the organisation.

For future AI Regulation, if the proposal recently adopted by the European

Parliament remains the same in this regard, national implementation laws may provide a clearer and more practical definition of the position of human oversight with respect to the position, the competencies required and tasks of these experts, analogously with the DPO provisions in Arts 37–39 of the GDPR.

Member states could also at that point take the opportunity to provide detailed instructions on the content and quality of the thorough audit records AI developers and deployers should keep that supervisors may go through in order to determine whether an AI system has complied with certain standards and how it arrived at a given choice. Another potential area of improvement would be introducing provisions that would enhance the ability of human supervisors to influence or overturn the judgments made by AI systems.

5 Concluding remarks

As with the case of data protection compliance and enforcement, regarding the case of future AI regulation the author strongly believes that the best and simultaneously only way forward for European legislation is to continue to strengthen and affirm the fundamental rights of individuals through elaborated protection and enforcement mechanisms, as has been the case with EU legislation practice thus far. There is something to be said for continuing consideration of technological and market developments the EU legislators practice for well over thirty years – starting with the incredible effect the Data Protection Directive produced in its heyday, laying the groundwork for the GDPR and its interdisciplinary approach to ensuring fundamental rights in the context of personal data based economy.

While the USA and the South East Asian technological powers continue to lead the way in technological development, outperforming Europe in the sense of both numbers of new products and services as well as founding numerous successful startup companies, the EU has chosen a different approach, sacrificing a small part of the otherwise very broad development and market freedom to ensure a constitutional prerogative – a safe haven of civil and human rights in a world increasingly threatened by the spectre of digital dictatorship and a techno-totalitarian state. By so doing, it has promoted liberal civil and the individual's rights and freedoms at a time where technology-fuelled tendencies towards authoritarianism, religious fundamentalism and outright dictatorship are slowly creeping back from the almost forgotten past.

There are certainly both positive and negative net effects of the contemporary 'digital' legislation adopted by the EU in the last decade in a range of areas from intellectual property and competition to data protection and cybersecurity. The functioning of the digital single market has accelerated economic and social changes, with its effects being visible on many fronts. These effects can be seen in the job market – where automation and digitalisation have had a chilling effect on the number of traditional jobs, even among the recognised professions. The new regulatory requirements have created massive work opportunities for a new generation of interdisciplinary experts combining legal, technical and operational knowledge – there could be as many as 1.2 million data protection officers working today in the EU (extrapolating from a recent national survey in Croatia and similar surveys in other EU countries like Poland and Czechia).

These experts, together with information security advisers and future AI human oversight experts, play a crucial role not only in ensuring organisational compliance, performing the mandated tasks and assessing the risk for the rights and freedoms of the individuals whose data is being processed, but also in raising the awareness of organisations, their employees and partners, and individuals whose data is processed as well. They must be better protected, resourced and educated in order to respond to the demands of tomorrow, especially with the proliferation of AI products and services.

As the level of rights-awareness in the general population grows, this will mean individuals' increased requests and inquiries about the processing of and access to their data, which in turn will foster stronger oversight activities and a further raising of compliance standards. This is already happening with data protection practices as demonstrated by the higher number of both fines issued, and their amount.

REFERENCES

- Acharya, A. Llama 2: Meta AI's Latest Open Source Large Language Model. Retrieved 29 July 2023 from: https://encord.com/blog/llama2-explained/.
- Silver, D. et al. (2018). A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. Science, vol. 362, pp.1140-1144.
- Engelbrecht, D. (2023). The Future of AI and Ethical Implications. In: Introduction to Unity ML-Agents. Apress, Berkeley, CA.
- McGrath, R.: The Pace of Technology Adoption is Speeding Up. Harvard Business Review, 2013. Retrieved 29 July 2023 from: https://hbr.org/2013/11/the-pace-of-technology-adoption-is-speeding-up.
- Roberts, M. (2018). Censored: Distraction and Diversion Inside China's Great Firewall. Princeton University Press.
- Wu, Y.: Understanding China's New Regulations on Generative AI. Retrieved 29 July 2023 from: https://www.china-briefing.com/news/understanding-chinas-new-regulations-on-generative-ai-draft-measures/.
- Pouget, H., O'Shaughnessy, M.: Reconciling the U.S. Approach to AI. Retrieved 29 July 2023 from: https://carnegieendowment.org/2023/05/03/reconciling-u.s.-approach-to-ai-pub-89674.
- Voigt, P., von dem Bussche, A. (2017): The EU General Data Protection Regulation (GDPR): A Practical Guide, Springer
- Cooper, J.C. (2022): Antitrust & Privacy: It's Complicated, University of Illinois, Journal of Law, Technology & Policy, Vol. 2022, pp. 343-397.
- Lambert, P. (2016): The Data Protection Officer: Profession, Rules and Role. CRC Press, Taylor&Francis.
- Schuett, J. (2023). Risk Management in the Artificial Intelligence Act. European Journal of Risk Regulation, pp. 1-19.
- Pukhainen, E., Väyrynen, K. E. (2021): The Benefits and Challenges of Technology Neutral Regulation – A Scoping Review. Twenty-fifth Pacific Asia Conference on Information Systems, Dubai, UAE.
- Lerch, P. (2023): All Agents Created Equal? The Law's Technical Neutrality on AI Knowledge Representation, 14 J. Intell. Prop. Info. Tech. & Elec. Com. L. 108
- IAPP: The GDPR Demands 75k DPOs. Retrieved 29 July 2023 from: https://iapp.org/media/pdf/DPA-Whitepaper.pdf

- Brook, C.: Half A Million DPOs in Place One Year Post-GDPR. Retrieved 29 July 2023 from: https://www.digitalguardian.com/blog/half-million-dpos-place-one-year-post-gdpr.
- Katulić,T. (2018): Transposition of EU Network and Information Security Directive into national law, 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO) Proceedings, pp. 1143-1148
- Moore, W., Frye, S. (2019): Review of HIPAA, Part 1: History, Protected Health Information, and Privacy and Security Rules. Journal of Nuclear Medicine Technology December 2019, 47 (4) 269-272
- National Institute of Standards and Technology (NIST) 2020: The NIST Privacy Framework: A Tool for Improving Privacy through Enterprise Risk Management. Retrieved 30 July 2023 from: https://www.nist.gov/privacy-framework/privacy-framework.
- Karanja, E. (2020): The role of the chief information security officer in the management of IT security. Information and Computer Security, Vol. 25 No. 3, pp. 300-329.
- Calvi, A., Kotzinos, D. (2023): Enhancing AI fairness through impact assessment in the European Union: a legal and computer science perspective. FAccT '23: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, pp. 1229–1245
- Hamon, R., Junklewitz, H., Sanchez, I., Malgieri, G., De Hert, P. (2022). Bridging the Gap Between AI and Explainability in the GDPR: Towards Trustworthiness-by-Design in Automated Decision-Making in IEEE Computational Intelligence Magazine, vol. 17, no. 1, pp. 72-85
- European Data Protection Board (2021): EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Retrieved 30 July 2023 from: https://edpb.europa.eu/our-work-tools/our-documents/edpbedps-joint-opinion/edpb-edps-joint-opinion-52021-proposal_en.
- Open Rights Group (2023): 26 Civil Society Groups call on Government to scrap Data Protection and Digital Information (DPDI) Bill. Retrieved 30 July 2023 from: https://www.openrightsgroup.org/press-releases/26-civil-society-groups-call-on-government-to-scrap-data-protection-and-digital-information-dpdi-bill/
- None Of Your Business (NOYB) (2023): EC gives EU-US Data Transfers Third Round at the CJEU. Retrieved 30 July 2023 from: https://noyb.eu/en/european-commission-gives-eu-us-data-transfers-third-round-cjeu

- Mazurek, G., Malagocka, K. (2019): Perception of privacy and data protection in the context of the development of artificial intelligence. Journal of Management Analytics, Volume 6, 2019.
- Engler, A. (2023): The EU and U.S. diverge on AI regulation: A transatlantic comparison and steps to alignment. Retrieved 30 July 2023 from: https://www.brookings.edu/articles/the-eu-and-us-diverge-on-ai-regulation-a-transatlantic-comparison-and-steps-to-alignment/#anchor8.
- Commission Nationale Informatique & Libertés (CNIL): Certification des compétences du DPO. Retrieved 30 July 2023 from: https://www.cnil.fr/fr/certification-des-competences-du-dpo-0.
- EDPB: Launch of coordinated enforcement on role of data protection officers. Retrieved 30 July 2023 from: https://edpb.europa.eu/news/news/2023/launch-coordinated-enforcement-role-data-protection-officers_en.

Chapter 4

Ethical challenges in AI governance and regulation in the EU: Lessons held by the fourth industrial revolution for Africa

David Mhlanga

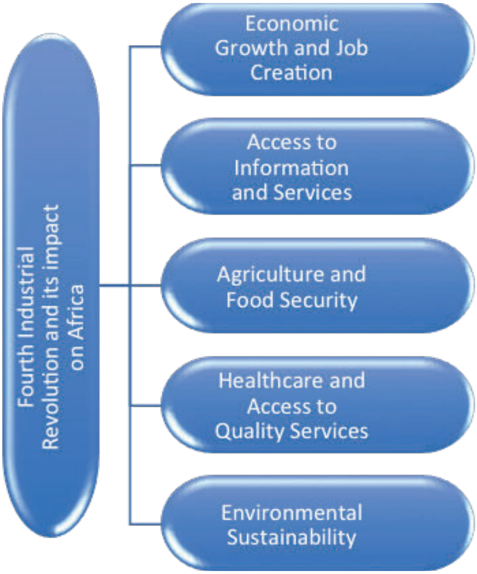
1 Introduction

The rapid advancements of artificial intelligence (AI) have revolutionised various industries and reshaped the global landscape, leading to the emergence of the Fourth Industrial Revolution (Shahroom & Hussin 2018; Koh et al. 2019). As AI becomes increasingly integrated into our societies, it brings with it a multitude of ethical challenges that demand careful consideration and effective governance and regulation (Du & Xie 2021; Mhlanga 2023). This is particularly crucial in the European Union (EU) where policymakers and stakeholders have been grappling with the ethical implications of AI and striving to establish a robust framework for its responsible development and deployment (Horgan et al. 2020; Holmes et al. 2021). With Africa standing on the cusp of its technological transformation and seeking to harness the potential of AI, it is essential to examine the ethical challenges encountered by the EU, and the actions taken to address the challenges to learn valuable lessons and adopt suitable strategies for effective AI governance and regulation. Africa's unique socio-cultural context and developmental challenges require a tailored approach that considers the region's specific needs and aspirations while ensuring that the ethical implications are adequately addressed. The purpose of this paper is to delve into the ethical challenges encountered by the EU in AI governance and regulation and identify lessons from these experiences with the Fourth Industrial Revolution that Africa can draw on. By closely examining the EU's journey in grappling with the ethical dimensions of AI, African policymakers and stakeholders can gain valuable insights into best practices, potential pitfalls, and regulatory frameworks that promote responsible and ethical AI adoption in the era of the fourth Industrial Revolution.

2 The fourth industrial revolution and its impact on Africa

The Fourth Industrial Revolution (4IR) is characterised by the fusion of digital technologies, such as artificial intelligence, robotics, the Internet of Things, blockchain, and advanced data analytics (Ndung'u & Signé 2020; Mhlanga 2022). Its impact on Africa holds the potential to transform various sectors, enhance economic growth and address societal challenges. Figure 1 below outlines the Fourth Industrial Revolution and its impact on Africa.

Figure 1: **The Fourth Industrial Revolution and its impact on Africa**


Fourth Industrial Revolution and its impact on Africa
Economic Growth and Job Creation
Access to Information and Services
Agriculture and Food Security
Healthcare and Access to Quality Services
Environmental Sustainability

Source: Author's analysis

The impact of the Fourth Industrial Revolution and its impact on Africa is shown in many scenarios which include its influence on economic growth and job creation, Access to Information and Services, Agriculture and Food Security, Healthcare and Access to Quality Services and Environmental Sustainability. These points are expanded on below.

2.1 Economic growth and job creation

The 4IR offers significant opportunities for economic growth in Africa. It enables the development of new industries and business models, fostering innovation and entrepreneurship (Magwentshu et al. 2019; Ramakrishna et al. 2020). By embracing digital technologies, African countries can leapfrog traditional stages of development and accelerate economic progress. The digital economy can contribute to GDP growth by promoting e-commerce, digital financial services, and technology-driven sectors. The 4IR creates fertile grounds for innovation and entrepreneurship in Africa. With digital technologies becoming more accessible and affordable, individuals and businesses can develop and deploy innovative solutions tailored to local needs. This, in turn, fosters a culture of entrepreneurship, as aspiring entrepreneurs can leverage digital platforms to launch and scale their businesses. Another important point is that embracing digital technologies enables African countries to bridge the infrastructure gap. By investing in broadband connectivity and digital infrastructure, governments can facilitate access to information and digital services, empowering individuals and businesses to participate in the global digital economy. However, there are concerns about job displacement due to automation (Graham 2019; Ye & Yang 2020). While some low-skilled jobs may be at risk, the 4IR also creates new job opportunities. African countries should prioritise investment in education and skills development to equip their workforce with the necessary digital skills. This will enable individuals to participate in the digital economy and reduce the risk of unemployment.

By embracing digital technologies, African countries can leapfrog traditional stages of development and accelerate economic progress.

2.2 Access to information and services

The 4IR has the potential to bridge the digital divide in Africa through the expansion of mobile connectivity, affordable smartphones, and Internet access. More Africans can benefit from digital services, including improved access to education, healthcare, financial services, and government information. A further aspect is that by leveraging digital technologies governments can improve service delivery and citizen engagement, leading to more inclusive and efficient governance (Bouzguenda et al. 2019; Sakolkar 2023). One of the most significant areas impacted by the expansion of access to digital tools is education. Online learning platforms and resources hold the potential to reach remote areas where traditional educational infrastructure is lacking. With access to digital educational content, Africans can acquire knowledge and skills to improve their livelihoods, bridge educational gaps, and pursue personal and professional growth. Digital financial services are another area where the 4IR has made significant strides in Africa. Mobile money platforms have gained popularity, allowing individuals to access banking services, make payments, and conduct financial transactions using their smartphones. This has facilitated financial inclusion, especially for the unbanked population, permitting them to participate in formal economic activities, save money securely, and access credit. Further, leveraging digital technologies can lead to more efficient and inclusive governance. Governments can use online portals and platforms to provide citizens with information, services, and opportunities for engagement. Online portals for accessing government information, applying for licences and permits, and participating in public consultations can enhance transparency, reduce corruption, and strengthen citizen–government interactions. These efforts can contribute to more inclusive and responsive governance systems.

2.3 Agriculture and food security

Agriculture is a vital sector in Africa, and the 4IR can revolutionise farming practices and enhance food security. Technologies like precision agriculture, remote sensing, and data analytics can improve crop yields, reduce waste, and optimise resource utilisation. Small-scale farmers can access market information, weather forecasts and financial services through mobile platforms, enabling them to make informed decisions and improve productivity. In addition, blockchain technology can enhance supply chain transparency, reducing food fraud and ensuring fairer prices for farmers. Precision agriculture leverages technologies like GPS, sensors and drones to monitor and analyse various factors affecting crop growth, such as soil conditions, moisture levels and nutrient requirements. By applying inputs precisely where and when they are needed, farmers can optimise resource use, reduce waste and increase crop yields. This approach helps minimise environmental impacts while maximising productivity.

Remote sensing technologies, including satellite imagery and aerial surveys, combined with data analytics, provide valuable insights into crop health, yield predictions and disease detection. Farmers can monitor their fields more effectively, identify potential issues at an early stage, and take proactive measures to prevent crop losses. Data analytics can also facilitate data-driven decision-making, allowing farmers to adopt optimal farming practices. Mobile platforms and digital solutions can bridge information gaps for small-scale farmers. By providing access to market information, weather forecasts and agronomic advice, farmers can make informed decisions regarding crop selection, the timing of planting, and selling their produce at the right time. Moreover, digital financial services such as mobile banking and microloans enable farmers to access credit, manage finances and invest in their agricultural activities more efficiently. Blockchain technology offers a transparent and secure platform for recording and verifying transactions. It can enhance supply chain transparency, traceability and accountability, particularly in agricultural value chains. By recording each transaction from farm to fork, blockchain enables consumers to verify the authenticity and origin of their food while ensuring fairer prices for farmers. This can help eliminate food fraud, improve market access and build trust among stakeholders. To effectively harness the potential of 4IR technologies in agriculture, it is crucial to invest in capacity-building programmes and knowledge-sharing platforms. Training farmers on the use of technology, data management and sustainable farming practices can empower them to adopt new approaches effectively. Collaborative initiatives involving farmers, researchers, governments and private sector actors can facilitate knowledge exchange and foster innovation in the agricultural sector.

2.4 Healthcare and access to quality services

The 4IR can address healthcare challenges in Africa by improving access to quality services and strengthening healthcare systems (Mbunge 2020; Mazibuko-Makena 2021). With the help of 4IR technologies, healthcare providers can remotely diagnose various conditions and provide appropriate treatment plans. This approach reduces the need for patients to travel long distances to receive specialised care, saving time and money. Remote consultations, virtual clinics, and remote monitoring of chronic conditions can significantly improve access to quality healthcare services. AI algorithms can analyse vast amounts of medical data, such as patient records, medical images, and genetic information, to assist healthcare professionals with diagnosing diseases accurately. AI-powered diagnostic tools can aid in early detection and timely intervention. Further, AI can be utilised in drug discovery processes, accelerating the development of new treatments, and improving healthcare outcomes. Telemedicine and mobile health applications can provide remote diagnosis, treatment, and health monitoring, particularly in underserved rural areas. One of the outstanding 4IR technologies which can help considerably with health is Artificial Intelligence. This technology can assist with disease diagnosis, drug discovery, and personalised medicine.

Moreover, the collection and analysis of health data can help identify disease patterns, support epidemiological research, and enable proactive public health interventions. In addition, telemedicine enables healthcare professionals to remotely diagnose, treat and monitor patients using communication technologies. This approach is particularly beneficial in underserved rural areas where access to healthcare facilities is limited. By leveraging mobile health applications, individuals can receive medical advice, access educational resources, schedule appointments, and receive reminders, thereby promoting proactive healthcare management.

2.5 Environmental sustainability

The 4IR presents opportunities for Africa to pursue sustainable development and mitigate the adverse effects of climate change. Smart grids, renewable energy technologies, and energy management systems can increase energy efficiency and promote clean energy adoption (Iris & Lam 2019; Iris & Lam 2021). Advanced data analytics can optimise resource usage and reduce waste in industries. Remote sensing and satellite imagery can also aid in environmental monitoring and conservation efforts, including wildlife preservation and deforestation prevention. The deployment of smart grids can revolutionise energy distribution and consumption by enabling the efficient integration of renewable energy sources, such as solar and wind power, into the existing power infrastructure. These grids can facilitate the bidirectional energy flow, empowering consumers to generate and sell excess electricity back to the grid. By incentivising the adoption of renewable energy technologies,

> "The 4IR brings powerful data analytics tools that can be harnessed to optimise resource usage, reduce waste, and inform evidence-based decision-making."

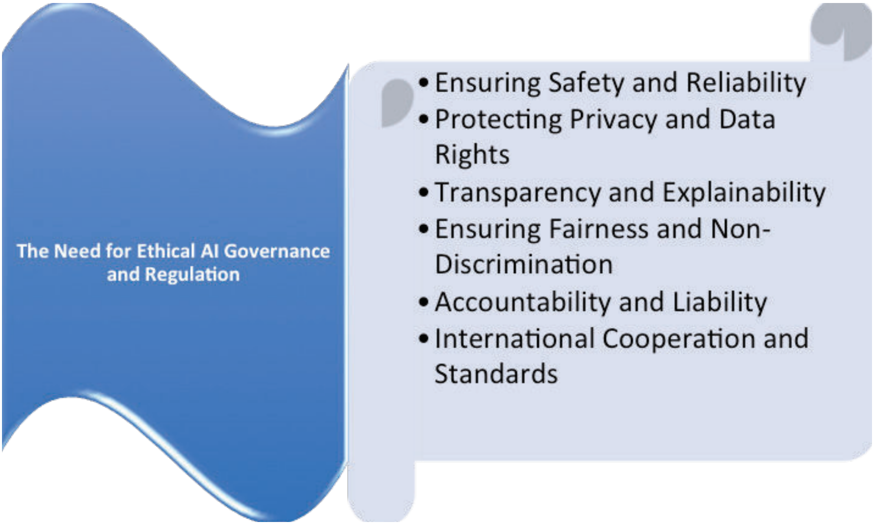
Africa can reduce its dependence on fossil fuels, decrease greenhouse gas emissions, and enhance energy security. Implementing energy management systems in industries and buildings can optimise energy usage, lower wastage, and improve overall energy efficiency. These systems employ advanced sensors, real-time data analysis, and automation to monitor and control energy consumption, identify inefficiencies, and suggest energy-saving measures. By optimising resource utilisation, businesses can significantly reduce their environmental footprint and achieve cost savings. The 4IR brings powerful data analytics tools that can be harnessed to optimise resource usage, reduce waste, and inform evidence-based decision-making. By analysing large datasets, organisations can gain insights into patterns, trends and inefficiencies, allowing them to identify areas where improvements can be made. For example, predictive analytics can help optimise agricultural practices, minimise water usage and optimise crop yields. Data-driven insights can also support sustainable urban planning, transportation systems, and waste management. Remote sensing technologies, satellite imagery, and unmanned aerial vehicles (UAVs) can play a vital role in monitoring and conserving the environment. These tools can be used to track deforestation, illegal logging, and encroachment on protected areas. They can further aid in wildlife preservation efforts, such as monitoring animal populations, tracking migration patterns and detecting poaching activities. By leveraging these technologies, Africa can strengthen its conservation strategies and protect its natural resources.

While the 4IR offers immense potential, Africa faces various challenges in taking advantage of its benefits. These challenges include limited infrastructure, such as reliable electricity and Internet connectivity, the digital skills gap, data privacy concerns, and cybersecurity risks. To address these challenges, African governments and stakeholders must prioritise investments in infrastructure development, education and skills training, policy frameworks, and cybersecurity measures. African policymakers and stakeholders can also gain valuable insights into best practices, potential pitfalls and regulatory frameworks that promote responsible and ethical AI adoption and other technologies in the Fourth Industrial Revolution Era.

3 The need for ethical AI governance and regulation

The rapid development and deployment of artificial intelligence (AI) technologies have raised concerns about their potential impact on society, privacy, and human rights (Donahoe & Metzger 2019; Leslie et al. 2021; Mhlanga 2023). The need for ethical AI governance and regulation has become increasingly important to ensure that AI systems are developed and used in a responsible and accountable manner (Mhlanga 2023). Figure 2 below summarises the need for ethical AI governance and regulation.

Figure 2: **The Need for Ethical AI Governance and Regulation**



The Need for Ethical AI Governance and Regulation

- Ensuring Safety and Reliability
- Protecting Privacy and Data Rights
- Transparency and Explainability
- Ensuring Fairness and Non-Discrimination
- Accountability and Liability
- International Cooperation and Standards

Source: Author’s analysis

3.1 Ensuring safety and reliability

AI systems, especially those operating in critical domains like autonomous vehicles or healthcare, must meet high standards of safety and reliability (O'Sullivan et al. 2019; Atakishiyev et al. 2021). Ethical AI governance and regulation should establish guidelines and standards for the testing, certification and ongoing monitoring of AI systems to ensure they are safe and reliable in their operation. Mechanisms should also be in place to address risks associated with malicious use or adversarial attacks on AI systems. To achieve this, several key measures can be implemented like testing and certification. AI systems should undergo rigorous testing and certification processes before being deployed in critical domains. This involves a comprehensive evaluation of a system's performance, including its ability to handle various scenarios and potential edge cases. Testing should cover a wide range of conditions, environments and inputs to assure robustness and reliability. Certification authorities can establish standards and benchmarks that must be met before an AI system is deemed safe and reliable for deployment. Once an AI system has been deployed, continuous monitoring is essential to identify any potential safety or reliability issues that may arise during its operation. Real-time data collection and analysis can help detect anomalies, errors, or performance degradation. Regular maintenance and updates should be performed to address any issues identified and make sure the system remains up-to-date and effective.

Another important aspect is to establish ethical AI governance frameworks and regulations, which are crucial to guiding the development and deployment of AI systems. These frameworks should include guidelines and principles that prioritise safety and reliability. They can address issues like transparency, explainability, accountability, and fairness in AI decision-making processes. Ethical AI governance frameworks should be developed collaboratively with input from experts, industry stakeholders, and the public. AI systems may be vulnerable to malicious attacks or adversarial manipulation. Safeguards should be implemented to protect AI systems from such threats. This involves incorporating security measures, such as robust encryption, authentication mechanisms, and intrusion detection systems. In addition, ongoing research and development efforts should focus on developing AI models that are more resilient to adversarial attacks. The other important aspect is to ensure the safety and reliability of AI systems, where it is essential to establish strict data privacy and protection measures. AI systems often rely on sensitive data, such as patient records or personal information, and it is vital to handle and store this data securely. Implementing strong data encryption, access controls, and privacy-enhancing technologies can help mitigate the risks associated with data breaches and unauthorised access. Lastly, collaboration and knowledge sharing are critical. Collaboration between industry, academia and regulatory bodies is essential for advancing the safety and reliability of AI systems. Sharing best practices, research findings, and lessons learned can accelerate progress in this field. Establishing platforms for knowledge exchange and fostering open dialogue among stakeholders can facilitate the development and implementation of effective safety and reliability measures.

3.2 Protecting privacy and data rights

AI often relies on vast amounts of personal data to train and operate effectively. Ethical AI governance and regulation should prioritise the protection of privacy and data rights. This includes implementing robust data protection frameworks, obtaining informed consent for data collection and usage, and ensuring that AI systems are designed with privacy in mind (Mhlanga 2022). Clear guidelines should be established for the responsible handling of personal data, including anonymisation and secure storage practices. Protecting privacy and data rights in the context of AI is essential to address concerns related to the collection, storage and use of personal data. Several key considerations and measures can be taken to prioritise privacy and data rights. One of these is robust data protection frameworks. Ethical AI governance and regulation should establish comprehensive data protection frameworks aligned with existing privacy laws and regulations. These frameworks should outline the rights and protections afforded to individuals concerning their data. They can include principles such as purpose limitation (data should be collected for specific and legitimate purposes), data minimisation (collecting only the necessary data), and data retention limits. Another important aspect is informed consent. Obtaining informed consent is crucial while collecting and using personal data for AI systems. Individuals should be informed about the types of data that will be collected, how it will be used, and any potential risks or implications. Consent mechanisms should be transparent, easily understandable, and provide individuals with the option to withdraw their consent at any time. In cases where consent cannot be obtained, such as with anonymised and aggregated data sets, alternative approaches should be explored to ensure privacy and data protection.

AI systems should be designed with privacy in mind from the early stages of their development. Privacy by design principles can guide the implementation of privacy safeguards into the architecture and functionality of AI systems. This includes incorporating privacy-enhancing technologies, such as differential privacy or federated learning, to protect sensitive data and minimise the risks of re-identification and this should be followed by responsible data handling. Clear guidelines and standards should be established for the responsible handling of personal data in AI systems. This includes assuring that data is anonymised or de-identified whenever possible to protect individual privacy. Anonymisation techniques, such as removing direct identifiers and applying data aggregation or perturbation methods, should be employed to minimise the risk of re-identification. Secure storage practices, encryption, and access controls should also be implemented to safeguard personal data from unauthorised access or breaches.

Third-party data sharing is another aspect that needs attention. When AI systems rely on third-party data sources, it is crucial to establish data-sharing agreements that prioritise privacy and data rights. These agreements should define the purpose and scope of the data sharing, specify data protection measures, and ensure compliance with relevant privacy regulations. Clear guidelines and mechanisms should be in place to assess the privacy practices of third-party data providers and make sure they meet the required standards. Ongoing monitoring, auditing and accountability mechanisms should be established to ensure compliance with privacy and data protection requirements. This could involve regular assessments of data handling practices, privacy impact assessments, and audits of AI systems' data usage and storage. Organisations should be accountable for any breaches or mishandling of personal data and take appropriate corrective actions. By prioritising the protection of privacy and data rights in AI systems, we can foster trust among users and mitigate the potential risks associated with the collection and usage of personal data. This approach warrants that AI technology is deployed responsibly and ethically, respecting individuals' privacy and promoting data protection.

3.3 Transparency and explainability

The black-box nature of some AI algorithms and models leads to concerns about accountability and decision-making processes. Ethical AI governance and regulation should encourage transparency and explainability in AI systems. This involves providing clear documentation of how AI systems make decisions, enabling individuals to understand the reasoning behind AI-generated outcomes. The ability to explain AI decisions is particularly critical in domains like healthcare, finance and criminal justice. Transparency and explainability in AI systems are essential to address concerns surrounding accountability, fairness and trust. AI developers should provide comprehensive documentation that explains the design, architecture and functioning of AI models. This documentation should include details about the data used for training, the preprocessing steps applied, the specific algorithms or techniques employed, and any biases or limitations associated with the model. This allows stakeholders, including users, regulators and auditors, to gain insights into how the AI system operates. Encouraging the use of interpretable AI models can significantly enhance transparency and explainability. Models such as decision trees, linear models, or rule-based systems provide clear rules or explanations for their predictions or decisions. While more complex models like deep neural networks may lack inherent interpretability, techniques like feature importance analysis, attention mechanisms, or layer-wise relevance propagation can provide insights into the model's decision-making process.

The other issue is that AI systems should be designed to generate explanations or justifications for their outputs or decisions. These explanations can take the form of textual or visual descriptions that outline the key factors, features or evidence that contributed to a specific outcome. By providing explanations, users and stakeholders can better understand the reasoning behind AI-generated decisions, with the decision-making process thereby becoming more transparent and accountable. Independent auditing of AI systems can be conducted to evaluate their transparency and explainability. Auditors can review the documentation, assess the model's behaviour, and verify its compliance with transparency standards. Algorithmic auditing can help identify any biases, unfair practices, or lack of explainability in AI systems, providing valuable insights for improvement.

User-friendly interfaces are also important. AI systems should incorporate user-friendly interfaces that allow individuals to interact with and understand the system's outputs. Visualisations, interactive dashboards or summary explanations can facilitate users' comprehension of AI-generated results, empowering them to question or challenge outcomes when necessary. Ethical AI governance and regulation should establish guidelines and standards that explicitly require transparency and explainability in AI systems, especially in critical domains like healthcare, finance and criminal justice. Regulators can mandate documentation, auditing or explainability requirements to assure that AI systems are accountable and adhere to acceptable standards of transparency. Promoting education and awareness initiatives can enhance understanding and appreciation for transparency and explainability in AI. This includes educating AI developers, users and decision-makers about the importance of transparency, the challenges involved, and the techniques available to achieve explainability. By fostering a culture that values transparency and accountability, stakeholders can actively advocate for more transparent AI systems. Prioritising transparency and explainability in AI systems can foster trust, enable better decision-making, and mitigate the risks of biased or unfair outcomes. It empowers individuals to understand and challenge AI-generated decisions, contributing to the more responsible and accountable deployment of AI technology.

3.4 Ensuring fairness and non-discrimination

AI systems can inadvertently perpetuate the biases and discrimination present in the data they are trained on. For example, facial recognition systems have been shown to have higher error rates for women and people with darker skin tones. Ethical AI governance and regulation should promote fairness and non-discrimination by making sure that AI systems are trained on diverse and representative datasets and by regularly auditing and monitoring their performance to identify and mitigate biases. Ensuring fairness and non-discrimination in AI systems is crucial to avoid perpetuating biases and to uphold ethical standards. There are key considerations and measures to prioritise fairness and non-discrimination. One aspect involves diverse and representative datasets. AI systems should be trained on broad and representative datasets that accurately reflect the diversity of the population they are intended to serve. Care should be taken to include data from various demographic groups, ensuring adequate representation of different genders, races, ethnicities, ages and socio-economic backgrounds. By incorporating diverse data, AI models are more likely to be fair and unbiased in their decision-making processes.

Bias detection and mitigation are also important. Regular audits and monitoring of AI systems should be conducted to detect and mitigate biases. This involves analysing a system's outputs and evaluating the impact on different demographic groups. Bias detection methods such as statistical analysis or fairness metrics can help identify disparities and potential sources of bias. If biases are identified, steps should be taken to understand their root causes and address them through model retraining, data augmentation, or algorithmic adjustments. Explainability in AI systems plays a crucial role in ensuring fairness. By providing explanations for decisions and predictions, individuals can better understand how the system arrived at a particular outcome. This transparency allows for the identification and examination of any discriminatory patterns or biases. Explanations can also help individuals affected by AI decisions to seek recourse or challenge potentially unfair outcomes.

Regular evaluation and auditing are also important. Periodic evaluation and auditing of AI systems should be conducted to assess their fairness

> "Ensuring fairness and non-discrimination in AI systems is crucial to avoid perpetuating biases and to uphold ethical standards."

and non-discriminatory performance. Independent audits can help identify potential biases and evaluate whether the system adheres to established fairness standards. These evaluations should involve experts with varied backgrounds to provide objective assessments and recommendations for improvement. Stakeholder Engagement including diverse perspectives and voices of stakeholders in the development and evaluation of AI systems is essential to address potential biases and ensure fairness. Collaboration with domain experts, community representatives and impacted individuals can provide valuable insights and contribute to the development of more inclusive and unbiased AI systems.

Regulatory guidelines are critical. Ethical AI governance and regulation should establish clear guidelines and standards for promoting fairness and non-discrimination in AI systems. These guidelines should specify the requirements for data collection, model training, auditing and evaluation processes to assure fairness. Regulators can also enforce transparency in the reporting and disclosure of AI systems' performance, including any identified biases or discriminatory impacts. AI systems should be designed with the capacity for continuous learning and improvement. As new biases are identified and societal understanding evolves, AI models and algorithms should be updated and retrained to address these concerns. Regular feedback loops, user input and ongoing research efforts can add to the iterative improvement of AI systems' fairness and non-discriminatory performance.

3.5 Accountability and liability

Determining accountability and liability in cases where AI systems cause harm or make incorrect decisions is a complex issue. Ethical AI governance and regulation should provide clarity on the allocation of responsibility and liability between AI developers, operators and users. Legal frameworks should be updated to address emerging challenges and ensure that individuals affected by AI systems have avenues for seeking recourse and remediation. Expanding accountability and liability in the context of AI systems is indeed a crucial aspect of ethical AI governance and regulation. As AI technology becomes increasingly integrated into various aspects of society, it is important to establish clear guidelines on determining responsibility and liability when AI systems cause harm or make incorrect decisions. There are some key considerations and actions that can contribute to addressing this complex issue. The clear allocation of responsibility is crucial. Ethical AI governance frameworks should clearly define the roles and responsibilities of different stakeholders involved in the development, deployment and use of AI systems. This includes AI developers, operators and users. Clarifying who is accountable for different aspects of AI systems will make it easier to determine liability in the case of harm or incorrect decisions.

Another important aspect is that AI systems should be designed and developed in a way that promotes transparency and explainability.

More importantly, updated legal frameworks are critical. Existing legal frameworks should be revised and updated to account for the unique challenges posed by AI technology. This may involve creating new legislation or adapting existing laws to address the specific issues related to AI. Legal frameworks should consider aspects such as data privacy, security, transparency, explainability and fairness when determining liability and providing avenues for seeking recourse. AI developers and operators should conduct thorough risk assessments during the entire lifecycle of AI systems. This includes identifying potential harms and risks associated with the system and implementing measures to mitigate those risks. By taking a proactive approach to risk assessment, developers and operators can demonstrate their commitment to accountability and minimise the chances of harm occurring.

Another important aspect is that AI systems should be designed and developed in a way that promotes transparency and explainability. Users and affected individuals should have access to understandable explanations of how AI systems make decisions or cause harm. This transparency helps with determining liability and fosters trust between stakeholders. Individuals affected by AI systems should have avenues for seeking recourse and remediation in the event of harm. This may involve establishing complaint mechanisms, dispute resolution processes, or even compensation frameworks to address the consequences of AI-related harm. These mechanisms should be easily accessible, fair and efficient.

Again, international collaboration is another crucial aspect. Given the global nature of AI technology, collaboration between different countries and jurisdictions is essential. International cooperation can help in harmonising legal frameworks, sharing best practices, and addressing challenges related to accountability and liability on a broader scale. Continuous monitoring and evaluation are also critical. Ethical AI governance frameworks should include provisions for the continuous monitoring and evaluation of AI systems. This allows for the identification of potential harm or incorrect decisions and enables timely action to rectify issues. Regular audits and assessments of AI systems can also contribute to accountability and liability determination. By addressing these considerations and implementing appropriate measures, society can work towards expanding accountability and liability in the realm of AI systems. This ensures that the benefits of AI are maximised while minimising potential harm and providing individuals with avenues for recourse and remediation when needed.

4 International cooperation and standards

The global nature of AI development and deployment means that ethical AI governance and regulation should involve international cooperation and the establishment of common standards. Collaboration between governments, industry stakeholders, academia and civil society organisations is required to develop shared principles, guidelines and best practices for ethical AI. International forums and organisations can play a valuable role in facilitating dialogue and coordination on AI governance and regulation. International cooperation and the establishment of common standards are indeed crucial for ethical AI governance and regulation. Given the global nature of AI development and deployment, it is essential to foster collaboration between different stakeholders to ensure the responsible and ethical use of AI. There are many key aspects and benefits of international cooperation and standards in the context of AI, including shared principles and values. International cooperation enables the development of shared principles and values that can guide the ethical use of AI across different countries and jurisdictions. By bringing together diverse perspectives, governments, industry stakeholders, academia and civil society organisations can work towards consensus on fundamental principles, such as fairness, transparency, accountability and human rights, which should underpin AI systems globally.

It is also important to ensure that legal frameworks are harmonised. Collaboration between countries can help harmonise legal frameworks related to AI governance and regulation. Aligning laws and regulations will make it easier to address the challenges posed by AI on an international scale. Harmonisation can facilitate smoother cross-border collaboration, assure the consistent protection of individual rights, and prevent regulatory fragmentation that could hinder innovation and the responsible use of AI. International cooperation allows for the development of consistent standards and guidelines for ethical AI. These standards can cover various aspects of AI development and deployment, including data privacy, security, bias mitigation, explainability and accountability. Common guidelines provide a unified framework for AI practitioners and organisations to follow, fostering responsible AI practices worldwide. Collaborative efforts enable the sharing of knowledge, best practices and lessons learned from different countries and organisations. By sharing experiences and expertise, stakeholders can learn from one another and build their capacity to address emerging AI challenges effectively. This knowledge sharing can occur through international forums, conferences, workshops, and collaborative research initiatives.

Ethical dilemmas in AI often transcend national boundaries. Issues such as algorithmic bias, autonomous weapons, privacy concerns and the impact of AI on employment require international cooperation to develop comprehensive and globally applicable solutions. By engaging in international dialogue and collaboration, stakeholders can work together to tackle these complex ethical challenges collectively. International cooperation enables policy coordination on AI governance and regulation. Forums and organisations on the global level can facilitate discussions and provide platforms for policymakers to exchange ideas and align their approaches. This coordination helps with avoiding potential conflicts and promoting a cohesive global response to the ethical challenges brought by AI. International collaboration contributes to building trust among countries and stakeholders. Through open and inclusive dialogue, diverse perspectives can be heard and considered while formulating AI policies and standards. This trust-building process fosters a sense of shared responsibility and ownership, strengthening the effectiveness and acceptance of international AI governance initiatives. Promoting international cooperation and standards in AI governance and regulation is essential for addressing the global impact of AI and assuring its responsible and ethical development and deployment. By working together, stakeholders can establish a common understanding of ethical AI principles, share best practices, and create a harmonised framework that promotes the positive and inclusive use of AI technology worldwide.

5 Ethical considerations in AI research and development

Ethical AI governance and regulation should encourage responsible research and development practices. This includes fostering a culture of ethics and accountability among AI researchers and developers. Ethical considerations, such as human rights, social impact and the potential consequences of AI deployment, should be integrated into the entire AI lifecycle, from the design phase to deployment and ongoing monitoring. The need for ethical AI governance and regulation is paramount in addressing the challenges and potential risks associated with AI technologies. By establishing clear guidelines and frameworks, society can harness the benefits of AI while minimising its negative impacts. Ethical AI governance and regulation should prioritise fairness, non-discrimination, privacy protection, transparency, safety, accountability, international cooperation, and ethical considerations throughout the AI lifecycle. This will ensure that AI technologies are developed and used in a manner that is line with societal values, human rights, and the best interests of individuals and communities. To further expand on ethical considerations in AI research and development, below we delve into some key aspects:

5.1 AI governance and regulation in the European Union (EU)

The EU has been at the forefront of developing AI governance and regulation, having recognised the need to address the ethical concerns associated with this transformative technology. Several key ethical challenges have emerged in the EU's AI landscape, including bias and discrimination. AI systems are vulnerable to inheriting and perpetuating any biases present in the data used to train them, leading to discriminatory outcomes. The EU has been actively working on ensuring that AI systems are transparent, fair and unbiased, aiming to prevent discrimination based on gender, race or other protected characteristics. AI systems can indeed inherit and perpetuate any biases present in the data used to train them, which can result in discriminatory outcomes. This issue has gained significant attention in recent years, and efforts have been made to address it. The EU has been leading the way in developing regulations and guidelines to ensure transparency, fairness and unbiased AI systems.

5.2 The Artificial Intelligence Act in the EU

In April 2021, the European Commission proposed the Artificial Intelligence Act aimed at establishing a comprehensive framework for AI regulation within the EU. It states that AI systems which can be used in different applications are analysed and classified according to the risk they pose to users. The different risk levels will mean regulation (European Parliament 2023). A key objective of the proposed legislation is to prevent discrimination based on gender, race or other protected characteristics. The Act emphasises that AI systems should be developed and used in a manner that respects fundamental rights and prohibits any form of discrimination. To ensure fairness and transparency, the Act includes provisions such as data quality and bias assessment requirements. AI systems must be trained on representative and diverse datasets, and developers must conduct regular risk assessments to identify and mitigate any biases in the data or algorithms. The proposal encompasses four primary objectives that form the foundation of the EU's vision for AI governance.

5.3 First pillar: Upholding rights and safety

The AI Act is a strong framework that was created to make sure that AI systems, whether they are employed, deployed or marketed within the Union, abide by strict safety criteria and follow the established legal framework on fundamental rights and Union values. This pillar provides legal certainty, which is anticipated to encourage investment and innovation in AI by clearly defining the specifications for AI systems as well as the responsibilities of all parties involved in the value chain. It additionally strengthens the governance and enforcement of current laws about fundamental rights and safety standards that apply to AI systems by giving relevant authorities new authority, resources and unambiguous guidelines for compliance evaluation. To avoid internal market fragmentation brought on by possibly incompatible state frameworks, this pillar is essential for the EU's digital single market policy. It is intended to assure fairness, protect everyone, and strengthen Europe's industrial foundation and competitive advantage in AI.

5.4 Second pillar: Safeguarding trustworthy AI

The EU AI Act's second pillar introduces a proportionate risk-based approach to policing the creation, promotion and application of AI systems across the Union. By coordinating regulations that apply to AI systems depending on their potential hazards, this pillar aims to strike a balance between innovation and protection. It imposes tight limitations and safety measures on the use of remote biometric identification systems in law enforcement while outlawing some AI practices regarded as damaging and inconsistent with Union values. The risk-based strategy described in this pillar tries to make sure that high-risk AI systems, which pose serious dangers to people's health, safety, or fundamental rights, adhere to the laws requiring reliable AI. Before being approved for sale in the Union market, these systems must pass stringent conformity evaluation procedures. The EU AI Act uses a risk-based approach to encourage responsible innovation while protecting people and supporting Union values in the rapidly changing field of AI.

5.5 Third pillar: Enforcing AI regulations across the EU

The EU AI Act's third pillar, which addresses governance and enforcement, aims to improve the application and enforcement of current rules about basic rights and safety standards that apply to AI systems. Creating the European Artificial Intelligence Board, this pillar introduced a vehicle for cooperation on the Union level while establishing a governance system on the level of the member states. By utilising the knowledge and resources of member states, the governance structure assures the uniform national enforcement of the AI Act. The European Artificial Intelligence Board simultaneously promotes collaboration, harmonises procedures, and guarantees uniform regulatory enforcement across the Union. The EU AI Act intends to increase accountability and ensure that AI systems deployed within the Union function within legal bounds, protecting people's rights and promoting public trust in AI technologies. It does so by building strong governance and enforcement procedures.

5.6 Fourth pillar: Building a Single Market for AI

The EU AI Act's fourth pillar is to facilitate the growth of a unified market for ethical, secure and reliable AI applications while averting market fragmentation. This pillar seeks to foster an atmosphere that promotes creativity, investment and the use of AI technologies. It highlights the significance of maintaining fair competition, reducing entry obstacles, and creating level playing fields for AI systems within the Union. The Act encourages the unhindered circulation of AI applications across member states and the harmonisation of regulations and standards for AI systems. Further, it encourages cooperation and information sharing among pertinent authorities to address the issues raised by AI systems with cross-border ramifications, especially those utilised in public administrations. The fourth pillar of the EU AI Act aims to unlock the full potential held by AI technologies while protecting individual rights and enhancing Europe's digital competitiveness by allowing the growth of the single market for AI.

6 The EU's General Data Protection Regulation (GDPR)

GDPR stands for General Data Protection Regulation. All EU member states are subject to the GDPR, which went into effect in 2018. It provides a comprehensive data protection framework and includes requirements about AI systems while establishing tight guidelines for the gathering, processing and storage of personal data. The GDPR places a strong emphasis on the defence of individual privacy rights and holds businesses responsible for upholding data security while utilising AI technologies. AI frequently uses enormous volumes of personal data, which raises questions about privacy and data protection. The GDPR has made tremendous progress in defending people's rights to their privacy and fostering responsibility in the handling of personal data. The regulation outlines the broad responsibilities of data controllers and the "processors" that handle personal data on their behalf. These obligations include the need to put in place the proper security safeguards reflecting the level of risk associated with the data processing operations they carry out.

The ethical issues relating to AI governance and legislation have received considerable attention from the EU.

The EU emphasises the significance of creating AI that is intended to enhance human skills rather than replace or negatively impact people. Human rights, dignity and well-being should be prioritised by AI systems while also assuring their accountability, transparency and explainability. The GDPR is comprehensive legislation that applies to the processing of personal data, including data processed by AI systems, although not expressly targeted at AI. It creates data protection guidelines, including those that guarantee data processing is transparent, equitable and lawful. Individuals' rights and privacy must be protected by GDPR standards for AI systems handling personal data. Ethical Principles for Reliable AI - Guidelines for the ethical development and use of AI have been released by the High-Level Expert Group on AI of the European Commission. These standards offer a framework for ensuring that AI upholds fundamental rights, privacy and societal values for developers, users and regulators. They stress how crucial it is for AI systems to have human agency, responsibility and explainability. For faulty products, including AI systems, the Product Liability Directive outlines liability guidelines. The Directive enables persons to seek compensation from the manufacturer or other parties in the supply chain if an AI system caused harm or damage because of a flaw. This law makes sure that safety and accountability are given top priority during the development and implementation of AI systems. These laws and rules demonstrate the EU's dedication to promoting the creation of AI that is centred on people while preserving their rights, dignity and well-being. The desire is to establish a legal framework that addresses possible hazards related to AI technologies and encourages responsibility, transparency and explainability.

6.1 Transparency and explainability

AI systems need to be transparent, which means that people should be aware when they are engaging with AI and comprehend the implications of those interactions. Explainable artificial intelligence is something that can offer human-level explanations for its decisions and actions, and the EU is working to encourage its development.

6.2 High-risk AI regulation

A legislative framework that would explicitly target high-risk AI systems has been proposed by the EU in the form of the AI Act. Certain applications of artificial intelligence, like those used in essential infrastructure, transportation and healthcare, would be subject to stringent standards if this act were passed.

Conformity assessments, which would include testing and documentation, would be performed on AI systems that posed a high level of risk to guarantee that the systems comply with safety, transparency and accountability criteria.

7 European Commission's ethics guidelines for trustworthy AI

The European Commission released guidelines for creating and using reliable AI in 2019. These recommendations centre on seven fundamental ideas: human agency and oversight, technical robustness and safety, privacy and data governance, openness, diversity, and non-discrimination; as well as societal and environmental well-being and accountability. By adhering to these guidelines, AI will be compliant with moral standards and uphold basic rights. EGE stands for the European Group on Ethics in Science and New Technologies. The European Commission President's independent advisory council is called the EGE, which was created 1991. Commission Decision (2021/156) establishes the group's legal mandate. The College of Commissioners as a whole and the President of the European Commission receive reports from the EGE which offers unbiased guidance to the European Commission on moral matters including science, technology and AI. It has published findings and opinions on AI ethics, emphasising the need for openness, justice and regard for human dignity. The EU develops its AI governance and regulation policies considering the EGE's suggestions. The European Commission has launched an initiative called AI Watch to track the growth, adoption and effects of AI across the EU. It attempts to offer information, analysis and viewpoints on AI-related subjects, including moral issues. AI Watch supports evidence-based policymaking and makes sure that ethical concerns are addressed in AI governance by monitoring the advancement and difficulties of AI deployment.

The ethical issues relating to AI governance and legislation have received considerable attention from the EU.

With an emphasis on the defence of individual rights, openness, justice and accountability, these instances show the EU's dedication to resolving ethical concerns in AI governance and legislation.

7.1 AI Infrastructure and Governance Capacities in Africa

Many examples can be observed in Africa wherein AI infrastructure is progressively being implemented across the continent. This section will provide several illustrations of the deployment of artificial intelligence (AI) and culminate with an examination of the governance frameworks that are presently being instituted. Zipline, a drone delivery business, using AI technology to enable unmanned aerial vehicles (UAVs) to independently navigate and transport vital medical resources to isolated regions of Rwanda. The enhanced accessibility to healthcare services resulting from this development has had a notable impact on individuals residing in rural regions, who were previously required to undertake extensive journeys to acquire vital prescriptions and immunizations. In the context of Kenya, artificial intelligence (AI) is being employed to facilitate the creation of diagnostic tools and forecast disease epidemics. As an illustration, AI-enabled algorithms are now employed in the analysis of medical pictures, including X-rays and CT scans, to aid in the diagnosis of diseases such as malaria and tuberculosis. Furthermore, artificial intelligence (AI) models are currently being employed to evaluate data about disease transmission patterns.

The enhanced accessibility to healthcare services resulting from this development has had a notable impact on individuals residing in rural regions

This enables the prediction of potential outbreak occurrences, along with the identification of their probable locations. Consequently, early intervention and prevention measures can be implemented to mitigate the impact of such outbreaks. Several firms are currently utilizing artificial intelligence (AI) in their operations. One such example is AfyaDoc, a startup based in Kenya. AfyaDoc has successfully created an application that employs AI technology to diagnose malaria by analyzing photographs of the eye. The application employs machine learning techniques to examine the ocular blood vessels, hence enabling the detection of indications of malaria infection. The AfyaDoc platform is presently undergoing pilot testing in several hospitals and clinics across Kenya. Another company, known as m-health, is also involved in the field. Kenya is a domestic enterprise in Kenya that is leveraging artificial intelligence (AI) to construct a predictive framework for anticipating disease epidemics. The method utilizes data about weather patterns, climatic conditions, and population density to discern geographical regions that exhibit a higher propensity for the occurrence of disease outbreaks. Subsequently, the system possesses the capability to transmit notifications to healthcare professionals situated inside these regions, enabling them to proactively undertake precautionary actions.

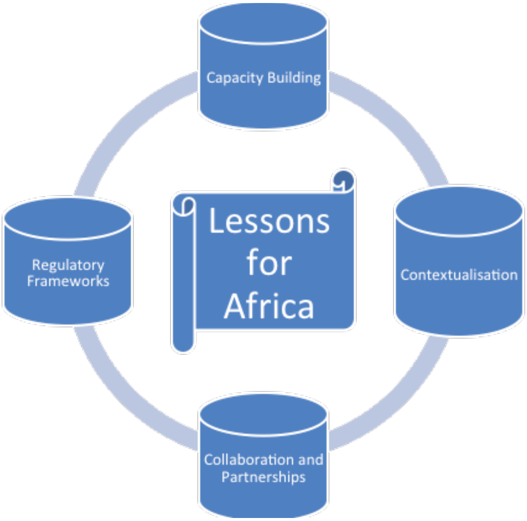
7.2 Governance Initiatives in Africa

South Africa harbours numerous AI research institutions and companies, with the government demonstrating a proactive stance in the formulation and implementation of AI policies. The South African Artificial Intelligence Institute (SAAI) was established by the government in 2020 to foster the responsible development and utilization of artificial intelligence (SAAIA 2023). The SAAI is now engaged in multiple endeavours, which encompass the formulation of ethical principles for the appropriate use of AI, provision of assistance for AI research and development, and facilitation of enhanced comprehension of AI among the general public. Rwanda has emerged as a prominent figure in the realm of AI policy development. The government has established collaborative partnerships with prominent international organizations, like the World Bank and the African Development Bank, to formulate comprehensive plans and policies for AI development. The Rwanda Artificial Intelligence for Development (RAFID) initiative was introduced by the government in 2018 to utilize AI to address significant issues in the country, including poverty and healthcare accessibility (Benoit, 2022). Kenya has demonstrated notable advancements in the establishment of a comprehensive legal framework for AI development. The implementation of the Data Protection Act in 2019 within the country is a significant stride towards establishing a favourable setting for the advancement of artificial intelligence. The legislation establishes fundamental guidelines about the acquisition, retention, and use of individuals' personal information, which are imperative in the advancement of artificial intelligence.

8 Lessons held by the fourth industrial revolution for Africa

Africa, with its diverse social, economic and technological landscape, can draw valuable lessons from the EU's experiences with AI governance and regulation. These lessons are outlined in Figure 3 below.

Figure 3: **Lessons held by the Fourth Industrial Revolution for Africa**



Source: Author's analysis

Contextualization: Africa should develop an AI governance and regulation framework that reflects its unique needs, challenges and cultural context. Localising AI ethics principles and policies will ensure that the benefits of AI are harnessed in a way that aligns with African values and priorities. Collaboration and Partnerships: The EU's approach to AI governance involves multi-stakeholder collaborations among policymakers, industry experts, civil society organisations, and academia. Africa could replicate this collaborative model by engaging various stakeholders to foster dialogue, knowledge sharing, and the co-creation of ethical AI frameworks. However, implementing this model in Africa faces certain obstacles. These include limitations in resources, which can affect the development and deployment of AI technologies and policies. Political will is also a crucial factor, as effective collaboration requires a strong commitment from government leaders. Additionally, the digital divide disparities in access to technology and digital literacy can hinder widespread participation in these collaborative efforts. To overcome these challenges, African countries need to focus on building infrastructure, enhancing digital literacy, and fostering a political environment conducive to technological innovation and ethical governance. This approach will help in realizing the full potential of AI for the continent's development.

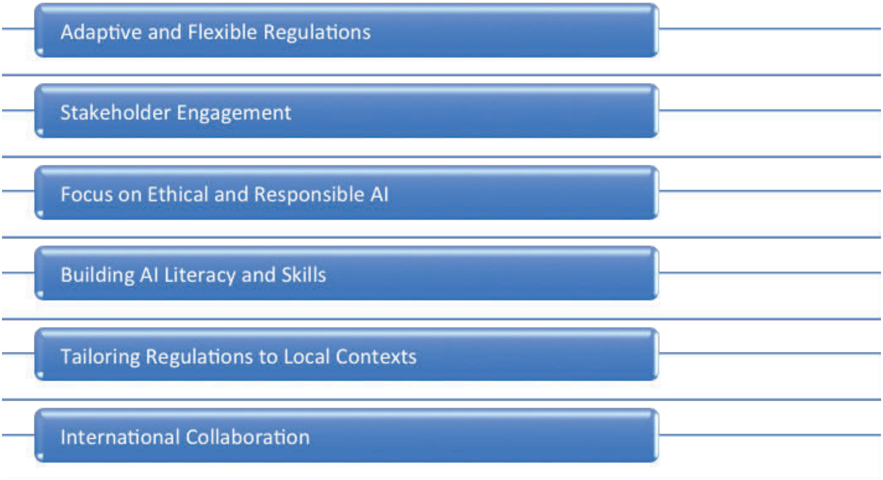
Capacity Building: Investing in capacity building and technical skills development is crucial for Africa to effectively govern and regulate AI.

Strengthening educational programmes, research initiatives, and public awareness campaigns could empower African nations to navigate the ethical challenges of AI and make informed policy decisions. Regulatory Frameworks: Africa can learn from the EU's regulatory frameworks, such as the GDPR, to develop comprehensive legislation that protects individuals' privacy rights, promotes transparency and ensures accountability in the use of AI. Customising these frameworks to suit Africa's unique needs will be essential.

Balancing The Need For Regulation With The Promotion Of Innovation

Balancing the need for regulation with the promotion of innovation in AI technologies in African nations involves a comprehensive approach. This balance is crucial to ensure that AI develops in a way that is both beneficial and responsible. Here's how this can be achieved as outlined in Figure 4 below.

Figure 4: **Balancing the Need For Regulation and the Promotion Of Innovation**


Adaptive and Flexible Regulations
Stakeholder Engagement
Focus on Ethical and Responsible AI
Building AI Literacy and Skills
Tailoring Regulations to Local Contexts
International Collaboration

Source: Author's Analysis

As depicted in Figure 4, the process of achieving a balance between the necessity for regulation and the promotion of innovation in Africa necessitates the implementation of specific measures, such as the development of adaptive and flexible regulatory frameworks. The implementation of regulatory frameworks that possess adaptability to the swift speed of technical advancements in the field of artificial intelligence. Regulations must possess a certain degree of flexibility to adapt to emerging advancements, thereby circumventing the formulation of too rigid guidelines that may swiftly become obsolete. Promoting a “sandbox” methodology that facilitates AI developers to engage in controlled experimentation and innovation, while operating inside a relaxed regulatory framework, thereby aiding in the identification of effective regulatory solutions.

Another crucial measure is to enhance stakeholder engagement. The regulation approach should incorporate a diverse array of stakeholders, encompassing AI developers, corporations, academia, and civil society. The adoption of this inclusive approach guarantees that regulations are firmly rooted in actual realities and encompass a wide range of opinions. Frequently engaging in consultations with the technology industry and innovators to gain comprehensive insight into their specific issues and requirements, ensuring that regulatory measures are conducive rather than inhibitory. Another crucial aspect to consider is the emphasis on ethical and responsible artificial intelligence (AI). The regulatory framework should prioritize the ethical utilization of artificial intelligence (AI), with particular attention to concerns surrounding privacy, data protection, and justice. The establishment of a foundation of trust and safety surrounding artificial intelligence (AI) technologies is of paramount importance in ensuring their acceptance and advancement. The implementation of principles and standards for responsible artificial intelligence (AI) is crucial in promoting openness and responsibility while avoiding excessive limitations. Furthermore, it is crucial to cultivate AI literacy and develop the necessary skills. The allocation of resources towards educational and training initiatives aimed at cultivating local proficiency in the field of artificial intelligence (AI). This facilitates the development of a well-informed workforce

The adoption of this inclusive approach guarantees that regulations are firmly rooted in actual realities and encompass a wide range of opinions.

capable of generating innovative solutions within the confines of established regulatory frameworks. The objective is to enhance public comprehension of artificial intelligence (AI) to cultivate a conducive atmosphere for the advancement and use of this technology.

Adapting regulations to suit specific local contexts is of paramount importance. The formulation of legislation that is customized to address the distinct requirements, obstacles, and preferences of African nations. This entails not simply imitating Western ideas, but rather developing techniques that are tailored to the specific circumstances. When examining the ramifications of artificial intelligence (AI) on regional economies, cultures, and civilizations within the context of regulatory frameworks... Another significant issue to consider is international collaboration. Engaging in collaborative efforts with other nations and international entities to get insights from globally recognized models of artificial intelligence (AI) regulation. The act of participating in discussions about international standards and norms is crucial to guarantee the compatibility and competitiveness of African artificial intelligence (AI) technology within the global arena. Through the implementation of these measures, African nations have the potential to establish a regulatory framework that effectively protects society and cultivates an ecosystem conducive to the flourishing of artificial intelligence (AI) innovation. The adoption of a balanced approach is crucial in harnessing the promise of artificial intelligence (AI) as a tool for promoting economic and societal progress, while simultaneously addressing and minimizing associated hazards.

In conclusion, the ethical challenges arising in AI governance and regulation in the EU are multi-faceted and complex, as shown in the discussion above. These challenges encompass issues such as bias and discrimination, privacy

Another crucial aspect to consider is the emphasis on ethical and responsible artificial intelligence (AI).

and data protection, accountability and transparency, and safety and security. The EU has been at the forefront of addressing these challenges, having recognised the need to develop comprehensive frameworks that mitigate the risks associated with AI while maximising its benefits. For example, the EU has been actively working to make sure that AI systems are transparent, fair and unbiased, aiming to prevent discriminatory outcomes based on gender, race or other protected characteristics. The EU's General Data Protection Regulation (GDPR) has also played a pivotal role in safeguarding individuals' privacy rights and promoting accountability in the use of personal data in the context of AI. By examining the EU's experiences, Africa can gain valuable insights into the key ethical challenges and considerations that must be addressed in the adoption and regulation of AI. However, it is important to acknowledge that Africa's path to AI governance and regulation will differ from that of the EU. Africa's unique socio-cultural context, diverse economies, and varying levels of technological infrastructure call for a nuanced approach. Although lessons from the EU can serve as a foundation, African nations must tailor their strategies to align with their own particular needs, aspirations and values.

9 Conclusion

Africa's emergence as a potential centre for the Fourth Industrial Revolution and AI technology comes with a considerable responsibility to address the associated ethical considerations. While recognising the advantages of AI in driving economic expansion and innovation, it is crucial to approach its integration carefully, considering the potential hazards and ethical implications. Learning from the European Union's experience, African nations can establish resilient frameworks for AI governance and regulation that prioritise ethics. Central considerations include addressing biases within AI systems by prioritising diverse and representative datasets to ensure equitable applications of AI. Safeguarding user privacy through robust data protection legislation is essential for building trust and promoting the widespread adoption of AI technologies. Moreover, creating accountability frameworks will assure that developers and users of AI systems are held responsible for their actions, promoting transparency and responsible AI deployment. Strengthening AI system security and resilience is also vital for mitigating cyber threats and preventing misuse and disruptions. By proactively addressing these ethical concerns and implementing well-considered regulations, Africa can create an environment conducive to the responsible and ethical utilisation of AI. This would foster trust among investors, entrepreneurs and citizens, leading to comprehensive and sustainable economic expansion across the continent. In short, Africa's progress in leveraging the Fourth Industrial Revolution and AI technology is intertwined with its ability to address ethical considerations. Drawing lessons from the EU, African nations hold the potential to lead in ethical AI deployment, fostering inclusive and sustainable advancements for the continent.

REFERENCES

- Atakishiyev, S., Salameh, M., Yao, H., & Goebel, R. (2021). Explainable artificial intelligence for autonomous driving: A comprehensive overview and field guide for future research directions. arXiv preprint arXiv:2112.11561.
- Bouzguenda, I., Alalouch, C., & Fava, N. (2019). Towards smart sustainable cities: A review of the role digital citizen participation could play in advancing social sustainability. Sustainable Cities and Society, 50, 101627.
- Benoit(2022) Rwanda inaugurates the first African Center for the Fourth Industrial Revolution (C4IR) dedicated to research and development in artificial intelligence. Available online: https://www.actuia.com/english/rwanda-inaugurates-the-first-african-center-for-the-fourth-industrial-revolution-c4ir-dedicated-to-research-and-development-in-artificial-intelligence/
- Donahoe, E., & Metzger, M. M. (2019). Artificial intelligence and human rights. J. Democracy, 30, 115.
- Du, S., & Xie, C. (2021). Paradoxes of artificial intelligence in consumer markets: Ethical challenges and opportunities. Journal of Business Research, 129, 961-974.
- European Parliament (2023) EU AI Act: first regulation on artificial intelligence. Available online: https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence
- Graham, M. (Ed.). (2019). Digital economies at global margins. MIT Press.
- Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., ... & Koedinger, K. R. (2021). Ethics of AI in education: Towards a community-wide framework. International Journal of Artificial Intelligence in Education, 1-23.
- Horgan, D., Romao, M., Morré, S. A., & Kalra, D. (2020). Artificial intelligence: power for civilisation–and for better healthcare. Public health genomics, 22(5-6), 145-161.
- Iris, Ç., & Lam, J. S. L. (2019). A review of energy efficiency in ports: Operational strategies, technologies and energy management systems. Renewable and Sustainable Energy Reviews, 112, 170-182.
- Iris, Ç., & Lam, J. S. L. (2021). Optimal energy management and operations planning in seaports with smart grid while harnessing renewable energy under uncertainty. Omega, 103, 102445.
- Jordan, A., & Schout, A. (2006). The coordination of the European Union: Exploring the capacities of networked governance. Oxford University Press.
- Koh, L., Orzes, G., & Jia, F. J. (2019). The fourth industrial revolution (Industry 4.0): technologies disruption on operations and supply chain management. International Journal of Operations & Production Management, 39(6/7/8), 817-828.

- Leslie, D., Burr, C., Aitken, M., Cowls, J., Katell, M., & Briggs, M. (2021). Artificial intelligence, human rights, democracy, and the rule of law: a primer. arXiv preprint arXiv:2104.04147.
- Magwentshu, N., Rajagopaul, A., Chui, M., & Singh, A. (2019). The future of work in South Africa. McKinsey and Company.
- Mazibuko-Makena, Z. (2021). The potential of Fourth Industrial Revolution technologies to transform healthcare: The question of access for the marginalized. Leap 4.0. African Perspectives on the Fourth Industrial Revolution: African Perspectives on the Fourth Industrial Revolution, 287.
- Mbunge, E. (2020). Effects of COVID-19 in South African health system and society: An explanatory study. Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 14(6), 1809-1814.
- Mhlanga, D. (2022). Human-centred artificial intelligence: the superlative approach to achieve sustainable development goals in the fourth industrial revolution. Sustainability, 14(13), 7804.
- Mhlanga, D. (2023). Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning. Education, the Responsible and Ethical Use of ChatGPT Towards Lifelong Learning (February 11, 2023).
- Ndung'u, N., & Signé, L. (2020). The Fourth Industrial Revolution and digitization will transform Africa into a global powerhouse. Foresight Africa Report, 5(1), 1-177.
- O'Sullivan, S., Nevejans, N., Allen, C., Blyth, A., Leonard, S., Pagallo, U., ... & Ashrafian, H. (2019). Legal, regulatory, and ethical frameworks for the development of standards in artificial intelligence (AI) and autonomous robotic surgery. The international journal of medical robotics and computer-assisted surgery, 15(1), e1968.
- Ramakrishna, S., Ngowi, A., Jager, H. D., & Awuzie, B. O. (2020). Emerging industrial revolution: Symbiosis of industry 4.0 and circular economy: The role of universities. Science, Technology and Society, 25(3), 505-525.
- Sakolkar, P. C. (2023). Impact of Digital Transformation on the Indian Government Regarding Service Delivery and Citizen Engagement.
- SAAIA (2023)The South African Artificial Intelligence Association (SAAIA)SA Artificial Intelligence Industry Association Launches in Pretoria. Available online: https://saaiassociation.co.za/sa-artificial-intelligence-industry-association-launches-in-pretoria/
- Shahroom, A. A., & Hussin, N. (2018). Industrial Revolution 4.0 and education. International Journal of Academic Research in Business and Social Sciences, 8(9), 314-319.
- Ye, L., & Yang, H. (2020). From digital divide to social inclusion: A tale of mobile platform empowerment in rural areas. Sustainability, 12(6), 2424.

Chapter 5

An AI foundation model for education

José-Miguel Bello y Villarino

1 Introduction

Last May, a group of Spanish researchers and representatives of civil society released a report entitled (in Spanish) "Eight proposals for the education system to avoid lagging behind in the data revolution" (Gortazar and Ferrer 2023). One of the main claims in the report was that, "Spain is a country with a broad culture of administrative data generation, but this occurs in a very fragmented way and with very little use for research or strategic purposes" (Gortazar and Ferrer, 2023, 7). Most scholars would consider this statement to be a fair evaluation of the situation in Spain, sadly one common to many other countries in Europe. The authors particularly noted with regret the "monopoly" of the public sector (the "administration") over education data and then proceeded to present a series of policy recommendations.

In this article, I focus on one of their central suggestions – promoting the data sharing of government-held education data with researchers and the private sector – and the assumptions underlying this recommendation. I argue that this view represents a degree of conventional wisdom that will not fully serve the interests of society in the educational data space. As an alternative, I suggest a different policy approach to the use of education data for strategic purposes. This alternative approach builds on the design, creation and deployment by the public sector of an artificial intelligence (AI) foundation model specifically developed with education data and for education purposes.

Concretely, Gortazar and Ferrer refer to previous policy work (Almunia and Rey Biel, 2021) to argue that the public entity owning the data will be better off sharing it with researchers for broader use, instead of holding it for its own (more limited) purposes. In their view – one shared with many computer scientists, policy advisors, consultants, industry and some think tanks – doing data analysis is not one of the core tasks of the State and, therefore, other entities are better placed to work with that data generated within the public sector (Gortazar and Ferrer, 2023, 9). All that is required of the State is to create the mechanisms to make the data available to others.

Contrary to these views, I argue that there is a much more promising path for use of the data generated by the public sector in the education space. In this article, I call for the development of an education-specific foundation AI model driven and directed by public authorities that can collate, process and use this data, and then provide non-government entities with access to this foundation model through Application Programming Interfaces (APIs). Such an approach avoids placing all the underlying data in the public domain or making it accessible by default to any interested party. Several benefits may be anticipated from this alternative approach.

First, this approach mitigates the risks involved in developing education applications from other existing foundation models (or their future iterations). Using Bert (Google), GPT4 (OpenAI) or Claude (Anthropic) – to name just a few of the most popular ones – trained on data mainly collected from the Internet to develop tools for education purposes can expose users to unnecessary risks.

Second, this approach will promote the use of education data that will be jurisdiction-specific. That is, the foundation model will consider not only the specificities of the curricula in that jurisdiction, but also the society's views on education.

Third, a foundation model would allow a notable degree of control over the AI systems developed for that jurisdiction for specific education-related tasks. Creating a foundation model involves developing an upstream filter responsive to education expertise and students' interests, which trickles down into any systems developed from that foundation model. This offers the developer of that foundation model a say in the ways the publicly held education data can be used for education purposes.

Yet, before entering into the details of this proposal, it is necessary to understand both options: The one based on making publicly held education data accessible to researchers and developers; and the one involving the development of a foundation model as a first step, before making that model accessible to researchers, private developers and public authorities.

In section 2 of this article, I explore the logic and departure point of these two approaches. Section 2 presents the type of foundation model envisaged here, explaining its characteristics and elements, including its jurisdiction-specific aspects. Section 3 explores options for the development of such a system, making special reference to the need for participatory approaches where different types of expertise are present. This section also covers the possible paths to deliver such a model, namely a pure public approach or a public-private partnership, with the purpose of informing policy options. Section 4 anticipates possible uses of an education-specific foundation model to illustrate the advantages of this approach. Finally, in the conclusion I present how this could happen in the immediate future (3–5 years) and the broader policy implications of this approach.

2 The two approaches to publicly held education data

AI is already transforming educational methodologies and outcomes in ways that, even if obvious given the nature of education, were not be anticipated ten years ago. As Gulson et al noted (2022) AI is just an acceleration of a longer project of datafication—that is, turning things and events into numerical data that can be added to large databases. This datafication has changed many aspects of education, from delivery and assessment, to policy design and allocation of resources.

As they note, modern education systems have been based precisely on datafication, as they depend on getting information about how students perform in different fields and then giving them credentials that confirm the accuracy of that information—what we have called until recently "big data". That big data is now combined with AI, a technology that can analyse and respond to education-raleted big data faster and better. As a result we are facing a technology that promises improvement and disruption from the smallest to the most important decisions made by individuals and organizations in the education sector (Gulson et al, 2022: 3). From the student deciding what parts of the curriculum they should revise, and then actually providing the tools to revise it; through the administrator assessing if a class should be added a support teacher, and then finding the most suitable one for the task; to the policy maker envisaging a reform of the curriculum and then assessing its long terms effect, all education-rlated matters can be within the scope of decision-assisted AI.

Yet, in all those case, any AI system developed will operate on the basis of a learning process from existing data. Concretely, for education-related AI to be effective and accurate, it will need data mainly held by the public sector. The decision about how to grant access to that data can shape how that improvement and disruption "from the smallest to the most important decisions" will happen. For the purpose of this article, and leaving aside other option of working with synthetic data or the option of private parties buying data from the public sector; this decision starts with a disjunctive: on the one side there are the advocates of open access to education data and on the other those who believe that the the public sector should use that position of prevalence in other ways.

The advocates of granting open access to education data believe that two things will happen after data is placed in the public domain. First, that it will be properly used. Second, that economic and social value will be extracted from it. These assumptions are inextricably linked to the principles developed in the Open Data Directive (2019). The Directive essentially focuses on the "reuse" of public-sector information – therefore increasing the supply of dynamic data – with the target of making it more easily available for all, including startups and SMEs.

The Directive's purpose is to support the more thorough use of that data by researchers and the private sector, while promoting competition and transparency in the information market. Essentially, if the data is available to all – once processed to ensure the necessary limitations attached to privacy and other data-related requirements – individuals and companies will extract more value from that data, driven by research objectives (e.g., universities) or market forces (e.g., companies aiming to place products using that data on the market).

In general terms, I believe their approach to the use of public data is adequate and well informed. Still, I doubt that these outcomes are either desirable or efficient in the education context.

2.1 Challenges for an Open-Data approach in the education context

The reuse of data makes perfect sense in the contexts envisaged by proponents of the Open Data Directive. Data from geographical datasets, land registry entries, statistical information or related to the legal sector (judgements, records, regulations etc.) can only offer limited value to the public sector. Their potential applications can be fully exploited solely if placed in the public domain. These are precisely the examples used by the Commission in its initial proposal and then reflected in Article 13(1) of the Directive referring to high-value datasets (geospatial, earth observation and environment, meteorological, statistics, companies and company ownership and mobility).

Moreover, the datasets initially identified as high-value contain limited (if any) personal or privacy implications (geospatial, earth observation and environment, meteorological, statistics), or refer to areas where the public interest would generally override the need to protect the privacy of individuals (e.g., company ownership).

Yet, the data from the education sector are much more than isolated data points. Education data, attached to individuals – even without any processing – tell personal stories. Data for primary and obligatory secondary schooling are also compulsorily collected, which means no person is likely to be able to avoid having their data shared. Even the idea of consent may be inadequate to counter the barriers to informed consent and any issues around the use of data, particularly from under 18-year-olds, namely, children. Also on higher levels of education, data collection is often unavoidable: marks, attendance or the choice of subjects at public universities would lie within the scope of that open-data approach.

Even the data that are collected in the first place are, in themselves, intrinsically connected to education policy. Consider, for example, data from assessments of student and teachers. How often is that data collected (how many comparable exams, tests, evaluations take place every year)? What kind of assessments are conducted (PISA-style across the board assessments, targeted assessments of some groups of students, assessments focusing only on certain subjects such as Mathematics and language)? How broad are these assessments (local, regional, state level)?

These are all policy-related questions that have shaped the data available today. Any use of such data from an open-data perspective is irrelevant if deprived of these considerations.

Similarly, the type and amount of data that the public sector is holding related to education in other non-assessment related domains can be closely attached to the views the State has about education and its existing education policy. Consider for example payment scales and incentives to education staff; levels of support to individual students and groups of students; the socioeconomic background of families; educational infrastructure, transport and meals at school; the language of schooling; timetables and holidays; changes in the curriculum; changes in the delivery method etc. This list of issues virtually has no end.

Behind the existence of these datasets, there are many cumulative decisions over the years that have determined whether the data should be collected (e.g., in relation to the socioeconomic background of families) or even the existence of the data itself (e.g., to (not) have a comparable exam for all students finishing primary school). These are all technical issues that shape education policy, but also reflect the political 'sensibilities' of the government in place and the broader view in a country about its education system. Even when limiting the comparison to the European Union, the role of education in society can be very different in Finland (Kosunen 2018, 69–70) , France (Gueudet et al. 2018, 42–44) and Ireland (Department of Education and Skills – Ireland 2011, 48–49).

2.2 Education data are not only about education

Education data collection, processing and use is then an issue of policy and politics. The way the data are collected, aggregated and processed needs to reflect that divergence of expertise (data science, education professionals), while considering the social and political implications of the way the data are collected and used in any system.

Further, the way the data are processed and interpreted can influence decisions about education delivery and policy that, taken at a given point in time, can generate changes in the real world that are then reflected in the data continuously collected by the system. For example, AI-based support systems for tutoring non-native students based on technologies developed in other jurisdictions with different curricula may only be felt in some types of tests but not in others.

Government policies beyond the education domain can also be based on education data. Migration policy can be determined by professional profiles needed in the future which can be anticipated based on current registrations in technical colleges. Even our systems of government may depend on education data, from the highest level (e.g., constitutional reform of electoral systems favouring the D'Hondt method instead of a first-past-the-post one may require some levels of complex understanding of mathematics) to the more mundane aspects of their operation (e.g., in Spain chairs of a voting station panel must have completed higher secondary education). This interconnectedness between education, government decision-making, policy design and societies strengths the arguments for a more controlled approach to the public use of education-driven data.

Finally, from a practical perspective it may be impossible to anonymise all relevant data, while keeping it meaningful for extensive use. This would favour opting for an intermediary step between the data held by the administration and the developers of AI systems for specific uses. Such an intermediary role could be played by an education-specific AI foundation model

3 An education-specific AI foundation model

3.1 What is a foundation model?

In this article, I have consistently referred to "foundation models", although it should be noted that this is just a proposed terminology in a rapidly evolving field. Indeed, several other names are commonly used, sometimes with slightly different implications. These include "General Purpose AI" (GPAI) and "large language models" (LLMs) or "large multimodal models" (LMMs). Nonetheless, in my view, the term "foundation" describes relatively well the idea that they are the basis (the 'foundation') for other AI systems that are 'fine-tuned' from that foundation model for particular purposes.

Even though in some understandings it is assumed that the models can both operate independently or form the base for other applications (Jones, 2023), for the purposes of this paper I assume that foundation models are developed with the main intention of forming the basis for other applications. This was the origin of the term as defined by researchers at the Stanford Institute for Human-Centered Artificial Intelligence in 2021 (Bommasani et al. 2021, sec. 1): "foundation model is any model that is trained on broad data (generally using self-supervision at scale) that can be adapted (e.g., fine-tuned) to a wide range of downstream tasks". In that paper, foundation models are opposed to "narrow AI systems", that is, ones developed for a specific purpose.

The amendments put forward at the European Parliament (2023) regarding the Proposal for an AI Act (European Commission 2021) use both the terminology of foundation models and general-purpose AI. The proposed new recital 60 e notes that:

> Foundation models are a recent development, in which AI models are developed from algorithms designed to optimize for generality and versatility of output. Those models are often trained on a broad range of data sources and large amounts of data to accomplish a wide range of downstream tasks, including some for which they were not specifically developed and trained. The foundation model can be unimodal or multimodal, trained through various methods such as supervised learning or reinforced learning. AI systems with specific intended purpose or general purpose AI systems can be an implementation of a foundation model, which means that each foundation model can be reused in countless downstream AI or general purpose AI systems. These models hold growing importance to many downstream applications and systems.

Foundation models are assumed to require massive amounts of data and computing resources and, therefore, cannot easily be developed. However, recent research experiences show that it is possible to significantly develop existing open-source models with a more limited investment (Touvron et al. 2023), perhaps opening the door to more players in this space. This field is currently dominated by the US companies OpenAI, Google, Meta and Anthropic and the Chinese ones Baidu and Huawei, together with the Beijing Academy of Artificial Intelligence (BAAI). Several other less relevant foundation models have also originated in the European Union and Israel.

3.2 What is the risk of using foundation models as the basis of AI systems for educational purposes?

Ever since these foundation models became popular, many have paid attention to their use in education. Among the scholarship, the work of Blodgett and Madaio (2021) is particularly interesting for this discussion. In this article – with a very explicit title that paraphrases Bommasani et al.'s: "Risks of foundation models in education" – Blodgett and Madaio present the risks of using foundation models for educational purposes, structured around six ideas:

1. The use of technology in education has tended to favour an economic logic at the expense of better outcomes (large-scale educational technologies are developed for efficiency reasons).
2. Similarly, education-related technology has traditionally given prevalence to the interest of learners from developed countries, favouring English-centric and North American perspectives. This has been accompanied by the reinforcement of existing social hierarchies and pedagogical biases.
3. Foundation models risk standardising teaching methods and content, homogenising education. This could, in turn, exacerbate existing educational disparities. Training data for such models, derived from publicly available sources on the Internet may inadvertently reinforce toxic cultures and dominant language norms, potentially marginalising and excluding minority identities.
4. Using foundation models in education might bypass the involvement of key stakeholders like teachers. This limited stakeholder involvement, together with the capital-intensive nature of AI development, could erode the agency of teachers and deviate from learner-centred educational paradigms.
5. Over-reliance on foundation models could lead to an oversimplified understanding of the complexities of teaching and learning, overemphasising pattern-matching at the expense of genuine comprehension.
6. The drive to simplify the field by making teaching and learning ‘legible’ to AI might leave out vital, nuanced aspects of education.

These six arguments are often repeated with different nuances in the more recent literature in this domain.

3.3 What makes a foundation model education-specific?

Developing an AI foundation model specifically for educational purposes and using mainly educational data calls for a strategic approach. As the model being developed is not meant to be restricted to a particular purpose, the scope of what it should be able to do is not clearly pre-determined ex-ante. Its objective should be to meaningfully process the data it will be trained with in order to generate a diversity of relevant outputs.

There are different dimensions that must be considered at the design and development levels. Here I present them following seven axes adapted from the ISO/IEC standard 22989(2022) on AI concepts and terminology

Data Collection: The collection should be presided by a principle of ensuring diversity. That is, the data should cover all relevant levels of education (e.g., pre-schooling, primary, secondary, tertiary) and it should aim to include all data subjects, not being limited by the accessibility of the data. For example, some schools in certain regions with different languages may have records in other languges. Economics of access to the data should not preclude having a data collection process as expansive as possible.

In terms of contents, at least the following material should be collected:

1. Educational Content: Documents referring to curricula, textbooks, scholarly articles, educational websites, lecture notes, video transcripts, and any other educational materials.
2. Interaction Data: This could be chat logs from online educational platforms, student queries, feedback etc.
3. Assessment Data: Data on student tests, exams, assignments and recorded in-class activities, such as corrections to homework.
4. Data from different actors involved in education normally held by the management of schools or of the education agency (salaries, promotions, support, socioeconomic background).

Data Cleaning and Preprocessing: Nomally this will require a process of standardisation of the data, segmenting it education level, subject, or any other relevant categorisation and by data format. At this level, it is common to consider removing as much as possible personally identifiable information (PII). The natural objective for this would be to maintain privacy. Yet, for a public-held foundation model this step may not be necessary as the anonymisation could also happen at the output end of the model.

Developing the model: This is the part where the partnerships with private sector companies may be essential. Basically, public entities could develop a new model specific for the educational setting or establish a partnership with one of those companies who have already develop existing multimodal models (MMMs) that could be a basis for the education one.

Training the model: Once a model is developed it will require several iterations of training on the educational data. Depending on the volume of data and the model's size, this could require significant computational resources. Training could also be accompanied by reinforcement learning from human feedback. This implies that the model may reinforced its learning from human preferences. This technique allows a "reward model" to learn from human feedback and improve in the following iterations of training. This human reinforcement could be designed to incoporte the views of the subject matter experts in the education domain (educators, policy makers, managers, experienced and inexperienced students, etc.)

Evaluation and redevelopment: At this stage the model would be ready for an initial evaluation that would allow the decision makers to understand where the project is heading. To deliver this evaluation two steps are necessary. First an evaluation metric that aligns with the desired objectives should be created. Second the proper evaluation must be conducted with the feedback from educators and students not involved in the model's design. This iterative process ensures that the model's performance improves over time and it allows the refinement of the model.

Testing of outputs outside of the evaluation metrics: The model should be evaluated and tested for other considerations before making it available for fine-tuning in narrow AI systems. Among others, it should be accompanied by clear guidelines on data privacy and security and its outputs assessed for representational harm (e.g., assuring that the model does not reinforce existing prejudices). Also the model could be tested for adversarial attacks, that is attacking the model to deliver outputs that it was not intended to generate. This adversarial testing, could determine the need of adversarial training (i.e., developing adversarial models to train the model on how to defend itself). If the system does not test well, the developing should go back to the developing and training stage.

Continuous Improvement: Finally, throughout the lifecycle of the model, the whole system should be monitored and stakeholders involved in its continuous improvement. This will involve creating the mechanisms for all actors using the foundation model (developers of narrower AI systems and users of those narrower systems) can provide their input to the entities involved in the development, so the foundation model could be improved. This could be as simple as a compulsory feedback tool that should be incorporated in any narrow AI systems that use the API to the model.

3.4 Why does each jurisdiction need to develop its own education foundation model?

While each country, whether in Europe or elsewhere, crafts its education policy based on its unique socio-political context, there is a certain degree of convergence, especially within the EU. This level of alignment is necessary to ensure mobility, collaboration, and the mutual recognition of qualifications.

However, beyond that limited degree of convergence, diversity in education policy is a reflection of how societies prioritise, structure and execute educational objectives. This diversity can be observed both within a country (among different regions or groups) and among different countries, particularly when contrasting countries in Europe and within the EU.

As foundation models would be based on these different educational objectives, it would be necessary for each member state to develop its own model and, sometimes, where there is great disparity in terms of curriculum and languages (e.g., in Belgium) it may even be necessary to create these foundation models on the regional level. For example, in Germany education is largely the responsibility of the individual federal states (Länder), leading to differences in curricula, assessment methods, and teacher training across the country.

It also cannot be excluded that socio-cultural differences, political factors and economic disparities could even make developing one single foundation model on the national level impossible. Yet, the similarities within a country, the elevated costs of developing a model, and the increasing returns of training systems with more data could create the right incentives for intra-state coordination.

4 A participatory approach to developing a foundation model

Two main considerations arise with the use of education data to shape education policy or other types of interventions in the education sector. The first is that broader social effects may be felt much later in time and be outside of the model or system developed based on the education data. These delayed effects impose the need to use existing data. It is necessary then that all stakeholders (and particularly decision-makers) are aware of the limitations and biases built into the data and the subsequent AI models developed from there.

Anticipatory innovation governance essentially involves upstream governance measures that allow us to deal with uncertainty in the way technologies may develop in the future (Tõnurist and Hanson, 2020). In a recent conference paper, co-authored by several experts from the education sector, we advocated for participatory approaches to the governance of general-purpose AI in the context of education (Swist et al. 2023). There we argued that it is necessary to favour the development of education technologies in the AI space that are informed by stakeholders.

In this section, I first look at the three objectives of education before exploring the types of expertise that would be required to develop a responsible AI foundation model with education data and for education purposes.

4.1 The three objectives of education

In a recent paper with Gulson (Gulson and Bello y Villarino 2023), we recall the work of Cranston, Kimber, Mulford, Reid and Keating (2010) on the purposes of education to better understand the modes of governance for AI in the education context.

1. Democratic equality: Education should equip learners to be engaged and skilled citizens. They should be capable of making personal decisions while also contributing to group choices. Emphasis should be on fairness and broader concepts of societal justice.
2. Social efficiency: This objective ties back to the human capital theory. It underscores the role of education in preparing individuals for their professional lives. The main goal here is fostering a robust and thriving economy. From this viewpoint, education is seen as a communal asset that has evident broader benefits. However, it is closely connected to personal gains, like obtaining qualifications.
3. Social mobility: This aspect views education chiefly as a personal asset. It is perceived as something of value where the qualifications and the way they are provided offer some people an advantage. While this is often tied to economic results, such as accessing certain employment positions, it can also relate to one's status in society.

It is important to keep these three dimensions of education in mind since they are a reminder of the different stakeholders involved in education – from the individual learner to broader society – which should also play a role in developing a foundation model specific to education.

4.2 The types of expertise involved in this process

I previously noted the need to develop a foundation model with adequate representation of the different stakeholders in the education sector. I suggested, in line with prior work with my colleagues, following a participatory model. In this subsection, I proceed to identify those stakeholders who will play a pivotal role in determining how the model will function, who it will serve, and which outcomes it will produce (among others, see: Williamson 2016; Zeide 2017; Celik et al. 2022; Holmes, Bialik, and Fadel 2023)

These stakeholders will often have overlapping and sometimes conflicting interests. Collaboration, open communication, and iterative feedback loops among them all are essential for the successful development of a foundational model. I have structured them in their relative order of importance, albeit it is important to know that their input is not equally relevant in the different parts of the process. Developers should have a better insight about what is possible. Policymakers should be able to balance different priorities. Students should be the central concern when assessing risks and harms and educators should be the first point of input in terms of usability.

- **Society as a whole:** Following the first of the three objectives of education, the starting point for the design of such a system is to take society's preferences into consideration. Both in terms of democratic equality and social efficiency, it is the collective made up of all members of the society who will be the potential beneficiaries of the advantages the model may generate. Further, as taxpayers, they will play a role in financing it.
- **Policymakers and Government Agencies:** As the entities responsible for translating the common interest in policies, elected and appointed policymakers and public employees are the most important individual stakeholders in this process. They should drive the initiative and ensure that the interests of the other stakeholders are accounted for.
- **Students:** As the chief beneficiaries of the AI education system, they will be the most affected by the model. They will interact directly with AI systems developed from the foundation model. They will be impacted the most in terms of social mobility from the benefits (or harms) derived from changes in education policy or administrative decisions shaped by the model. The learning tools and systems to assist educators developed from the foundational model will also most directly impact them.
- **Educators (Teachers, Tutors, Trainers):** They will integrate and use the AI tools developed from the foundation model. They will also be particularly concerned with monitoring the outputs of the model that will feed into the different education systems.
- **Parents and Guardians:** They will be especially concerned with the quality, ethics and effectiveness of the model. As the guardians of minors, they will also play a role in demanding that the model is aligned with their children's educational goals.
- **Educational Institutions (Schools, Colleges, Universities):** They will be the core environment where the applications of the foundation models will be deployed. If used to developed tools for policy decisions, these institutions could also be directly impacted by the model.
- **AI Developers and Engineers:** As the main individuals in the design, development and maintenance of the system, they should possess suitable understanding of the technical aspects but also the needs and objectives of the model. They work closely with educators to tailor the system to the needs of the classroom.
- **Research and Academic Community:** They should provide the model's foundations in terms of its pedagogic purposes and anticipating

possible approaches in its development. They should also play a role in monitoring impacts and effects, and facilitating improvements via the development of new methodologies.

- **Commercial Entities (EdTech Companies):** They will play an important role in developing the tools that would use these foundation models. An important part of the know-how for developing and testing the models would originate from these companies, which also have a vested interest in developing the best possible foundation model as that would determine the quality of their products.

5 The deployment of an education-specific foundation model: two examples of possible system-specific uses

Now, I turn to the possible uses of such a foundation model. As noted above, foundation models have no concrete purpose and can be fine-tuned as AI systems for specific tasks. For the purposes of this paper, I will remain neutral about who could possibly do that fine-tuning, as the key point here is that the public entity that led the development of the foundation model will control access to that model and impose whichever restrictions it deems appropriate.

In any case, the two examples offered below of AI single-purpose systems developed from the foundation model seem differently suited for private and public development. The first one is individual tutoring. In the context of education, there is nothing more appealing than the possibility of developing individualised tutoring that can be adapted to the needs of each student. Since different providers may be interested in developing tutoring systems adapted to the jurisdiction where the model was developed, it may be an area where there is a role for the private sector. As these fine-tuned systems would mainly target parents, private schools or the public education systems as buyers, some degree of competition seems desirable.

On the other hand, the second example involving an AI system supporting teachers to develop classroom materials based on the curriculum for students with differentiated needs – for example through a chatbot for teachers, as currently developed based on GPT4 in the state of South Australia in Australia – would be an AI system that is more properly suited to be developed by public authorities through participatory methods.

5.1 Individual tutoring

The "two-sigma problem" in education refers to a phenomenon observed by the educational researcher Benjamin Bloom in his 1984 study. In that study, Bloom found that students who had received one-on-one tutoring, using mastery learning techniques, performed two standard deviations (or two sigmas) better than students who had been given traditional classroom instruction. The 'problem' is how to scale this benefit to larger populations noting the resource-intensive nature of one-on-one tutoring.

The biggest challenge of the two-sigma problem is the resource requirement of one-on-one tutoring. AI systems can be scaled to reach millions of students simultaneously, breaking down the barriers of cost and accessibility. At the same time, knowing that the system is based on a foundation model that has been appropriately designed, trained and tested through participatory methods involving relevant stakeholders in education in that society is a much better guarantee that the biases built into the learning experience for all students, regardless of their background or previous knowledge, would have been better considered than in an off-the-rack tutoring system developed based on an obscure foundation model.

AI foundation models developed in the same jurisdiction where the tutoring is taking place, and trained with relevant data for the students in that jurisdiction through a participatory method that understands the needs and circumstances of the inputs used for the training, can play a pivotal role in making individual tutoring possible, relevant and safe. Besides, unlike human tutors, AI systems could be available around the clock, allowing students to learn at their own pace and at times most convenient for them, which may be very relevant for students from less privileged backgrounds.

AI can also tailor learning to the individual needs of each student, just like a human tutor would. By assessing a student's existing knowledge, strengths, weaknesses, and learning style, the AI can present the relevant material it has been trained with in the most appropriate manner for that particular student and, having learned common problems with that material based on the experience of other students and the materials previously developed by teachers to address those problems, it could address the individual needs of students in both an adapted yet tested manner.

Such a system could also provide immediate feedback. It could gather student input in real time and give instant affirmations, corrections and explanations, shown to facilitate learning (Shute 2008). That feedback could also be used for the system to design adaptive learning pathways, based on the continuous assessment of a student's progress. A system of this nature could adjust the curriculum in real time, ensuring that students are always working at the edge of their competency. This is in line with Vygotsky's "Zone of Proximal Development", which emphasises the importance of presenting challenges that are neither too easy nor too hard (Shabani, Khatib, and Ebadi 2010).

Further, a foundation model trained with videos of classes and presentations, recordings of lessons and images contained in education materials or produced by teachers and students could provide the basis for a much richer AI-driven tutoring system. This could favour engagement through interactivity in a way that could be a key factor in the success of tutoring.

5.2 Specific curriculum development

The other example is centred around the work of teachers. The time constraints teachers experience is a well-documented challenge in education that can affect primary and secondary school teachers in multiple ways (Hargreaves 2001, 95–116). Two concrete aspects of the teaching role are particularly affected by time constraints. First, some time-consuming tasks, such as preparation and the marking of assessments, are less flexible, while others, such as collaboration with colleagues and planning and preparation, tend to be prioritised less.

Although collaboration can lead to shared resources, ideas, and teaching strategies, time constraints can impede these collaborative efforts. Limited time also makes it difficult for teachers to adequately prepare lessons, differentiate instruction for various student needs, and integrate different resources into the classroom effectively.

An AI system could be developed from a foundation model to help a teacher prepare materials for individual classes. Since it would be based on a jurisdiction-specific model, the contents proposed would be adapted to the curriculum. Given that it could also be used to generate materials (e.g., activity sheets), it could cater to children with differentiated capacities as well.

While not a substitute for teacher collaboration, having a system that feeds from a model trained on a myriad of data including the work of other teachers could provide an alternative to peer-to-peer or mentor–mentee collaboration. AI systems could scrape and analyse vast amounts of educational content, filtering and suggesting materials that align with specific learning objectives, student proficiencies, or interests. This could help in dynamically generating a curriculum.

6 Conclusion: The way forward

In this article, I have identified the opportunity that foundation models offer to transform education in the EU. This approach strategically extracts value from the data generated over the years in our public systems while opening the door for private involvement. It achieves those objectives without jeopardising public control of the privacy of the data.

Although not noted explicitly in the text, a publicly-owned foundation model was also a proposal for an enhanced system of governance for education data and, indirectly, Ed-Tech. Public authorities will determine the configuration and design of the foundation model and likely retain control. This is equivalent to holding the key to access that education data. Concretely, sections 4, 5 and 6 outlined a path for policy implementation. Public authorities can find on those pages a tool for evidence-based policymaking.

The described approach can also offer an industrial policy dimension. The foundation model's development can be undertaken directly by the public sector, but will be more likely developed in collaboration with private entities with some experience in the domain. In terms of efficiency, it would probably be easier to resort to the usual companies that have already developed large language models. However, given the strong influence of local characteristics in the education domain, these foundation models provide an opportunity for companies to exploit the comparative advantage attached to their local knowledge and expertise in their natural operating environment.

Finally, as these foundation models need to be fine-tuned in systems with narrower purposes (like the individual tutoring system of the one tasked with the generation of personalised curriculum development materials), this would create new opportunities for the local industry. Many small and medium-sized enterprises would be incapable of using the education data even if made accessible to them, mainly due to the large investments that would be needed to handle the data. This, in turn, would favour the position of large international corporations. However, foundation models simplify access to that data, creating an intermediary between the desired task and the data that was used to train the models through APIs. This would increase competition in the market and governments could even generate some revenue if they decided to charge for access to those APIs.

REFERENCES

- Almunia, Miguel, and Pedro Rey Biel. 2021. Towards a Cultural Change in Data Management in Spain: A Reform Proposal. https://www.esade.edu/ecpol/en/publications/data-management-spain/ (August 1, 2023).
- Blodgett, Su Lin, and Michael Madaio. 2021. "Risks of AI Foundation Models in Education." http://arxiv.org/abs/2110.10024 (August 28, 2023).
- Bommasani, Rishi et al. 2021. "On the Opportunities and Risks of Foundation Models." arXiv preprint arXiv:2108.07258.
- Celik, Ismail, Muhterem Dindar, Hanni Muukkonen, and Sanna Järvelä. 2022. "The Promises and Challenges of Artificial Intelligence for Teachers: A Systematic Review of Research." TechTrends 66(4): 616–30.
- Cranston, Neil et al. 2010. "Politics and School Education in Australia: A Case of Shifting Purposes." Journal of Educational Administration 48(2): 182–95.
- Department of Education and Skills - Ireland. 2011. "National Strategy for Higher Education 2030 - Report of the Strategy Group."
- Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on Open Data and the Re-Use of Public Sector Information (Recast). 2019. OJ L http://data.europa.eu/eli/dir/2019/1024/oj/eng (August 1, 2023).
- European Commission. 2021. Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts.
- European Parliament. 2023. "Amendments Adopted by the European Parliament on 14 June 2023 on the Proposal for a Regulation of the European Parliament and of the Council on Laying down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))(1)." https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html (August 29, 2023).
- Gortazar, Lucas, and Álvaro Ferrer. 2023. Ocho Propuestas Para Que El Sistema Educativo No Se Quede Atrás En La Revolución de Los Datos. https://doi.org/10.56269/20230510/LC.
- Gueudet, Ghislaine, Laetitia Bueno-Ravel, Simon Modeste, and Luc Trouche. 2018. "Curriculum in France." In International Perspectives on Mathematics Curriculum, IAP, 41–69.
- Gulson, Kalervo, and José-Miguel Bello y Villarino. 2023. "AI in Education (Holding Title)."

- Gulson, Kalervo N., Sam Sellar, and P. Taylor Webb. 2022. Algorithms of Education: How Datafication and Artificial Intelligence Shape Policy. Minneapolis, University of Minnesota Press.
- Hargreaves, Andy. 2001. Changing Teachers, Changing Times: Teachers' Work and Culture in the Postmodern Age. A&C Black.
- Holmes, Wayne, Maya Bialik, and Charles Fadel. 2023. "Artificial Intelligence in Education." In Data Ethics : Building Trust : How Digital Technologies Can Serve Humanity, Globethics Publications, 621–53. https://doi.org/10.58863/20.500.12424/4276068 (August 29, 2023).
- Jones, Elliot. 2023. "Explainer: What Is a Foundation Model?" https://www.adalovelaceinstitute.org/resource/foundation-models-explainer/ (August 28, 2023).
- Kosunen, Sonja. 2018. "Access to Higher Education in Finland: Emerging Processes of Hidden Privatization." Nordic Journal of Studies in Educational Policy 4(2): 67–77.
- Shabani, Karim, Mohamad Khatib, and Saman Ebadi. 2010. "Vygotsky's Zone of Proximal Development: Instructional Implications and Teachers' Professional Development." English language teaching 3(4): 237–48.
- Shute, Valerie J. 2008. "Focus on Formative Feedback." Review of educational research 78(1): 153–89.
- Swist, Teresa et al. 2023. "How Might the Possible Futures of General Purpose AI Models in Education Be Governed? A Participatory Agenda Setting Proposal." ChatLLM23 Sydney, 23 April 2023.
- Tõnurist, Piret, and Angela Hanson. 2020. Anticipatory Innovation Governance: Shaping the Future through Proactive Policy Making. Paris: OECD. https://www.oecd-ilibrary.org/governance/anticipatory-innovation-governance_cce14d80-en (August 28, 2023).
- Touvron, Hugo et al. 2023. "Llama 2: Open Foundation and Fine-Tuned Chat Models." http://arxiv.org/abs/2307.09288 (August 28, 2023).
- Williamson, Ben. 2016. 15 European Educational Research Journal Digital Education Governance: An Introduction. SAGE Publications Sage UK: London, England.
- Zeide, Elana. 2017. "The Structural Consequences of Big Data-Driven Education." Big Data 5(2): 164–72.

Chapter 6

Towards emerging technologies and e-government: The case of Croatia

Sabina Hodžić

1 Introduction

Digital technology is the foundation for transforming the economy, society and government. Therefore, the key to this transformation is the introduction of various emerging technologies. Today, different types of emerging technologies can be found on the market, such as artificial intelligence (AI), machine learning, 5G and the Internet of Things, biometrics, virtual reality, blockchain, robotics, natural language processing, quantum computing, and others. They all aim to increase the efficiency and transparency of the governing system. In the current situation in Croatia, the digital transformation was accelerated by the COVID-19 pandemic, affecting both businesses (home-based work) and administration (new online administrative services). Emerging technologies thus form the core of the administration's organisational structures. Moreover, their role is to improve the functioning of e-government and the e-participation model. As stated in a study by Hodžić et al. (2021), EU governments should strive to implement an open government approach and use digital communication channels intended for new technologies to provide consistent public information to their citizens. As a result, some EU member states (Estonia, Denmark, Finland, Germany) have invested significant financial resources in emerging technologies to automate key administrative tasks, improve public service delivery, and promote transparency and accountability.

The aim of this study is to present the situation along with the prospects for the further development of new technologies supporting e-government in Croatia. In addition, the study considers the practices being implemented. Together with the benefits, the study also highlights the obstacles given that significant financial resources are required for implementing the emerging technologies.

The study is organised as follows. After a brief introduction, Section 2 describes the development and current state of e-government in Croatia, with a focus on the E-Croatia 2020 strategy and the Digital Croatia 2032 strategy. The challenges and opportunities of emerging technologies like artificial intelligence and blockchain technology are presented in Section 4, coupled with a detailed SWOT analysis for blockchain technology in tax administration. Section 6 provides policy recommendations for Croatia, while section 7 presents a conclusion.

2 The development and current state of e-government in Croatia

Digital transformation is on the rise, impacting every aspect of life, human communication, business operations, and the functioning of the economy in modern society. The development of e-government has been monitored by the UN's E-Government Development Index since 2001. E-government is a consequence of three processes of socio-economic progress. These are the technological revolution, the change in administration and the orientation of government and politicians towards cost reduction, efficiency and closing the gap between themselves and citizens with the help of emerging technologies. On the EU level, there are several indicators of the present state of digitalisation, with the Digital Economy and Society Index (DESI) being a composite index that measures the progress of EU member states towards a digital economy and society using relevant indicators of digital performance. It is divided into five main dimensions (connectivity, digital skills, Internet usage, integration of digital technologies, digital public services), where each is subdivided into several sub dimensions, in turn made up of individual indicators.

According to the latest available report DESI (2022), Croatia ranks 21st out of 27 EU member states. Between 2017 and 2022, the DESI index rose slightly more than that of the EU. In addition, emerging technologies have continued to gain popularity among Croatian companies: 35% of them use cloud solutions, 43% use electronic invoices, and 9% use AI technologies. As a result, in the "integration of digital technologies" dimension, Croatia ranks 14th (36.7) among EU member states (36.1). The overall score for the "Integration of digital technologies" dimension is shown in Table 1.

Table 1: **Overall score for the dimension – Integration of digital technology**

	CROATIA			EU
	DESI 2020	DESI 2021	DESI 2022	DESI 2022
SMEs with at least a basic level of digital intensity (% of SMEs)	NA	NA	50% (2021)	55% (2021)
Electronic information sharing (% of enterprises)	26% (2019)	26% (2019)	24% (2021)	38% (2021)
Social media (% of enterprises)	22% (2019)	22% (2019)	24% (2021)	29% (2021)
Bigdata (% of enterprises)	10% (2018)	14% (2020)	14% (2020)	14% (2020)
Cloud (% of enterprises)	NA	NA	35% (2021)	34% (2021)
AI (% of enterprises)	NA	NA	9% (2021)	8% (2021)
ICT for environmental sustain-ability (% of enterprises having a medium/high intensity of green action through ICT)	NA	75% (2021)	75% (2021)	66% (2021)
e-Invoices (% of enterprises)	12% (2018)	43% (2020)	43% (2020)	32% (2020)
SMEs selling online (% of SMEs)	21% (2019)	30% (2020)	29% (2021)	18% (2021)
e-Commerce turnover (% of SME turnover)	9% (2019)	14% (2020)	13% (2021)	12% (2021)
Selling online cross-border (% of SMEs)	10% (2019)	10% (2019)	13% (2021)	9% (2021)

Source: DESI report – Croatia, 2022.

The above data allow the conclusion that Croatian companies are taking advantage of online trade opportunities, i.e., 29% of SMEs sell online (above the EU average of 18%), while 13% of all SMEs sell cross-border, with 13% of sales coming from the online segment. As for new technologies among Croatian companies, 35% of them use cloud solutions, 43% use e-invoices, and 9% use AI technologies.

In the “digital public services” dimension, with 53.6 points, Croatia ranks 23rd among EU member states (67.3 points). The Croatian public administration offers a wide range of online services through the national web portal e-Citizen, used over 33.5 million times in 2021. Nevertheless, Croatia is still underperforming in the availability of digital public services, with a score of 69 for digital public services for citizens vs. the EU average of 75 and 68 for businesses vs. the EU average of 82. Better results are recorded in the “open data” sub dimension (84% vs. 81% in the EU).

These results are part of the digital platform e-Citizens in Croatia. As a central point for all public sector information and services, it provides useful information and services for EU citizens, but also for all other foreign nationals with temporary residence in the Republic of Croatia. This platform consists of three main components: the central government portal system, the national identification and authentication system and the personal use box system. By issuing the electronic identity card (eID) with an identity certificate, the Ministry of the Interior has enabled access to all electronic services. The main task of the e-Citizens digital platform is to speed up communication between citizens and the public administration and to increase the transparency of the public sector. The services that can be found on the platform are:

- Certificate of no criminal proceedings - enables submission of an application for the issuance of a certificate of no criminal proceedings for the purpose of employment, exercise of social welfare rights, exercise of health, disability or pension insurance rights or other purposes.
- Certificate of Good Conduct (criminal record certificate) - enables submission of an application for the issuance of a certificate of good conduct (criminal record certificate) upon entry into force of the Croatian Government Decision on Calling Local Elections and until the submission of candidacies for local elections.
- Real Property Registration and Cadastre Joint Information System - enables the issuance of excerpts from the land registers and the book of deposited contracts.
- e-Registration for marriage - enables online submission of the notification of the intention to enter into a civil marriage, reservation of date and location within official premises, and online payment of administrative fees.
- e-Newborn - enables parents to register their child’s name and regulate the status of a newborn in the state records.
- e-Registration for life partnership - enables online submission of the notification of the intention to register a life partnership in the official premises.

- Electoral Register - serves to verify one's entry in the electoral register of the Republic of Croatia.
- Electoral Register - e-Temporary registration - enables you to submit an online request for a change of the voting location in Croatia and abroad, as well as online registration for voting (for persons without a Croatian ID).
- Registers of non-profit organisations - enables retrieval of electronic records from the Register of Associations, Register of Foreign Associations, Register of Political Parties, Register of Foundations, Register of Foreign Foundations, Register of Legal Persons of the Catholic Church and Register of Religious Communities in Croatia as well as the Register of National Minority Councils, Coordination Bodies of National Minority Councils and National Minority Representatives.
- e-Case - enables parties, attorneys-in-fact and other interested parties participating in court proceedings to get informed about the course and dynamics of case resolution in regular and appellate proceedings, i.e. provides them with access to basic information on court cases.
- e-Bulletin board - enables viewing of the online bulletin boards of courts and other competent bodies in Croatia.
- Commercial court register (Companies register) - a public register comprising information and documents on entities registered in accordance with the law, kept by commercial courts.
- Organised Land - enables search and browsing of basic land registry data and basic cadastral alphanumerical and graphic data for all users, without the need for registration.
- Insolvency Register - an electronic register established in accordance with Regulation (EU) 2015/848 of the European Parliament and of the Council of 20 May 2015 on insolvency proceedings, in order to enhance the provision of information to relevant creditors and courts about initiated insolvency proceedings.
- e-Communication - a service for electronic communication between participants in court proceedings and the courts.
- e-Enforcement - a service enabling the submission of motions for enforcement to municipal courts on the basis of an authentic document.
- e-Tax Administration – citizens and businessess can fill in and submit requests for issuance of a tax certificate and requests for issuance of/ entry of changes into the tax card by logging in to the system.
- e-Visitor - enables registration and de-registration of tourists.
- e-Nautics - enables the registration of the arrival of a foreign vessel or a Croatian boat in Croatian territorial waters and the issuance of an

electronic confirmation of the payment of fees for navigation safety upon registration of a yacht or boat (the so-called "vignette").

- e-Crew - allows charter companies to perform mandatory registration of crew and passenger lists online autonomously by the moment of departure of the vessel, at the latest.
- and others.

In addition to citizens, the digital platform also offers electronic services for businesses. This is important in order to increase the efficiency and productivity of services and to gain access to information and markets. For example, the registration of a company and other changes can be made online.

The idea of this digital platform is a distributed and decentralised system in which each ministry or other public institution develops, implements and maintains e-services in its respective area of responsibility. In addition, this will bring public administration closer to citizens and businesses through the use of the Internet. Apart from disadvantages such as high investment costs, lack of skilled labour and others, there are numerous advantages, e.g. lower labour costs, improved efficiency and higher quality of services and transparency.

A project to modernise the Shared Services Centre (CDU), funded by the National Resilience and Recovery Plan, was to begin implementing the blockchain platform in 2022. The CDU's main goal is to centrally manage and consolidate the state's information infrastructure, data, applications, operations, and horizontal processes to improve the transparency, accountability and efficiency of public administration. By the end of 2023, the state cloud is expected to provide interoperability with over 300 institutions (ministries, public institutions, local units and regional self-governments, and others). In particular, the use of new technologies in areas like public procurement and taxation will increase efficiency, improve transparency, and reduce opportunities for corruption and tax evasion.

The Regulation on Office Operation was adopted (Official Gazette 75/2021), which provides for the obligation to adapt, i.e., to establish, information systems for the office operation of the state administration. It enables complete office operation in electronic form and imposes a functional obligation to connect and exchange data with other information systems managed separately for certain administrative areas, as well as the possibility to connect and exchange data with the reporting system on the state of completion of administrative procedures. In addition, a platform of electronic services for e-signature and e-seal has been established, enabling electronic and mobile signing, as well as certification and verification of the validity of e-signature and e-seal within the scope of the activities of state and public administration bodies. By June 2022, 33 institutions were connected to the State and Public Administration platform, i.e., local units and regional self-governments.

According to an eGovernment benchmark study (2021), the level of digitisation in public administration is (e-administration) is 61% and the Republic of Croatia appears in 26th place in Europe (out of 36 countries studied). The study shows that Croatia is not sufficiently exploiting its ICT potential for the provision of public e-services and other public services. Although some areas in Croatia were already adequately digitised several years ago, the study indicates that the country still needs to create important conditions for improving the entire system of public e-services, including the regular monitoring of technological trends, implementation of advanced technological solutions in the digitisation of services and administrative procedures, and improvement of the existing information infrastructure and systems, especially with regard to basic registration systems.

3 The E-Croatia 2020 Strategy and Digital Croatia 2032 Strategy

In line with the European Commission's guidelines, i.e., the Digital Single Market Strategy, Croatia adopted an e-government strategy in 2017 called e-Croatia 2020. This strategy formed part of the government's e-government and digitalisation plan. The goal of the e-government strategy is to create interoperable e-government systems and services and reduce bureaucracy. To achieve this mission, the Croatian government must overcome several challenges. These challenges include training public administration employees in ICT, establishing one-stop shops for the real world, regulating business processes, arranging and collecting data in public registers, and developing a network that enables ultra-fast access (100 Mbit and above) across public institutions, central government, and self-government units. The financial costs will be covered by national funds and co-financing from the European Union under the 2014–2020 Multiannual Financial Framework – in collaboration with other ministries, public institutions, businesses and the academic community.

The main strategies guiding the e-Croatia 2020 strategy are:

1. the National Cyber Security Strategy (NCSS);
2. the Strategy for Broadband Development in the Republic of Croatia 2016–2020; and
3. European and national strategic contexts.

In order to monitor the development of e-government, various services for citizens and business have been developed. These services for both citizens and businesses include e-citizens, e-tax, e-health, e-schools, an e-permits, e-tourism etc. The preconditions for the development of e-services are electronic identification (eID), electronic documents (eDocuments), authentic sources, electronic safe (eSafe) and Single Sign On (SSO). In the area of finance and taxes, obligatory e-services are the following (e-Croatia 2020 Strategy):

1. “fiscalisation a service of the tax administration which collects information on every invoice the moment they are issued;
2. services, submission of forms via the eTax portal, including groups of services/forms such as value added tax, income tax and contributions (JOPPD form), profit tax, consumption tax and lottery and prize draw competitions;
3. e-customs refers to the calculation and collection of tax revenues from customs duties on imports and exports, better and higher quality control of excise goods subject to excise duties;
4. e-excise as of 1 September 2014, all excise duty payers and payers of special taxes are obligated to submit all forms electronically;
5. submission of the Reports on Receipts, Income Tax, Surtax and Contributions to Compulsory Insurance (JOPPD form);
6. electronic submission of all the available forms is obligatory for taxpayers classified as medium-sized and large enterprises within the meaning of the Accounting Act”.

Apart from all the advantages, the fundamental issue concerns the question of information, data security and personal privacy. This makes it important to determine regulations and/or adopt new regulations on the national and local levels for the digital economy and digital rights. Following the European legal framework context, the following directives, regulations and proposals are relevant: Directive 2006/123/EC on services in the internal market, Directive 2014/55/EU on electronic invoicing in public procurement, Regulation 910/2014 on electronic identification and trust services for electronic transactions in the internal market, Directive 2014/24/EU on public procurement, Directive 2011/24/EU on the application of patients’ rights in cross-border healthcare, Directive 2003/98/EC on the re-use of public sector information, amended by Directive 2013/37/EU, and Proposal for a Directive of the European Parliament and of the Council on the accessibility of public sector bodies’ websites.

Since joining the European Union, Croatia has also needed to comply with the European Convention on Human Rights and the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. The legal framework for the e-government model in Croatia is governed by the following legislation:

1. Constitution of the Republic of Croatia (Official Gazette Nos 56/90, 135/97, 8/98, 113/00, 124/00, 28/01, 41/01, 55/01, 76/10, 85/10);
2. Act on Personal Identification Number (Official Gazette No 60/08);
3. Act on the Protection of Personal Data (Official Gazette No 103/03); and
4. Act on Information Security (Official Gazette No 79/07).

The listed legislation establishes guidelines for solving problems in the digital market and certain rules for public administration.

As part of continuous improvement in the area of the digital economy caused by rapid ICT development, at the end of 2022 Croatia adopted a new strategy – Digital Croatia Strategy for the period until 2032. This strategy is a multi-sectoral strategic planning act aligned with the National Development Strategy until 2030 and the basic documents and policies of the EU and the Republic of Croatia, including the National Recovery and Resilience Plan of which it forms an integral part. The mission of this strategy is to strengthen inter-institutional cooperation and coordination for a successful digital transition of society and economy. To achieve this, the application of emerging technologies such as 5G/6G, artificial intelligence, machine learning, cloud computing, technology of large amounts of data, i.e., Big Data, and blockchain technology in the public and private sector are inevitable. It also remains open to the implementation of some future disruptive technologies that may appear in the future. Implementation of the aforementioned emerging technologies will enable the better processing and use of data, which in turn will contribute to the more efficient work of public institutions, creation of data-driven public policies, personalisation of public services, reduction of administrative burden, more efficient communication between public institutions and citizens, and better opportunities for collaboration between the public and private sectors.

To accomplish all the digitisation goals, the Croatian government must follow the principles of good governance and 'do no substantial harm'. It will also support a values-based and ethical approach as a foundation for the digital transformation. Following the European Green Deal strategy, it will give high priority to the protection of natural resources, taking into account that the introduction of new digital technologies will help reduce energy consumption and thus harmful gas emissions.

Pursuant to the defined vision and mission, four strategic goals were established in four priority areas, with the aim of digitising Croatian society, the public administration and the economy in the period up to 2032. These are the following (Official Gazette, No. 2/2023):

1. "A developed and innovative digital economy" – the goal is to support digital innovation centres and digitisation in micro, small and medium-sized enterprises through planned interventions, digitise public services for entrepreneurs, ensure the availability of anonymised data, transform

and strengthen the competitiveness of creative and cultural industries, and optimise Croatian tax and parafiscal legislation and administration;

2. "Digitalised public administration" – to be achieved by modernising government information infrastructure and advanced software solutions, achieving the full interoperability of public administration and enabling data access for citizens and businesses, strengthening organisational and human institutional capacity, digitising all key public services, and promoting digital services and customer support to citizens. This can be accomplished by further investing in the public administration's operational efficiency through the development of modern and effective internal digital resources (including hardware and software infrastructure, networked databases, digitised internal processes, digitally trained staff, and a strengthened organisational structure), the user experience of citizens, this includes administrative facilitation and easier access to services through the digitisation of key public services, covering the entire life situation of citizens and the business situation of legal entities, the application of the standards for the development of public e-services in the Republic of Croatia, the use of emerging technologies to ensure better use of collected data for both the creation of public policies based on real data and more personalised access to public services.
3. "Developed, available and used networks of very large capacities" – to be achieved by creating the conditions for spatial planning and the faster construction of networks, regulating the impact of land use costs on network expansion, supporting network expansion in areas where there is no commercial interest in investment, and promoting the use of high-speed services;
4. "Developed digital competencies for life and work in the digital age" which will increase the number of ICT experts in the labour market through planned actions, develop citizens' digital skills for living and working with ICT, and implement the digital transformation to support the development of the education and research system. In addition, the internationalisation of higher education and the labour market will help to attract more foreign students and experts in ICT. Ensuring the sustainability of foreign language study programmes and joint study programmes implemented by higher education institutions from the Republic of Croatia is an important prerequisite for the further internationalisation of the country's higher education system, as well as for improving the quality of higher education through greater integration into the European and global higher education space. In order to reap all the benefits of the digital transformation and increase the competitiveness and value of labour, it is necessary to work on the development of workforce competencies for the application of digital technologies in professions that are not IT and on the development of human resources for traditional industries and professions that are

adapted to the needs of the digital environment. Formal and informal educational programmes, created by applying the tools of the Croatian Qualifications Framework through the awarding of lifelong learning vouchers, will ensure the acquisition of the digital skills necessary for work for employed and unemployed people, including vulnerable groups like the young or long-term unemployed. Further, higher education institutions will be encouraged to implement shorter programmes that improve and renew the digital skills needed for the labour market and economic development. Employment policies and the legal framework for the modern labour market and economy of the future will also be improved, active employment policies will be further developed, and a special focus will be placed on the inclusion and preparation of the long-term unemployed for jobs as part of the digital transformation.

The financial framework for implementing this Strategy is included in the state budget, Multiannual Financial Framework of the European Union (for the periods 2014–2020 and 2021–2027), and European Mechanism for Recovery and Resilience (areas defined by the National Recovery and Resilience Plan 2021–2026). The means for implementation are planned in the financial framework of medium-term strategic planning acts. These are the National Public Administration Development Plan for the period 2022–2027; National Broadband Access Development Plan (2021–2027); National plan for the development of the judicial system (2022–2027); National plan for equalising opportunities for people with disabilities (2021–2027); National Health Development Plan (2021–2027); and National Island Development Plan (2021–2027). Accordingly, implementation of the strategic goals and public policy priorities of the Digital Croatia 2032 Strategy will rely largely on financing from available EU sources. Further, a financial framework is estimated for each targeted strategic goal. For example, for the strategic goal – A developed and innovative digital economy, approximately EUR 303 million is estimated, for the strategic goal – Digitalised public administration, around EUR 515 million is estimated, while for the strategic goal – Developed, available and used networks of very large capacities the estimate is around EUR 311.1 million, and for the strategic goal – Developed digital competencies for life and work in the digital age, around EUR 286 million. The final total budget amount for financing all goals and activities will be known after the negotiations with the European Commission and following coordination with other relevant ministries. Moreover, an important part of efficiently using the financial resources will be monitoring the implementation of the Digital Croatia 2032 strategy and the method and dynamics of reporting on its implementation and evaluation during the period (2022–2032) covered by the strategy.

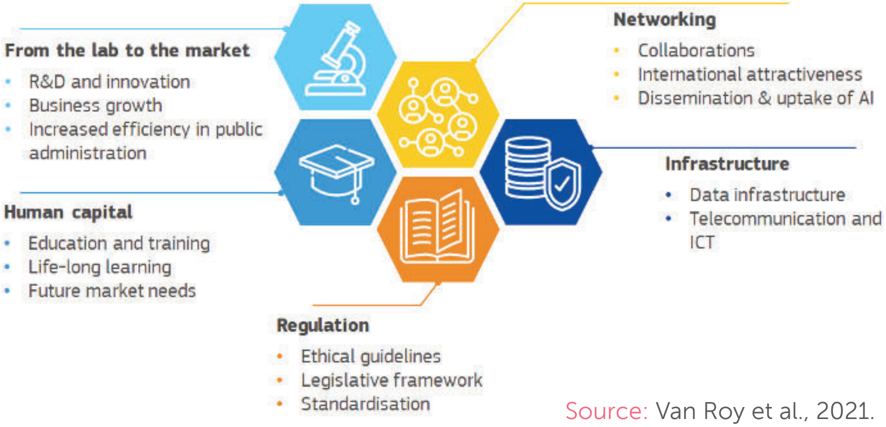
All of these goals are aligned with the EU's 2030 digital goals through concrete details of planned developments on the EU and national levels, key performance indicators to track progress towards meeting the digital goals, an annual cooperation cycle to monitor and report on progress, and cross-country projects combining EU, member state, and private sector investments.

4 Challenges and opportunities for emerging technologies – artificial intelligence and blockchain

Artificial intelligence and blockchain play a central role in the success of the green and digital transformations in Europe and in technological sovereignty. These are two of the most significant emerging technologies that will be responsible for a major impact on future societies and economies. Efforts to develop artificial intelligence and blockchain technologies have seen a rapid increase in recent years. At the same time, huge financial resources are being invested in the application and development of intelligent systems in various fields such as communication, commerce, healthcare, Internet research, production processes, education, financial services and others. In order to drive the progress of new technologies, the European Commission has taken several measures: The Horizon 2020 programme has allocated EUR 1.5 billion to AI for the 2018–2020 period; the Digital Europe programme, as part of the 2021–2027 Multiannual Financial Framework, will complement this by allocating a further EUR 2.5 billion to invest in and unlock the use of AI by businesses and public administrations. On the global level, investments in AI and blockchain amounted to EUR 80–85 billion between 2010 and 2019 (annual growth rate of 38%). As the EU's goal is to make the EU a world-class hub for AI, all EU member states and Norway signed a Declaration on Cooperation in Artificial Intelligence (2018) to work together on the opportunities and challenges of AI. By June 2021, 20 member states and Norway had adopted national AI strategies, while 7 member states were in the final drafting stages.

The European Commission (2019) defines AI as "systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals". This system can be purely software-based, acting in the virtual world (e.g., voice assistants, image-analysis software, search engines, speech- and face-recognition systems) or embedded in hardware devices (e.g., advanced robots, autonomous cars, drones or IoT applications). The opportunities for using such a system include learning from past experiences and applying the insights in future situations (e.g., intelligent routing, optimised energy usage), identifying patterns and meanings behind qualitative and quantitative data (e.g., understanding human language, performing facial recognition), and recognising and acting on environment changes (e.g., autonomous driving). As such, it embraces a broad group of domains, including machine learning, natural language processing, computer vision, robotics and automation, connected and automated vehicles, AI processing units, and AI services. AI systems are not only beneficial for public administration systems but for companies as well. Hence, they typically benefit from higher productivity and efficiency. Productivity is generally achieved via better decision-making processes, whereas efficiency is typically achieved via automating manual processes. The EU therefore seeks to put AI at the service of European citizens and the economy. To this end, the main initiatives are to prepare for the socio-economic changes brought about by AI, establish an appropriate ethical and legal framework, stay ahead of technological developments, and encourage adoption by the public and private sectors, with the goal of achieving investments of EUR 20 billion per year over the next decade. Figure 1 provides an overview of the policy areas for AI.

Figure 1: **Overview of important policy areas for AI**



Source: Van Roy et al., 2021.

In addition to the areas elaborated on below, AI also aims to address two challenges facing society: climate change and the COVID-19 pandemic. As regards the situation in Croatia, the Croatian government is still working on its national AI strategy. Although the working group, made up of experts from academia, business, civil society, and the public sector, has completed an AI strategy, it has not yet been approved by the Croatian government and is thus not publicly available.

Croatia realises a considerable share of its economic growth precisely through tourism and related activities. The application of AI to achieve the best possible guest experience and consistent, predictive communication with the guest through all available communication channels is an important factor via which Croatia should present itself to the world as a space where traditional experiences and modern technologies are combined. The best example in this area is a private company called Acquaint that has developed an AI system exclusively for hotels. It is a virtual receptionist named "Alexa" that, taking enormous amounts of data into account, selects three specific recommendations for the guest from the 750 services available. It is currently being tested at five Jadranka Group hotels on the island of Lošinj – Bellevue and Alhambra, both five-star hotels, as well as Aurora, Vespera and Punta, each with four stars. AI also offers the opportunity to increase hotel revenues in a very special way. It requires very little investment and relies solely on guest spending, which means there is no need to expand accommodation capacity or raise prices as traditionally occurs to increase revenues.

Another important technology the EU wishes to promote is blockchain. The European Commission (2020) defines blockchain as "a technology that enables people and organisations to agree on and permanently record transactions and information in a transparent way without a central authority". As such, it is mainly associated with financial services and cryptocurrencies, yet its potential is much greater. It can cover various areas – payments and international transactions, copyright and intellectual property protection, digital identity and elections, digitisation of processes, tracking of goods, and supply chains. Therefore, an analysis of its strengths, weaknesses, opportunities and threats (SWOT) is presented in Section 5. (see Table 2).

Despite these two new technologies offering many opportunities, there are also a lot of challenges. The biggest challenges in adopting and implementing these technologies are legal and operational. There are also challenges like commercial interests, risk aversion, the high cost of new technologies, data quality, privacy and data security requirements, data localisation requirements, financial inclusion concerns, and others.

5 SWOT analysis of blockchain technology

The blockchain was created in 2008 as part of the digital transition. It has the following components: a shared record of transactions, consensus on their verification, operating rules and encryption. According to Pierluigi's study (2021), the first blockchain is a distributed ledger of participants in a network; the second blockchain uses peer-to-peer technology, which allows users to connect without a central authority or control point; while in the third blockchain transactions are grouped into blocks that are linked to the blocks before them, which allows the underlying transactions to be traced. It is a subset of the broader distributed ledger technology whose purpose is to record asset transactions and their details simultaneously in multiple locations. It is also based on three concepts: the distributed nature of the ledger, a consensus mechanism, and cryptographic hash functions and digital signatures. Three main types of blockchain exist: public (or open), permissive (or private) and hybrid. Blockchain has a wide range of applications today and enables low-cost transactions with digital assets. It also creates the opportunity to create a digital tax administration consisting of a standardised electronic form for filing tax returns, the real-time cross-checking of files for fraud prevention, and transparent third-party data and financial visibility. The SWOT analysis presented in Table 2 therefore provides a high-level assessment of the elements of blockchain technology in tax administration.

Table 2: **SWOT analysis of the blockchain technology in tax administration**

Strenghts	Weaknesses
Faster and transparent transactions Information and communications services are readily available and well documented Lower costs of fulfilling tax liabilities Direct connection with taxpayers, no third-party mediator Higher efficiency Inviolate privacy	Data security problems Lack of information and telecommunications regulatory basis Underdeveloped information and telecommunications infrastructure in the rural areas of a country Lack of financial inclusion Lack of public access to the Internet Lack of public presentation and citizen awareness with digital tools – low technology maturity High energy costs
Opportunities	**Strenghts**
Simplification of tax procedures and reduction of costs to taxpayers Improvement of compliance risk management system Business process optimisation Rapid growth of the ICT sector Broader application of information and telecommunications in business and public administration Education and motivation of users to adopt blockchain technology Improved customer experience	Reduce dependence on tax advisors Insufficient financial funds for modernisation Insufficient knowledge and skills of employees Digital identity fraud High investment costs for implementations Willingness to adopt

Source: Owens and Hodžić, 2023.

Implementing blockchain as one of the emerging technologies in tax administration or public administration is generally a huge challenge. An overall result is that digital tax administration or public administration is introduced or significantly improved. One strength of the implementation is the faster and transparent transactions already facilitated by Blockchain 5.0 (no more than 2 seconds to complete the process); information and communication services are readily available and well documented, and the direct connection with taxpayers without third-party intermediaries, which means that neither financial institutions nor a clearinghouse are involved. On the other hand, weaknesses lie in data security issues since there is no guarantee that all of the data will be secured.

Four factors are important for integrating blockchain technology into tax administration: data redundancy, information transparency, immutability of data, and a consensus mechanism. Pursuing the potential of blockchain technology as such is relevant to the specific tax categories listed in Table 3.

Table 3: **Potential of blockchain for specific tax categories**

CATEGORY	DOMESTIC TAX	INTERNATIONAL TAX
Reporting obligations of the same information to multiple tax authorities and agencies	Payroll tax	Transfer pricing and country-by-country reporting
Third-party reporting obligations	Withholding tax	Withholding tax DAC6
Transaction tax	Value-added tax Sales tax Tax on property transaction	Customs, tariffs
Information sharing among tax authorities	Among federal, state and local governments: State Audit Report Programme (SARP) State Reverse File Match Initiative (SRFMI) Municipal Agency Partnering Programme	Among multiple countries: Bilateral Tax Information Exchange Agreement (TIEA) Multilateral Tax Information Exchange Agreement Automatic Exchange of Information

Source: Owens and Hodžić, 2023.

Apart from the advantages, this type of technology requires extreme coordination between authorities, plus there is the problem of hesitant and expensive implementation. Still, the most noteworthy application concerns VAT. This is the most important tax in all EU countries as it brings in the largest revenue to the state budget. It is a consumption tax on goods and services levied in each stage of the supply chain. The potential benefits of blockchain technology in VAT transactions include the following (Deloitte, 2017):

1) the administrative burden of companies is significantly reduced, thus saving time and the cost of accounting services;

2) all transactions are conducted in real time;

3) all transactions executed by smart contracts are tamper-proof and transparent;

4) a reduced risk of fraud and mistakes;
5) immediate insight into a company's finances;
6) rapid money transfers between businesses and the government;
7) the burden on taxpayers of receiving the VAT amount calculations in an invoice and the VAT amount due in a tax return is eliminated; and
8) opportunities for VAT fraud are drastically lowered because the same system facilitates the processing of VAT from a transactional perspective and, at the very same time, multi-dimension checks and verifications of the transaction, parties of the transactions, as well as the legal and business context of the transaction can be ascertained.

Overall, the potential of blockchain technology for VAT lies in it enabling the recording of sellers and buyers in real time (VAT) and that, due to smart contracts, all transactions executed on the blockchain are tamper-proof and transparent, which also reduces the risk of fraud and errors. This led the EU to propose a blockchain solution, with the most effective being the blockchain application VATCoin (Ainsworth et al., 2018).

Apart from its huge potential in VAT, blockchain technology holds potential in international trade, from manufacturing to shipping and distribution to customs clearance. The main role of customs clearance is to monitor the flow of goods to ensure the legality of trade and detect any smuggling activities. For this reason, Dubai Customs has developed an innovative cross-border e-commerce platform based on blockchain technology. The advantages of this platform are (Owens and Hodžić, 2023):

1) to consolidate clearance and easily reconcile inventory by optimising information sharing;
2) increasing efficiency by eliminating declaration preparation time and reducing the cost of e-commerce transactions;
3) identifying and certifying e-commerce companies;
4) improving flexibility for companies engaged in e-commerce;
5) reduce physical document submissions for imports into the mainland from bonded zones; and
6) to provide 100% visibility and traceability on e-commerce transactions to all stakeholders.

Similar to the VAT application, this also allows information to be sent in real time.

In summary, the key benefits of blockchain technology for tax administration are trust, transparency, operational efficiency, the ability to tokenise assets and values in the future, i.e., converting the rights to an asset into a digital representation of that asset, or as another component of the digital economy, interoperability, and privacy by ensuring that only authorised parties access the data.

6 Policy recommendations for Croatia

Based on the strategic goals and the analysis, several recommendations can be derived with respect to development needs and potential in the area of digital skills development and digital jobs:

1) increasing the number of ICT experts in the labour market;
2) raising the level of digital competencies and retraining the workforce from non-IT occupations to meet the needs of the labour market;
3) boosting the level of basic and advanced digital competencies of citizens for active participation in the digital society;
4) assuring the further digital transition of the education sector and establishment of programmes for students interested in ICT topics;
5) redefining enrolment quotas in higher education with the aim of increasing the number of people with an ICT-related diploma;
6) redefining enrolment quotas in secondary education with the aim of better preparation for studying and successfully completing STEM studies;
7) increasing the number of teachers and spatial resources in higher education institutions that train ICT experts;
8) attracting more foreign students and experts in the ICT field through the internationalisation of higher education;
9) supporting the development and application of digital tools in education in order to ensure equal opportunities for education and the acquisition of digital competencies for all citizens; and
10) encouraging the greater representation of women among ICT experts.

All of these recommendations are also in harmony with the European Commission’s guidelines, i.e., the Digital Compass 2030, which sets the following targets: at least 80% of all adults should have basic digital skills, 20 million ICT professionals should be employed in the EU, and the number of women among ICT professionals should be increased.

With regard to the situation in Croatia concerning the development of digital competencies and emerging technologies, several strengths are recognised, including:

1) the education sector has shown considerable readiness to implement digital technologies and acquire new digital competencies through the e-School project;
2) an advanced labour market analytics monitoring system has been established;
3) Croatia is an extremely desirable and safe place to live, namely, an important factor for attracting foreign ICT experts and digital nomads;
4) in April 2022, the voucher financial instrument for adult education was established; and
5) the national regulatory and financial framework encourages the development of digital education.

Nevertheless, there are some weaknesses, such as:

1) the insufficient number of ICT experts in the labour market considering the needs and possibilities of company growth;
2) the absence of a national plan for internationalisation, and for attracting and retaining digital professionals and talents;
3) the high tax burden on work and taxes on individual incomes of valuable employees;
4) the passive resistance to the principles of open education, new methods and techniques teaching, and challenging or misinterpreting the possibilities that e-learning methods and the application of new technologies bring improvements in the field the quality of education and continuous, flexible acquisition of new competencies, including digital ones; and
5) the insufficient data concerning the need for public administration employees to raise the level of their digital competencies.

To promote the faster application and construction of AI-based solutions, it is necessary to provide widely available resources with capabilities and provide platforms for the application of AI. It is further necessary for these resources to be cost-effective, available and to provide considerable capacity for computer processing and data storage. The government must accordingly create a framework in which the use of available cloud resources is encouraged by enacting regulations that enable the transfer and processing of data with such available resources. In addition, to facilitate access to government-managed data it is recommended that data be consolidated and managed from a single data management centre, typically part of a shared services centre operated by the state.

To support the creation of knowledge and skills in the field of AI, centres of excellence and competitiveness in the field of AI must be established using available EU funding frameworks and partial investments from local sources. The centres of excellence must be supported by the joint cooperation of the private, public and academic sectors. Therefore, it is expected that the government can support the setting up of at least one centre of excellence focused on applying AI in the public sector and the required number of centres of excellence and competitiveness for applying AI in particular sectors and industries.

7 Conclusion

AI is now present in various technologies and areas of human activity. The realisation that artificial intelligence does not really have to be intelligent in the general sense, but intelligent enough to solve problems, has led to a major shift from academic research to the application of techniques developed in AI. Accordingly, there are truly numerous examples of the use of intelligent systems today. At the same time, there are weak and strong AI systems. An example of weak AI is the simulation of intelligence (e.g., speech recognition), while strong AI is found in the properties of human intelligence.

Since joining the EU, the government of the Republic of Croatia has continuously increased its efforts towards the digital transformation and the provision of public electronic services to citizens and business. In all of the country's relevant strategic planning acts, digitalisation is highlighted as a priority, with one of the strategic goals of the National Development Strategy until 2030 being the "digital transformation of society and the economy". The digitisation of public administration is recognised as a priority of the Government Programme 2020–2024 and included in Objective 4.1. An efficient, transparent and resilient state, while in the National Plan for Reconstruction and Resilience 2021–2026 in subcomponent C2.3. Digital transformation of society and public administration.

Croatia should become a country with digitally and economically competitive businesses and a digitised public administration by 2032, and it is important that all levels of government and citizens are actively involved in digital processes. AI holds great potential for the development of society, the development of innovation, and the development of public services, yet it also raises new questions of responsibility, ethical use, and legal opportunities and frameworks. AI stakeholders in Croatia are also participating in the development of standards and frameworks that seek to establish the understanding and scope of responsible use of AI. The development and use of solutions based on AI must have an effective legal framework that, among others, harmonises the use of AI with the existing legal frameworks of the EU and the Republic of Croatia. The government must create and continuously harmonise such a framework, not only to protect citizens and organisations, but also to create a framework that enables the smooth development and growth of the business sector and AI solutions. Unhindered development means understanding the opportunities, but also creating new frameworks in which the Republic of Croatia can be competitive, attract investments and attract new development companies – startups working in the field of AI.

The digital transformation is a complex process that requires successful planning and implementation to ensure sufficient tangible and intangible resources. The application of AI in business implies the digital transformation and the concept of Industry 4.0. These topics are strongly interconnected, and their common goal in business is, of course, to improve business processes and business results. The application and implementation of AI in business improves the large amounts of data and information that become available. This includes structured data, such as data collected by various sensors and analytics systems, or unstructured data like data from cameras, social media and networks etc. This growing amount of available data is a key driver for the increased use of AI in business. Although there is greater use of AI in business, it is not yet suitable for all business processes. While many business processes and aspects can be automated, tasks that require judgement, prioritisation, compromise etc. require human intelligence. In the future, the impact on the economy will be even greater as AI will help increase production and sales, and thereby GDP. However, it is believed that society is still at a relatively early stage of AI adoption. Only a small number of companies have adopted a wide range of AI technologies and applications. Nonetheless, the situation is improving year by year. The potential of AI has been acknowledged by many countries, which have made initial investments and legislative changes.

REFERENCES

- Ainsworth, R. T., Alwohaibi, M., Cheetham, M., & Tirand, C. (2018). A VATCoin Solution to MTIC Fraud: Past Efforts, Present Technology, And the EU's 2017 Proposal, Tax Notes, 335. Retrieved 25 July 2023 from: https://scholarship.law.bu.edu/faculty_scholarship/1402
- Croatian Parliament. (2023). Official Gazette No. 2. Digital Croatia Strategy for the period until 2032.
- Croatian Parliament. (2021). Official Gazette No. 75. The Regulation on office operations.
- Deloitte. (2017). Blockchain technology and its potential in taxes. Retrieved 20 July 2023 from: https://theblockchaintest.com/uploads/resources/Deloitte%20-%20Blockchain%20Technology%20and%20its%20potential%20in%20Taxes%20-%202017%20-%20Dec.pdf
- European Commission.. (2022). Europe's Digital Progress (DESI) Report 2022 – Croatia. Retrieved 25 July 2023 from: https://digital-strategy.ec.europa.eu/hr/policies/desi-croatia
- European Commission. (2021). eGovernment Benchmark 2021 – Entering a New Digital Government Era. Luxembourg: Publications Office of the European Union.
- European Commission. (2020). Shaping Europe's digital future. Luxembourg: Publications Office of the European Union.
- European Commission. (2019). A definition of Artificial Intelligence: Main capabilities and scientific disciplines, Brussels.
- European Commission. (2018). Declaration of Cooperation on Artificial Intelligence. Brussels.
- Hodžić, S., Ravšelj, D., & Jurlina Alibegović, D. (2021). E-Government effectiveness and efficiency in EU-28 and COVID-19. Central European public administration review, 19, 159 – 180.
- Ministry of Public Administration. (2020). eGovernment Strategy 2020 (e-Croatia 2020). Retrieved 20 July 2023 from: https://rdd.gov.hr/UserDocsImages//MURH_migracija%20s%20weba//e-Croatia%202020%20Strategy%20-final.pdf
- Owens, J. & Hodžić, S. (2023). Blockchain Technology: Potential for Digital Tax Administration. Intertax, 50(11), 813-823.
- Pierluigi, M. (2021). Blockchain and Banking - How Technological Innovations Are Shaping the Banking Industry. Palgrave Pivot Cham.
- Van Roy, V., Rossetti, F., Perset, K., Galindo-Romero, L. (2021). AI Watch - National strategies on Artificial Intelligence: A European perspective, Luxembourg: Publications Office of the European Union.

Chapter 7

Back to the future: Can Bulgaria become Europe's AI hub?

Fabio Ashtar Telarico
University of Ljubljana,
Slovenia

1 Introduction

Over the last three centuries, humanity's technological development has proven time and again that the future often lies well beyond the boundaries of what seems plausible today. The lightning-fast spread of artificial intelligence (AI), which promises to be the next big thing in economics and beyond, may thus be no different. Changes on this scale often transform the world economy and entire countries within it, in a matter of years realising changes that previously had been advancing at a glacial pace. From outside the 'tunnel of plausibility' and the limitation it imposes on future-oriented policymaking, Bulgaria's successful adoption of AI technology feels less like a leap towards a new age and more like a skid back to the future. Namely, after the Soviet Union, Bulgaria was the most advanced of the European communist states and used to produce up to half of all the electronics made in the communist bloc (Petrov, 2023, p. 105). Indeed, according to

several analysts and academics, Bulgaria could become the continent's AI hub despite its peripheral position in the European Union's (EU) system of knowledge production (Jehlička, 2021). Even Bulgarian decision-makers seem to be aware of this, as the current National AI Strategy indicates (MS-RB, 2020).

Still, given its shrinking population and small economy, Bulgaria's ability to partake in the global debate on the use and abuse of AI is extremely limited and it must rely on the EU. Indeed, the Union has already taken a vocal stance on almost everything related to AI. These policies can however be both an enabler and an inhibitor for countries that failed to reap the benefits of previous waves of digitalisation and especially for Bulgaria, where the right policies could bring big gains from AI. In fact, the idea of letting the Union take on the burden of regulating AI is especially attractive for those smaller/poorer countries that lack the capability to implement their own rules. Yet, by adopting the 'maximum harmonisation' principle, the Act may stave off legitimate aspirations to set up national AI policies. Given the enduring inequalities amongst EU member states, this choice may end up penalising the Union's peripheral members.

Against this background, this policy paper recommends that Bulgarian and EU policymakers amend the proposed supranational policies in the light of a new guiding principle: guaranteeing EU member states' policy space, particularly when there are noteworthy national and local specificities that can be valorised to the benefit of all member states like in Bulgaria's case. Concretely, following the examples set by the United States and Japan, it makes four policy recommendations:

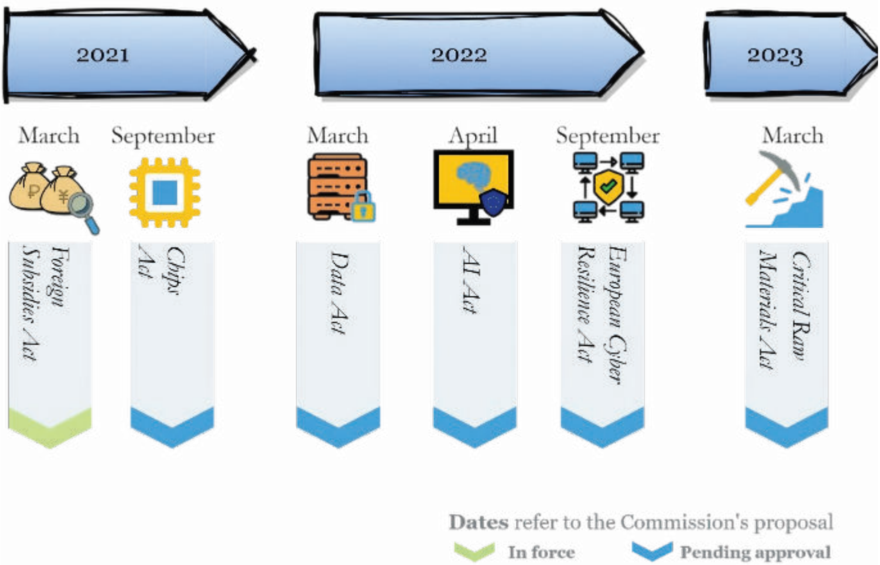
- **harmonising permissions and project delivery** for high-tech manufacturing and development;
- a **Union-wide semiconductor research agenda** imposing a wider scope of cooperation;
- **harmonised workforce training** to reduce spatial inequalities; and
- **allowing state-sponsored takeovers and investments** under intergovernmental coordination.

These recommendations (detailed in Section 4) are based on analysis of the EU's AI Act from the perspective of Bulgarian AI policy and the possible interactions between them (Section 2). Moreover, they consider the ongoing private and public initiatives to support the development of AI in the country and the key differences between the Bulgarian approach to the construction of an AI ecosystem and the prevailing practice in the most advanced EU countries (in Section 3).

2 A view from the periphery on the EU's digital and AI strategy

Much attention has been paid to the EU's AI Act, or more officially the 'Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (the Artificial Intelligence Act) and Amending Certain Union Legislative Acts.' The regulation's stated aims are to: create a level playing field, protect users, and ensure that the development of AI respects liberal values. However, the AI Act cannot be fully understood outside of the overarching framework of the EU's industrial strategy for the digital transformation of its member states' economy. This New Industrial Strategy for Europe (EU Commission, 2020a) is manifested in several other and initiatives on the EU level including, besides the AI Act, policies related to foreign capital (Foreign Subsidy Regulation, 2021), data management (Data Act, 2022), cyber security (Cyber Resilience Act, 2022), semiconductors (Chips Act, 2022), and more (Figure 1). Further, these areas are so deeply entangled that each policy's very ability to affect economic and political outcomes depends on full implementation of the other. For instance, to become a key player in the global race to the 'smartest' AI, the EU needs to ensure that local AI laboratories can thrive and recoup international competitors' head start while being able to protect their intellectual property. Then, the AI Act cannot really function without the consistent application of the Foreign Subsidy Regulation, especially vis-à-vis competitors like China and, to a lesser extent, Russia. Moreover, the European AI ecosystem/s ought to be shielded from possible acts of cyberwarfare in which Russia and China are much better versed than any EU country and, possibly, even the USA. The AI Act thus relies on the protection offered by the Cyber Resilience Act. Practically, the models underlying AI need expensive and sophisticated processors to 'learn' new information (in jargon, machine learning or model training) and generate content. Hence, the AI Act also depends on the reliable production of semiconductors that the Chips Act attempts to guarantee. Yet, those chips require rare earths and other commodities that EU countries import from declared competitors in the AI contest like China and Russia or unfriendly countries in Africa and Asia. This means that, from a supply-chain perspective, the Chips Act will remain a resounding, yet empty statement of little material impact if the Critical Raw Materials Act cannot safeguard imports of lithium, cobalt, coltan and other commodities (regarding this interaction, also see Timmers, 2022, pp. 26–33).

Figure 1 Infographic of the main EU legislative acts affecting the development of AI



Dates refer to the Commission's proposal

In light the closely woven nature of these policies, it is reasonable to expect that the AI Act's undertone and design will not vary substantively from those of the previous legislation on the digital economy. Namely, given the content of other segments of the EU's digital strategy, the AI Act suffers from a number of criticalities that ought to be addressed. This is especially true from the perspective of countries that hope to boost their domestic AI development like Bulgaria. Crucially, two set of aspects must be considered. First, the impending risk that the AI Act will be biased in favour of wealthier EU member states similarly to other items of legislation implementing the EU's digital strategy. Second, the AI Act's systematic compression of any and all space for national policy to foster and boost AI development.

2.1 Mistakes already made: The Chips Act and the Foreign Subsidies Act

Considering the AI Act in the broader framework of the EU's digital strategy, it might be surprising that these policies systematically overlook the existing economic inequalities amongst member states in high-tech sectors. Yet, economists and policy experts have been paying attention to the risk of worsening economic imbalances in the Union. For instance, the Brussels-based economic think tank Bruegel argued that "poorer EU countries risk being left behind" in the allocation of the Chips Act's promised subsidies (García-Herrero & Poitiers, 2023). Moreover, calls to strengthen peripheral states' capabilities in this area have been softly spoken (see the EU's rapporteur on the Chips Act: Nica, 2023, para. 21 on p. 18). Wealthier countries in the meanwhile are already cashing in: the largest European contract manufacturer and designer of semiconductors, STM, is to build a EUR 5.7 billion plant in France (Fleming, 2023) and the US chip behemoth Intel will invest EUR 17 billion in Germany (Intel, 2022).

Similarly, the Union is implementing stricter regulation of third countries' investment in EU countries, notably in the high-tech sector. The idea dates back to a 2019 white paper that addresses the potential distortionary effects of foreign mergers and acquisitions (M&A) investment in the EU Single Market (EU Commission, 2020b). Ostensibly, this move was aimed at reducing the risk that potentially hostile third countries could single-handedly take over EU firms and their intellectual property or stymie their development through unfair competition in the EU (Tilman Kuhn et al., 2022). However, in reality, the ensuing legislation (Foreign Subsidy Regulation, 2021) has had two other effects. First, it has constrained the already very limited foreign direct investment (FDI) going to the EU's periphery, most of which was originating in China. Second, it has helped wealthier countries shield their national champions from foreign acquisitions and competition while channelling the still-needed foreign capital into non-M&A investment in the most advanced economies. The data seem to corroborate this interpretation. In fact, the regulation has incentivised foreign investors to steer away from M&A and start "pouring" venture capital "into European tech start-ups" in Germany, France and the UK (Kratz et al., 2022, p. 3). As a result, the majority of incoming FDI in strategic and high-tech sectors in the EU mostly originates in Germany, France and, despite Brexit, the UK.

2.2 Conflating floor with ceiling: No room for national AI policies

Materially, the AI Act challenges policymakers in several ways as they seek to upscale their domestic AI ecosystems. The first paragraph of this section discusses the risks stemming from the Act's attempt to bring about the across-the-board, maximum total harmonisation of AI, wilfully ignoring national specifics. The second paragraph then delves into a few other controversial clauses that may render it very difficult for peripheral countries that failed to participate in the previous digital revolution to catch up on AI.

2.2.1 Total maximum harmonisation

Even though the debate on this issue remains open, it seems that the AI Act adheres to the 'maximum harmonisation' principle more strongly than previous regulation of the digital economy. Thus, the AI Act's pre-emptive effect "could have far-reaching consequences" for the development of national AI ecosystems (Veale & Borgesius, 2021, p. 108). Theoretically, a maximum harmonisation regulation "conflates the floor and ceiling" in setting a threshold for national policy "leaving the Member States with no room for manoeuvre" (Mańko, 2015, p. 19). Nevertheless, EU regulations usually only adopt this approach for a minority of their provisions. Most pieces of maximum-harmonisation legislation were hence instances of partial maximum harmonisation. Further, all aspects not covered by maximum harmonisation were left to minimum harmonisation, with the EU setting a floor that national regulators ought to reach, but otherwise does not bind them.

By contrast, the AI Act contains some provisions on high-risk AI systems and virtually no sort of regulation on all other applications. It is notable that only the former are "without prejudice to other user obligations under Union or national law" (AI Act, 2021, Art. 29, para. 2). The bulk of the AI Act would therefore implement a sort of "total maximum harmonisation" (Veale & Borgesius, 2021, p. 108 [emphasis added]) rarely seen before. In this case, member states would be unable to alter the regulatory incentive structure for domestic AI development in both the high-risk areas regulated by the Act and those the Act fails to regulate.

2.2.2 One size must fit all: The risk of failing to ensure equal opportunities

Turning now to clauses that could obstruct the development of healthy AI ecosystems in a country like Bulgaria, three broad factors should be mentioned: (a) the structure and functions of the national supervisory authorities; (b) the hinderances to their operation in smaller countries; and (c) the compromise on liberal values.

First, although national supervisory authorities are given an important role, their regulation is insufficient while also preventing national policymakers from improving it. In short, the Act foresees the designation of a national market supervisory authority (MSA) tasked with implementing AI rules. Surely, this choice is contestable in and by itself. For instance, such agencies are not necessarily independent of the government (Veale & Borgesius, 2021, p. 111). In addition, the very decision to set up an MSA may perversely incentivise underfunded and understaffed agencies to overlook user well-being. MSAs are in fact not bound to follow up on users' complaints (Market Surveillance

Regulation, 2019, Art. 11, paras. 3–7). Yet the criticalities are becoming even more visible from the perspective of poorer countries like Bulgaria. In fact, the regulation mandates that staffers' "expertise shall include an in-depth understanding of artificial intelligence technologies" (AI Act, 2021, Art. 73, para. 3). However, it will be very difficult to actually find people with such competencies already on the public sector's payroll and engaging external experts may prove extremely expensive. Further, like any other regulatory agency, national AI regulators will be vulnerable to regulatory capture, especially when the imbalance between the economic prowess of overseeing states and overseen business is evidently in favour of the latter, like in cases in Bulgaria. Finally, several EU countries' supposedly 'independent' regulators have an abysmal track record when it comes to efficiently overseeing market activities. In the case of Bulgaria, the Blue Blink think tank exposed notorious cases of regulatory failure and, arguably, the capture of the tobacco (Barova et al., 2019) and legacy-media (Telarico, 2021a, pp. 22–25) regulators. Still, more 'technological' agencies like the Council for Electronic Media fare no better: at least in one case, a regulator threatened a journalist (Neykova, 2017) and the Council menaced the independence of the public broadcaster vis-à-vis the government (Dimitrova & Viktorov, 2014) and the church (Hussein, 2020).

Several EU countries' supposedly 'independent' regulators have an abysmal track record when it comes to efficiently overseeing market activities.

Second, the AI Act ends up creating hinderances to the MSA's compliance assessment. Maintaining that "translation costs related to mandatory documentation and communication with authorities may constitute a significant cost for [AI] providers [...] of a smaller scale", the Act requires that member states accept documentation in a language "which is broadly understood by the largest possible number of cross-border users" (AI Act, 2021, Art. 73 on p. 35) Essentially, the very premise of this argument is misleading. In fact, most AI services are ultimately connected to the Microsoft-Alphabet-Meta "AI oligopoly" (Wallach, 2022). And yet, this false premise means that MSAs will have to accept documents in English and bear the translation costs themselves. Evidently, these provisions will disproportionately affect countries where English is not widely known and MSAs with smaller budgets. These effects are compounded by the decision to assign MSAs an impressive remit that includes the duty "to look for synthetic content on social networks, assess manipulative digital practices of any professional user, and scrutinise the functioning of the digital welfare state" (Veale & Borgesius, 2021, p. 111).

Third, according to a study commissioned by the EU parliament, the AI Act would permit more favourable treatment of private corporations as AI users than public authorities in the same role (Georgieva et al., 2022, p. 11ff). Curiously, this distinction is not known in earlier digital regulations such as the GDPR. It is thus unclear how exactly it will affect AI. However, it seems to further tilt the balance in favour of AI providers to the detriment of MSAs and national policymakers. In actuality, while the text bans police forces from utilising real-time biometric identification systems in public spaces, the private sector still has access to this technology. The prohibition on social scoring also applies solely to public authorities and does not cover the private sector. Finally, the Act would only protect predefined categories (children, elderly, physically/mentally disabled) from specific forms of harm (physical and psychological) intentionally caused. Overall, this means that countries where liberal values still struggle to affirm themselves also due to digital illiteracy, such as Bulgaria (Telarico, 2021b), could see a worsening trend. Worse still, national policymakers are legally barred from amending the AI Act's rules in any way

The AI Act would permit more favourable treatment of private corporations as AI users than public authorities.

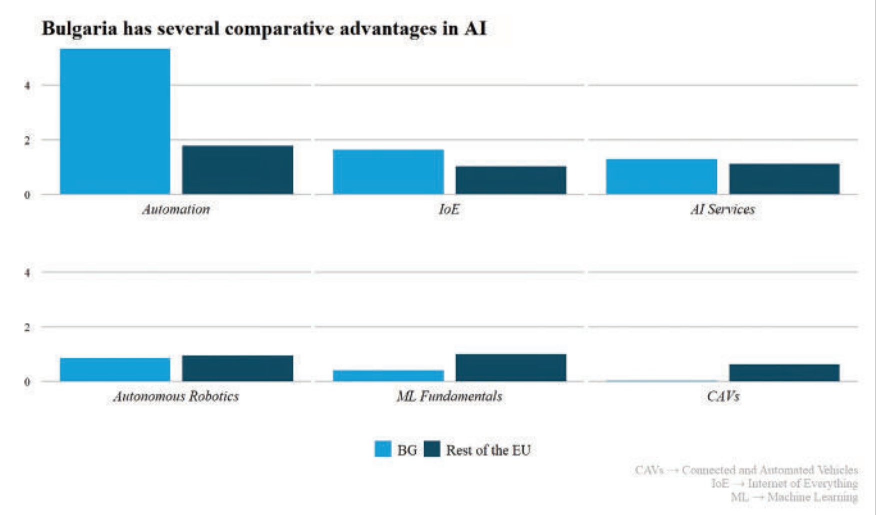
to accommodate their national specificities and assure stronger protection of the already dwindling liberal values. Such concerns are so pressing that even the Bundesrat, the upper house of Germany's parliament, raised them in its advisory opinion to the EU parliament. Namely, the Bundesrat opinion (2021, para. 53 on p. 19) demanded the introduction of "opening clauses that allow additional obligations, exceptions or derogations from the proposed regulation insofar as they are necessary to safeguard media pluralism". Given the level of monopolisation of the Bulgarian media landscape, such a clause would answer to a real need.

3 Just a matter of getting back to the future: Bulgaria's AI ecosystem

Although a little known fact abroad, Bulgaria has such a noteworthy industrial tradition in the fields of informatics and electronics that a professor at the University of Tennessee nicknamed it "Balkan Cyberia" (Petrov, 2023). This means it should not be surprising that recent studies on the AI economy in the EU ran contrary to the stereotype that depicts post-socialist, Eastern European countries as backward and fundamentally unable to innovate (Lopez-Cobo & De Prato, 2022). Instead, they draw the picture of a country that is well placed to contribute massively to the advancement of AI thanks to its well-trained workforce, top-notch researchers and solid tradition in the field of informatics. This is the reason for Bulgaria's enormous edge over the rest of the EU in the field of automation, and for its significant lead in the Internet of Everything and AI services (Figure 2).

However, the very fact that (almost) no one seems to be aware of this suggests that traditions and human capital are not sufficient to thrive in the age of computerisation, as it was once fashionable to call it. Rather, the key to Bulgaria's early giant leaps in informatics and more recent advancements lies in the construction of solid networks through which information, experience and knowledge can travel across borders. There is however a key difference between these two periods, as explained in the following paragraphs. Unarguably, in the 1980s the core of the network was the Bulgarian Communist Party (BCP), whereas, innovation and knowledge production are nowadays the result of the spontaneous, intrapreneurial or academic action of individuals who leverage their personal and professional networks to allow small segments of Bulgaria's workforce to express their potential.

Figure 2 Bulgaria and the rest of the EU's relative comparative advantages in key AI sectors



CAVs → Connected and Automated Vehicles
IoE → Internet of Everything
ML → Machine Learning

Source: Author's elaboration based on data from Righi et al., 2022.

3.1 Why Bulgaria's past may be the foundation for its future

Even though no one seems to remember it these days, it was such a well-known fact at the time that even the popular German magazine Der Spiegel (1982) informed its readers that "[m]ore than 70 per cent of advanced electronics in the entire Eastern Bloc come from Bulgaria". This figure was probably sensationalistic. In comparison, the Bulgarian Communist Party's (BCP) official estimates, known for being too rosy, put the figure at 45% in 1985 (Petrov, 2023, p. 105).

Notably, the BCP's protracted investment in its intelligence apparatus also played a vital role in Bulgaria's long-gone glory in the field of advanced electronics. After all, the first mass-produced Bulgarian computer, the Pravetz-82, was designed in 1982, imitating an Apple II that the Bulgarian secret service had manged to smuggle in from the West. Indeed, copying the hardware of these devices was not a mission impossible, and actually there were hundreds of Apple II clones across the West and non-aligned countries like Yugoslavia (Caruso, 1984; Bošnjak, 2021). Meanwhile, the USSR had the Agat-9 (Kaspersky Lab, 2014) and socialist Romania managed to put out a copy of the British ZX spectrum (Petrescu et al., 2012). Yet Bulgarian informatics earned an enviable reputation thanks to the ability of local software engineers. Somewhat anecdotally, Bulgarian PCs set themselves apart from others in the 'cloning' business after the 1985 International Symposium on Robotics in London. On that occasion, the relatively cheap IMKO-2 was used to control a robot arm (nicknamed ROBKO-01) through such a simple software interface that "even for specialists from the USA and Japan" that setup was "extremely impressive" (Pravetz Computers, 2017).

Nonetheless, the real explanation for Bulgaria's success story is a much less intriguing, bread-and-butter policy issue: education. Over 6,300 students had graduated from higher education courses in automation and informatics already in 1969–1971; by 1990, Bulgarian schools accommodated over 3,000 PCs; taught Basic, Logo, Pascal and other programming languages; organised lessons such as "Introduction to Cybernetics" and "Automation of Production" as part of the mandatory study of informatics (Petrov, 2023, pp. 122, 232). Eventually, with the fall of the Soviet Union and the end of one-party rule in Bulgaria, those young people who the BCP had drawn into the previously unimaginable world of hardware designing and software engineering witnessed the end of state-sponsored informatics. Left without subsidies and exposed to international competition after having lost access to Eastern Bloc buyers while facing a shrinking and impoverishing domestic market, the Bulgarian computer industry ceased to function properly. Some, who had an intellectual and practical predisposition for Schumpeterian creative destruction, went on to become "outstanding tech entrepreneurs" of the "Pravetz generation" (Fiscutean, 2016). Intuitively, this was a small minority of people with "the abilities and connections to remain involved in business and politics well after the end of the regime, sometimes

Bulgaria's digital economy represents over 7% of the country's gross domestic product (GDP), this being the second highest figure in the EU after Malta.

even increasing their power" (Petrov, 2023, p. 298). Others channelled their immense talent into much less "reputable" uses, eventually turning Bulgaria into the "biggest creator and distributor of computer viruses" in the early 1990s (Editors of Kompyutar za Vas, 1990, p. 1). According to experts, these were probably the most talented and idealistic members of their generation, "who had been 'on board with the [BCP] regime's dream' of a society where tiring and tedious tasks would have been automated to the benefit of all workers" (Petrov, 2023, p. 300).

Still, most of the 'Pravetz generation' moved on without having many opportunities to use, play with, or work on an up-to-date PC for years due to the deep economic crisis that marked the 1990s. Ultimately, the waste of so much talent and material progress was "a consequence of having developed a generation of young Bulgarians whose programming skills found few outlets" (Sudetic, 1990, p. 9) other than revengeful action against a type of capitalism that had wrecked their aspirations and whose minds were occupied by too many daily worries to keep on dreaming of a new world.

3.2 Bulgaria's present: A sprawling ICT sector and world-class AI research

Even though for many the splendours of the past are little more than a treasured memory, Bulgaria is still home to a small, healthy high-tech environment. On one hand, its information and communication technology (ICT) firms have experienced fast growth since the country joined the EU in 2006, and the tech park in Sofia hosts some of the world's most renowned software firms. On the other, research and development are greatly underfunded, especially when it comes to public institutions. However, the establishment of the Institute for Computer Science, Artificial Intelligence and Technology (INSAIT) in Sofia may mean that the light at the end of tunnel is finally in sight.

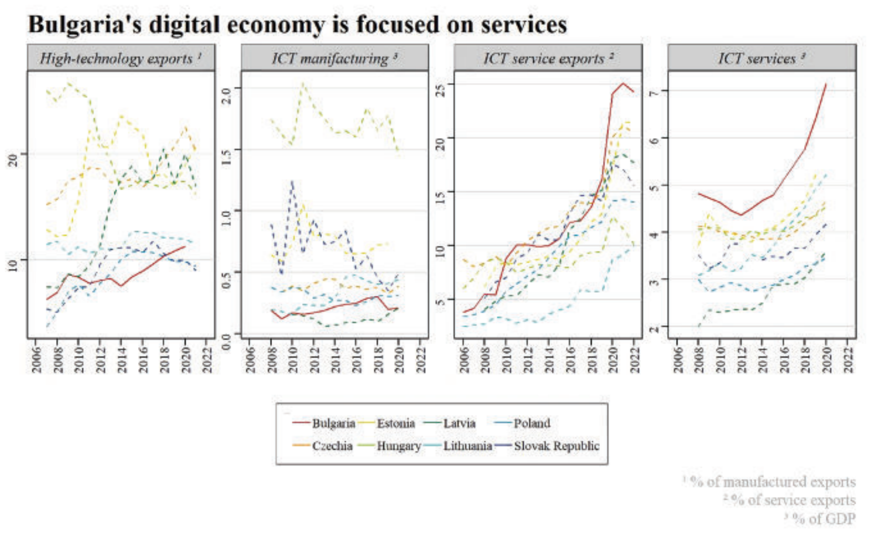
In a nutshell, while Bulgaria has solid fundamentals in AI, a sustainable AI ecosystem has yet to come.

3.2.1 A de-industrialising export-oriented ICT sector

Bulgaria's digital economy represents over 7% of the country's gross domestic product (GDP), this being the second highest figure in the EU after Malta (EUROSTAT, 2023). Yet, the country's ICT sector has shifted its focus over time. Instead of producing cutting-edge hardware, Bulgarian private firms tend to provide digital services domestically and internationally. Indeed, in this field Bulgaria is a leader amongst Eastern European EU member states (Figure 3). The productivity-enhancing possibilities of AI are, moreover, incentivising the sector-wide adoption of AI, already with "entire firms being entirely powered by AI in Bulgaria" (Nakov, 2023).

Still, this means the country's ICT sector is very exposed to two complementary risks. First, becoming a 'local branch' of multinational behemoths integrating hardware and software that can exploit Bulgaria's low wages to near-shore, low-paid tech service jobs. Second, a large chunk of the national economy relies on access to commodities and high-end manufacturing capabilities hosted by potentially unfriendly countries without which most ICT services could not be provided. Further, the demise of high-tech manufacturing is worsening the state of abandonment into which Bulgaria's industrial apparatus slipped during the 1990s. That would be unfortunate because the Chips Act suggests that the EU desperately needs the sort of cutting-edge manufacturing capabilities that Pravetz commanded in the 1980s.

Figure 3 Production of digital and tech-related services and products in Bulgaria and Eastern Europe



Source: Author's elaboration based on data from WB, 2022a, 2022b; EUROSTAT, 2023.

3.2.2 A source of fresh insights: INSAIT and AI research in Bulgaria

In April 2022, the Bulgarian capital Sofia's tech park witnessed the long-awaited inauguration of INSAIT, the first research institute exclusively focused on advanced research in computer science and AI. The institute was established by the University of Sofia in collaboration with ETH Zurich and Lausanne's Ecole Polytechnique Federal (EPFL), two of the best technical universities in Europe (INSAIT, 2022).

Notably, the initiator and driving force behind INSAIT is Martin Velchev, a Bulgarian expat currently employed at ETH and the first Bulgarian national to win an individual ERC grant, which are awarded only to world-changing, top-notch researchers in the world (Vechev et al., 2023). Admittedly, the almost untapped reservoir of fresh talent and consolidated expertise that the country has to offer in the realm of technology is one of the reasons why INSAIT was set up, of all places, in Sofia. However, INSAIT is also attempting to attract world-class researchers to Bulgaria by offering access to ETH's resources with little or no strings attached. Until now, these efforts seem to have borne fruit: world experts in machine learning and cybersecurity from the Massachusetts's Institute of Technology, ETH, EPFL, Yale, and Google Labs have already joined the staff (INSAIT, 2023a). Further, INSAIT offers a fully-financed PhD programme taught by professors who, in total, won 11 ERC grants and founded 9 deep-tech startups (INSAIT, 2023b). Potentially, INSAIT could harness ETH's embeddedness in the tech industry to attract and retain talent in loco while fostering the development of a more hands-down, production-capable AI ecosystem in Bulgaria.

Indeed, as hopeful as the very existence of INSAIT may be, it is a far cry from resolving the structural issues affecting Bulgaria's ICT sector. As a matter of fact, insofar as it is a private initiative, INSAIT's scholarship will favour the best and brightest regardless of the structural spatial inequality of opportunities amongst Bulgarian students: two-thirds of the students selected for the fully financed summer internship come from the capital, and none from provincial towns (INSAIT, 2023c). The aspiration to stop the brain drain obsessively repeated like a mantra by several top INSAIT staffers (Van Gool, 2022; Vechev et al., 2023) will also likely remain unrealised. After all, INSAIT is essentially a spin-off of the machine-learning research centre at ETH Zurich, which already hosts a number of Bulgarian-born experts in AI. There is hence a serious chance that INSAIT will mostly operate as a conduit of fresh ideas and young talents from Bulgaria to Switzerland rather than helping the former retain its human capital. Finally, INSAIT is perpetuating the enormous wage gap between researchers on the periphery and in the centre of the EU. Practically, doctorands will receive stipends of BGN 3–6,000 (EUR 1,500–€3,000), which is an excellent salary relative to the country's cost of living, but well below the average minimum PhD stipend in Europe (see data in Ahmadi, 2022). This makes it very unlikely that INSAIT will be able to attract international students to Bulgaria, except occasionally. It is also not a direct competitor of world-class institutions, more a 'branch'.

4 Key policies and targets to get Bulgaria back to the future

Taken together, the attempt at identifying the foreseeable effects of the EU's AI policy and the survey of AI research and the economy in Bulgaria paint quite a worrying picture. On one side, many of the brightest minds of the Pravetz generation, who had embraced the BCP's dream of emancipating automation and felt betrayed by shock-therapy capitalism, were either forced to take on unskilled jobs to earn a living or dedicated themselves to piracy. Simultaneously, the private sector that emerged from the post-socialist transition of the 1990s and 2000s benefitted from the immense stock of human capital produced by the real-socialist regime. However, its short-sighted maximisation of present value and profits led to the almost complete dismantlement of the country's once flourishing high-tech manufacturing. On the other side, the EU lacks a serious industrial policy, including with respect to electronics and AI, because "it drunk the neoliberal Kool-Aid and took it too seriously" (Tooze & Abadi, 2023). In fact, Bulgaria's Pravetz is not the only strategically important European tech company to have suffered upon being exposed to the global market's competitive pressure following deregulation: Philips, Alcatel, Nokia, Ericsson, Siemens, and many others also appear on this list (Gobble, 2014).

In the case of AI and related technologies, the EU is especially vulnerable due to very long and potentially unreliable supply chains and the lack of domestic production and development capabilities. This is accordingly an area in which the Union and its member states could reap enormous benefits simply by ditching some of the economic orthodoxy guiding its current economic policies. Obviously, this does not mean embracing the sort of totalitarian, statist policies that allow Chinese companies to push international competitors out of the market. Instead, it means following the path traced by the USA where policymakers were courageous enough to scrap economic dogmas that did not suit the country's development and are now being repaid with higher growth trajectories than seen in the EU (Rachman, 2023). Along this path, European and national policymakers should be guided by the core values of the liberal tradition that lies at the heart of the European common house: solidarity, equality of opportunities, and the valorisation of local peculiarities.

In this sense, national policymakers should pressure their colleagues in Brussels to rewrite the AI Act and amend the proposed Chips Act to introduce a new guiding principle and implement two reforms along two pillars. Intuitively, the main reference point for the EU should be the industrial and high-tech policy of its closest ally, the USA (and especially the Chips and Science Act, 2022). Clearly, EU policymaking does not always enjoy the scope and, sometimes, the resources that the federal government in Washington commands. However, much can be achieved by combining EU lawmaking power and peer-to-peer intergovernmental cooperation, particularly once the principle of total maximum harmonisation is replaced by a guarantee of sufficient national-policy space AI regulation. At least, insofar as the ability to diverge in regulating AI use and development does not compromise the single market's cohesion.

Consistent with this principle, national and EU policymakers should adopt a new high-tech policy paradigm and work towards the following substantial policy changes.

Harmonising permissions and project-delivery for high-tech manufacturing and development | Indeed, some of the key issues that the EU's Chips Act should address similarly to its US counterpart are strictly regulatory and, thus, relatively straightforward. For instance, the regulation should establish harmonised permit requirements for high-tech manufacturing. Ideally, cooperation between governmental and private actors will allow best practices in the field of permissions and project-delivery to be diffused, ultimately supporting projects boosting the EU's AI capabilities.

A Union-wide semiconductor research agenda imposing a wider scope of cooperation | In addition, the available research funding scheme Horizon Europe should be expanded to include an EU-wide semiconductor research agenda prioritising issues the private sector is unable to address like basic and theoretical research or prototyping. Possibly, the distribution of funds amongst participating institutions should be proportional to criteria such as the expected contribution to the final outcome and the comparative relative advantage that each EU country has in a relevant field. Moreover, in the spirit of European solidarity, these funds should favour projects allocating significant implementation responsibilities to institutions in lower-income countries whose staff have a proven track-record, such as INSAIT.

Harmonised workforce training to reduce spatial inequalities | Additional funds should also be earmarked for the implementation, on the national

Since the 1990s, Bulgarian decisionmakers have shown their lack of knowledge on public-investment managing in a capitalistic economy.

level, of free professionalising courses in microelectronics to prepare the workforce to take on jobs in semiconductor manufacturing. Crucially, the funds should be made conditional on proving that access to these courses guarantees equality of opportunities and bridges spatial inequalities.

Allowing state-sponsored takeovers and investments under intergovernmental coordination | As has already happened, the EU can craft exceptions to the founding treaties allowing for more direct state intervention in the markets. Indeed, this would be much more than a revanche of 20th-century Keynesianism. The Japanese Investment Corporation, a fund backed by the Tokyo's government, is in fact carrying out the acquisition of the semiconductor behemoth JSR for over EUR 6 billion. The stated aim is clear: "Japan wants to double down on its comparative advantage in materials [...] needed for semiconductor manufacturing" (Kharpal, 2023).

However, given the reality of the common market and the enormous inequalities in the financial prowess and fiscal stance of the 27 member states, some strings must be attached. Thus, the EU should sponsor a new sort of fiscal-coordination treaty establishing a venue for harmonising high-tech industrial policies that would qualify as subsidies under EU law. Such ad-hoc, intergovernmental coordination could serve both solidarity and equality of opportunities by requiring that the allocation of funds reflects existing comparative relative advantages, but dropping the rule of unanimity for authorising all investments. Still, since the 1990s, Bulgarian decisionmakers have shown their lack of knowledge on public-investment managing in a capitalistic economy. Even more so in high-tech sectors. Thus, the resources that these reforms would unlock ought to be allocated according to a carefully crafted catch-up plan. Possibly, the government's public investment agenda could mimick the steps that have led other European countries to success in IT manifacturing. More concretely, the following policies and benchmarks may be set.

Invest in STEM and ICT education | Each and every country that succeeded in reaping the benefits of the digital revolution since the early 2000s has placed a high value on education informatics and STEM (science, technology, engineering, and mathematics) disciplines (cf. Wolf & Terrell, 2016). The establishment of INSAIT is a step in the right direction. However, that is an elite institution that cannot produce the pool of skilled IT and AI professionals that Bulgaria needs. Thus, recovering one of the few effective policies of the real-socialist period (Petrov, 2023, pp. 122, 232), Bulgaria should invest in higher-education institutions around the country, especially outside of Sofia and other big cities to ensure that a booming high-tech industry can locally source its skilled workforce.

Create a high-tech oriented sovereign wealth fund | Due to the currency board, a peculiar arrangement governing its monetary policy, Bulgaria sits over an uninvested public wealth to the tune of several billion euros. But these funds can be earmarked for public expenses in times of need, as the pandemic showed. Thus, it is high time to create a sovereign wealth fund (SWF) that priorities high-tech investments. In this area, Bulgaria can take example from the Estonian Development Fund, the Nowegian Government Pension Fund Global, Singrapore's GIC, or the Qatar Investment Authority (see Engel et al., 2020) and even Russia's National Wealth Fund (Kondratov, 2014). All these SWFs prioritise investing in companies driving innovation and technological advancements while also facilitating knowledge transfer through the creation of joint ventures and ensure local high-tech companies gain access to global markets.

Supporting Innovation and investment with public funds | Following the example of Poland and other high-tech manufacturing hubs, Bulgaria should invecst in the establishment of a startup-support program arranging research grants and providing mentorship to startups and innovative businesses. In part, these funds should be invested in the creation of technology parks and incubators providing office space, mentorship, and networking opportunities for start-ups outside of Sofia.

Boosting export-oriented high-tech manufacturing | Unlike most big players in the high-tech sector, including peripheral ones like Poland, Bulgaria does not have a large domestic market. Thus, it cannot count on a significant customer base for locally produced high-tech products and services. However, Bulgaria is so strongly pro-free trade that its epeorts and imports worth several times more than its entire GDP. But these already good figures could improve further with the introduction of an e-Residency allowing foreign entrepreneurs to establish and manage a business in Bulgaria hasslefree and remotely (Kotka et al., 2016). In terms of benchmarks, Bulgaria's IT sector should priorities export-oriented productions and aspire to catch-up with Poland and Czechia in terms of high-tech manufacturing exports (see Figure 3) in the next decade.

Mobilising public resources and favouring large-scale cooperation will arguably allow all member states to express their full potential in the fields in electronics manufacturing and AI while also preventing unfair competition amongst EU member states.

5 Conclusion

While the vantage point of Bulgaria's high-tech industry could expose policymakers and analysts to pointless nostalgia for a past 'golden age', the reality is much different. Until now, EU policies concerning AI and related technologies, such as semiconductors, have exhibited a distinctive bias towards the largest and wealthiest member states. Crucially, this is not simply a worrying trend in light of the Union's weakening internal cohesion and widening inequalities, but also a tendency that runs contrary to economic rationality. In fact, the national peculiarities that EU regulations such as the AI Act endeavour to forcefully 'harmonise' determine a distribution of comparative relative advantages that runs contrary to these biases. Bulgarian and EU policymakers should hence take account of the significant leads that peripheral countries can take in driving innovation in some sectors and allow their talents and workers to express their whole potential.

Should the EU decide to do so, it would not just be a latecomer. In reality, neither global competitors like China and Russia nor close allies like the USA and Japan seem to be holding onto to the age-old belief in the infallibility of the markets. Instead, they are geared to a more 'dirigiste' model of the state's economic role, albeit to different degrees and with diverging ideological undertones. The EU must therefore decide to let go of some of the rules that helped it come together in the 1980s and 1990s to embrace a really future-proof industrial policy. By so doing, policymakers on all levels must stop looking at the future only from within the 'tunnel of plausibility' and attempt to imagine the disruptive effects that world-changing technologies can bring. In this process, Bulgaria would emerge as a key actor in the transformation of the EU from a club of backwards-looking, de-industrialising countries into a serious global actor.

Yet, to achieve this aim, the policies currently under discussion must be rewritten to the point of becoming unrecognisable to their drafters. Like other successful latecomers to previous industrial revolutions (cf. Gerschenkron, 1962), the EU can still succeed, but it needs to imitate those at top of the food chain. The EU should thus learn from the USA's Chip Act regarding the need for harmonising permits and project-delivery for high-tech manufacturing and development as well as establishing a Union-wide semiconductor research agenda imposing a broader scope of cooperation and harmonised workforce training to reduce spatial inequalities. Meanwhile, taking a page out from Japan's playbook, it should allow state-sponsored takeovers and investments, but not before having introduced a system of fiscal intergovernmental coordination.

While the USA is already financing the "first 'zettascale' supercomputer", which would be 1,000 times faster than the fastest supercomputer available today' (The White House, 2022), Japan is securing its future in the AI age by nationalising a key global player in the semiconductor industry. At the same time, the European parliament is discussing barely understandable details of AI regulation. Since being founded, the EU's strength has laid in its diversity and it is high time for policymaking to valorise these peculiarities to reduce the inequality of opportunities and assert itself on the global stage.

REFERENCES

- Ahmadi, S. (2022, May 5). Ph.D. in Ireland vs. Europe: A comparative overview. Sina Ahmadi. https://sinaahmadi.github.io/posts/phd-in-ireland-vs-europe-a-comparative-overview.html
- Barova, V., Antonov, P., Geshanova, G., Gavrailova, M., & Ivanov, H. (2019). Wolf in sheep's clothing: Tobacco industry's sponsorship and CSR in Bulgaria. Tobacco Prevention & Cessation, 5(Supplement). https://doi.org/10.18332/tpc/105308
- Bošnjak, R. (2021, June 30). IRIS 8—Racunar proizveden u BiH [IRIS 8 Computer made in BiH]. Bošnjak Rudolf Electric Cars. https://www.bev.ba/REFERENCE/computers/iris8/index.html
- Bundesrat. (2021). Beschluss des Bundesrates Vorschlag für eine Verordnung des Europäischen Parlaments und des Rates zur Festlegung harmonisierter Vorschriften für künstliche Intelligenz (Gesetz über künstliche Intelligenz) und zur Änderung bestimmter Rechtsakte der Union [Decision of the Bundesrat: Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (Artificial Intelligence Act) and amending certain acts of the Union]. Der Bundesrat. https://www.bundesrat.de/SharedDocs/beratungsvorgaenge/2021/0401-0500/0488-21.html
- Caruso, D. (1984, January 23). Customs officials seize 400 fake Apple computers. InfoWorld, 6(4), 17.
- Regulation (EU) 2022/2560 of the European Parliament and of the Council of 14 December 2022 on foreign subsidies distorting the internal market, Pub. L. No. 2021/0114/COD, 1 (2021). https://eur-lex.europa.eu/eli/reg/2022/2560/oj
- Der Spiegel. (1982, November 14). Milliarden Dollar Schulden in Moskau [Billions of dollars of debt in Moscow]. Der Spiegel, 40.
- Dimitrova, E., & Viktorov, K. (2014, March 12). Медийната комисия изслуша СЕМ за Волгин [The Media Commission heard CEM about Volgin]. Bulgarian National Television. https://bntnews.bg/bg/a/215128-medijnata-komisiya-izslusha-sem-za-volgin
- Editors of Kompyutar za Vas. (1990). От едитоите [Editorial]. Sp. Kompyutar Za Vas, 6(5–6), 1.
- Engel, J., Barbary, V., Hamirani, H., & Saklatvala, K. (2020). Sovereign wealth funds and innovation investing in an era of mounting uncertainty. In S. Dutta, R. Escalona, B. Lanvin, & S. Wunsch-Vincent (Eds.), The Global Index 2020—Who Will Finance Innovation? (pp. 89–102). World Intellectual Property Organization (WIPO). https://www.wipo.int/edocs/pubdocs/en/wipo_pub_gii_2020.pdf#page=138

- EU Commission. (2020a). Communication from the Commission: A New Industrial Strategy for Europe (COM(2020) 102 final). European Commission. https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593086905382&uri=CELEX%3A52020DC0102
- EU Commission. (2020b). White Paper on Levelling the Playing Field as Regards Foreign Subsidies. European Commission. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2020:253:FIN
- Proposal for a Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence, 2021/0106/COD, European Parliament, EU Council (2021). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206
- Proposal for a Regulation of the European Parliament and of the Council Establishing a Framework of Measures for Strengthening Europe's Semiconductor Ecosystem, 2022/0032/COD, European Parliament, EU Council (2022). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52022PC0046
- Proposal for a Regulation of the European Parliament and of the Council on Harmonised Rules on Fair Access to and Use of Data, 2022/0047/COD, European Parliament, EU Council (2022). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A68%3AFIN
- Proposal for a Regulation of the European Parliament and of the Council on Horizontal Cybersecurity Requirements for Products with Digital Elements, 2022/0272/COD, European Parliament, EU Council (2022). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52022PC0454
- Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011 (Text with EEA relevance.), Pub. L. No. 169, 2017/0353/COD 1 (2019). http://data.europa.eu/eli/reg/2019/1020/oj/eng
- EUROSTAT. (2023, June 1). Percentage of the ICT sector in GDP. EUROSTAT Data. https://ec.europa.eu/eurostat/databrowser/view/isoc_bde15ag/default/table?lang=en
- Fiscutean, A. (2016, February 12). How these communist-era Apple II clones helped shape central Europe's IT sector. ZDNET. https://www.zdnet.com/article/how-these-communist-era-apple-ii-clones-helped-shape-central-europes-it-sector/
- Fleming, N. (2023). How Grenoble has mastered industry–academia science collaborations. Nature. https://doi.org/10.1038/d41586-023-00109-x
- García-Herrero, A., & Poitiers, N. (2023). Europe's promised semiconductor

subsidies need to be better targeted (Blog Post) [Policy Brief]. Bruegel. https://www.bruegel.org/blog-post/europes-promised-semiconductor-subsidies-need-be-better-targeted

- Georgieva, I., Timan, T., & Hoekstra, M. (2022). Regulatory Divergences in the Draft Ai Act: Differences in Public and Private Sector Obligations. Publications Office. https://doi.org/10.2861/69586
- Gerschenkron, A. (1962). Economic Backwardness in Historical Perspective: Belknap Press.
- Gobble, M. M. (2014). European High-Tech Industry at the Crossroads. Research-Technology Management, 57(1), 2–8. https://doi.org/10.5437/08956308X5701001
- Hussein, P. (2020, June 24). Кошлуков пред СЕМ заради “Вяра и общество”, БНТ го излъчва онлайн [Koshlukov in front of CEM because of “Faith and Society”, BNT broadcasts it online]. 24chasa.Bg. https://www.24chasa.bg/bulgaria/article/8736500
- INSAIT. (2022, May 11). About INSAIT. Institute for Computer Science, Artificial Intelligence, and Technology. https://insait.ai/about-insait/
- INSAIT. (2023a). Faculty & PhD mentors | INSAIT. Institute for Computer Science, Artificial Intelligence, and Technology. https://insait.ai/phd-mentors/
- INSAIT. (2023b). INSAIT announces its international doctorate program with professors from CMU, EPFL, ETH Zurich, MIT, Yale. Institute for Computer Science, Artificial Intelligence, and Technology. https://insait.ai/insait-announces-its-international-doctorate-program-with-professors-from-cmu-epfl-eth-zurich-mit-yale/
- INSAIT. (2023c, June 27). INSAIT стартира лятна AI програма за ученици [INSAIT launches summer AI program for students]. Institute for Computer Science, Artificial Intelligence, and Technology. https://insait.ai/insait-стартира-лятна-ai-програма-за-ученици/
- Intel. (2022, March 15). Intel in Germany. Intel Corporation. https://www.intel.com/content/www/us/en/corporate-responsibility/intel-in-germany.html
- Jehlička, P. (2021). Eastern Europe and the geography of knowledge production: The case of the invisible gardener. Progress in Human Geography, 45(5), 1218–1236. https://doi.org/10.1177/0309132520987305
- Kaspersky Lab. (2014, September 25). Агат 9—Советский ответ Apple. Часть первая [Agat 9—The Soviet answer to Apple. Part One]. Khabr. https://habr.com/ru/articles/237789/
- Kharpal, A. (2023, June 26). Japan-backed fund to buy critical semiconductor

firm JSR for $6.3 billion as chip tensions rise. CNBC. https://www.cnbc.com/2023/06/26/japan-backed-fund-to-buy-semiconductor-firm-jsr-for-6point3-billion.html

- Kondratov, D. I. (2014). The government's role in the export of capital from Russia. Herald of the Russian Academy of Sciences, 84(5), 385–393. https://doi.org/10.1134/S1019331614050025
- Kotka, T., del Castillo, C. I. V. A., & Korjus, K. (2016). Estonian e-Residency: Benefits, Risk and Lessons Learned. In A. Kő & E. Francesconi (Eds.), Electronic Government and the Information Systems Perspective (pp. 3–15). Springer International Publishing. https://doi.org/10.1007/978-3-319-44159-7_1
- Kratz, A., Zenglein, M. J., Gregor Sebastian, & Witzke, M. (2022). Chinese FDI in Europe: 2021 Update: Investments remain on downward trajectory – Focus on venture capital (MERICS Report, pp. 1–19) [Working Paper]. Mercator Institute for China Studies (MERICS). https://merics.org/sites/default/files/2022-04/MERICS-Rhodium-Group-COFDI-Update-2022-2.pdf
- Lopez-Cobo, M., & De Prato, G. (Eds.). (2022). AI Watch Index 2021. Publications Office of the European Union. https://doi.org/10.2760/921564
- Mańko, R. (2015). Contract lawand the Digital Single Market: Towards a new EU online consumer sales law? (Policy Brief PE 568.322; In-Depth Analysis, pp. 1–32). European Parliamentary Research Service. https://data.europa.eu/doi/10.2861/16497
- MS-RB. (2020). Концепция за развитието на изкуствения интелект в България до 2030 г.: Изкуствен интелект за интелигентен растеж и проспериращо демократично общество [Concept for the Development of Artificial Intelligence in Bulgaria by 2030: Artificial intelligence for smart growth and a prosperous democratic society]. Council of Ministers of the Republic of Bulgaria. https://www.strategy.bg/StrategicDocuments/View.aspx?lang=bg-BG&Id=1338
- Nakov, S. (2023, April 28). В България има цели фирми, които са изцяло задвижвани само от изкуствен интелект [There are entire companies in Bulgaria that are entirely powered by artificial intelligence] (V. Spasova, Interviewer) [Bloomberg TV - Bulgaria]. https://www.bloombergtv.bg/a/17-v-razvitie/117765-v-balgariya-ima-tseli-firmi-koito-sa-iztsyalo-zadvizhvani-samo-ot-izkustven-intelekt
- Neykova, Z. (2017, July 6). Скандал в парламента заради Съвета за електронни медии [Scandal in the Parliament over the Council for Electronic Media]. Bgonair. https://www.bgonair.bg/a/2-bulgaria/72818-skandal-v-parlamenta-zaradi-saveta-za-elektronni-medii

- Nica, D. (2023). Report on the Chips Act. European Parliament. https://www.europarl.europa.eu/doceo/document/A-9-2023-0014_EN.html.
- Petrescu, A., Chirtoacă, G., & Badea, D. (2012). ZX Spectrum—Retrospectiva [ZX Spectrum—A retrospective] (DOI punct ZERO TV, Interviewer) [16:9]. https://www.youtube.com/watch?v=ngdbz3fp3wk.
- Petrov, V. (2023). Balkan cyberia: Cold War computing, Bulgarian modernization, and the information age behind the Iron Curtain. The MIT Press. https://direct.mit.edu/books/book-pdf/2128212/book_9780262373265.pdf.
- Pravetz Computers. (2017, October 3). История [History]. https://pravetz.bg/history.
- Rachman, G. (2023, June 19). Europe has fallen behind America and the gap is growing. Financial Times.
- Righi, R., Melisande, C., Sofia, S., Michail, P., Vazquez-Prada Baillet, M., Vazquez-Prada Prato, G. V.-P., López-Cobo, M., Benetta, A. D., Nigris, S. D., Desruelle, P., Brown, N. D., Gutierrez, E. G., Hupont-Torres, I., Plumed, F. M., Rodríguez, E. M., Nativi, S., Nepelski, D., Pineda, C., Rossetti, F., ... Tolan, S. (2022). AI Watch Index 2021 (http://data.europa.eu/89h/e3757f41-fe54-4330-946d-ae897686164f) [dataset]. JRC Data Catalogue. https://data.europa.eu/doi/10.2760/921564
- Sudetic, C. (1990, December 21). Bulgarians Linked to Computer Virus. The New York Times, 9.
- Telarico, F. A. (2021a). Demythologising the 'Russification' of Bulgarian media's treatment of civil society: Analysis of the transfer of Russian mainstream news media's cliches and rhetoric on CSOs upholding human rights and environmentalism to Bulgaria (pp. 1–72). BlueLink Foundation. https://doi.org/10.5281/zenodo.5903390.
- Telarico, F. A. (2021b). Digital Civic Cultures in Post-socialist South Eastern Europe: Lessons, Prospects and Obstacles After Thirty Years of Media (Il) literacy in the Region. In Дигитална гражданска компетентност и медийни стереотипи [Digital civic competence and media stereotypes] (1st ed., pp. 95–108). Polymona. https://fatelarico5.wixsite.com/website/chapter-2021-2.
- The White House. (2022, August 9). CHIPS and Science Act Will Lower Costs, Create Jobs, Strengthen Supply Chains, and Counter China. The White House. https://www.whitehouse.gov/briefing-room/statements-releases/2022/08/09/fact-sheet-chips-and-science-act-will-lower-costs-create-jobs-strengthen-supply-chains-and-counter-china/.
- Tilman Kuhn, Orion Berg, Marc Israel, & Kate Kelliher. (2022, September 7). The EU Releases its Second Annual FDI report showing increased momentum in FDI regulation and screening in the EU27. White & Case LLP. https://www.whitecase.

com/insight-alert/eu-releases-its-second-annual-fdi-report-showing-increased-momentum-fdi-regulation.

- Timmers, P. (2022). Digital Industrial Policy for Europe (pp. 1–74) [Policy paper]. Centre on Regulation in Europe (CERRE). https://cerre.eu/wp-content/uploads/2022/12/Digital-Industrial-Policy-for-Europe.pdf.
- Tooze, A., & Abadi, C. (2023, June 23). The United States vs. Europe (95). https://www.stitcher.com/show/ones-and-tooze/episode/the-united-states-vs-europe-304695570.
- Chips and Science Act, Pub. L. No. H.R.4346 (2022). http://www.congress.gov/bill/117th-congress/house-bill/4346.
- Van Gool, L. (2022, July 6). A pioneer in computer vision joins INSAIT [Web site]. https://insait.ai/a-pioneer-in-computer-vision-joins-insait/.
- Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act—Analysing the good, the bad, and the unclear elements of the proposed approach. Computer Law Review International, 22(4), 97–112. https://doi.org/10.9785/cri-2021-220402.
- Vechev, M., Tsankov, P., & Raychev, V. (2023, April 8). България като световен център в развитието на изкуствения интелект—Възможно ли е? [Bulgaria as a world centre in the development of artificial intelligence—Is it possible?] (Y. Dimitrov, Interviewer) [Television]. https://bnt.bg/news/balgariya-kato-svetoven-centar-v-razvitieto-na-izkustveniya-intelekt-vazmozhno-li-e-316805news.html.
- Wallach, W. (2022, June 16). The battle between autocracy and democracy has blinded us to the A.I. Oligopoly. Fortune. https://fortune.com/2022/06/16/ethics-autocracy-democracy-blinded-tech-oligopoly-artificial-intelligence-politics-wendell-wallach/.
- WB. (2022a). High-technology exports (% of manufactured exports): Bulgaria, Poland, Romania, Slovak Republic, Czechia, Hungary, Latvia, Estonia, Lithuania. World Bank Open Data. https://data.worldbank.org.
- WB. (2022b). ICT service exports (% of service exports, BoP): Bulgaria, Poland, Romania, Slovak Republic, Czechia, Hungary, Latvia, Estonia, Lithuania. World Bank Open Data. https://data.worldbank.org.
- Wolf, M., & Terrell, D. (2016). The High-Tech Industry, What is it and Why it Matters to our Economic Future. Beyond the Numbers: Employment and Unemployment, 5(8), 1–7.

Chapter 8

Artificial Intelligence in the digital transformation era: A Polish perspective

Konrad Sobański

1 Introduction

The 21st century has seen the unprecedented expansion of digital transition in all aspects of human life, in turn creating a new environment for societies and policymakers around the world, including the European Union. One of the most profound disruptions seems to result from a technology called Artificial Intelligence (AI). In general, AI is a computer science of creating intelligent machines that can learn and act automatically, and encompasses machine learning, neural networks, deep learning, large language models, and natural language processing (Goldman Sachs, 2023a; Rzeźnik, 2023). Artificial Intelligence aims to automate the decision-making processes of economic agents, increase the efficiency of operations, and promote economic growth. However, implementing AI technology is definitely a complex process and raises many concerns. In response to the emergence of AI, governments across the globe are designing their official approaches to this modern technology and formulating a regulatory framework for this area. It is an extremely difficult task given that policymakers are navigating in entirely ‘unchartered waters’. Nevertheless, governments need to reconcile the potential benefits expected from the application of AI with its risks to humankind in the long term. Detailed analysis of the latter is crucially important and should be at the forefront of rational regulatory planning.

The main aim of this chapter is to explore and identify challenges and opportunities related to the digital transformation through AI in Poland. First, the chapter analyses AI trends in Poland and the world. Second, the current state of AI policy and examples of AI companies in Poland are described. Third, the chapter presents the regulatory framework for AI on the Polish and EU levels. Fourth, challenges and opportunities for AI policy in Poland are discussed. Finally, the article formulates guidelines that could be used for governance and regulation in the field of AI.

2 Key country data on AI trends in Poland and the world

The term Artificial Intelligence relates to machines that mimic human intelligence. It consists of three categories. The first one, artificial narrow intelligence, also known as ‘weak AI’, performs specific tasks like voice recognition. The second and third categories, i.e., artificial general and super intelligence, are considered ‘strong AI’ and relate to machines having cognitive abilities similar or larger than those of humans. However, today there are no actual examples of strong AI (Goldman Sachs, 2023a; OECD. AI, 2023). Artificial intelligence in practice can take the form of a computer system (software), such as virtual assistants, search engines, recognition systems, or more physically in the form of autonomous cars, robots or the Internet of Things (Rzeźnik, 2023, pp. 5–6).

AI encompasses two major areas. The first one, machine learning (ML), focuses on developing algorithms to improve predictions and decisions made by computers based on experience and patterns gathered from large amounts of data. There are several subsections of ML: neural networks (NN), deep learning (DL) and large language models (LLM). A neural network is a mathematical model designed similarly to the human brain, comprising neurons (nodes) transforming inputs into outputs through computations. An example of NN is Google’s search algorithm. DL is a neural network with at least three layers of nodes, not requiring a labelled dataset and less dependent on human interaction. Autonomous driving and speech recognition are instances of DL. Large language models like ChatGPT are based on deep neural networks trained on large amounts of unlabelled data. The second area of AI is constituted by natural language processing (NLP) and concentrates on making computers able to understand human language. NLP is based on computational linguistics and statistical, machine learning, and deep learning models to understand words in a manner similar to humans. NLP applies two techniques: syntactic analysis which identifies the relationship between words in sentences, and semantic analysis which identifies the meaning of the words in a sentence. Examples of NLP are Google Translate, chatbots like Siri and Alexa (Goldman Sachs, 2023a; Rzeźnik, 2023, pp. 6–7).

In 1637, the philosopher and scientist R. Descartes stated that one day machines would make decisions and act in an 'intelligent' way, which might be taken as the foundation of the term Artificial Intelligence (Miernik, 2023). The history of AI technology is long since it spans from the 1950s when A. Turing developed a test to check a machine's ability to present intelligent behaviour, to the 1960s when the first AI chatbot ELIZA was created by J. Weizenbaum of the Massachusetts Institute of Technology. Yet, it was not until the 2000s that the development of AI technology accelerated with the advent of the Internet, Big Data and the high computing power of computer processors. There are several landmark examples of the application of AI since 2010, such as (Goldman Sachs, 2023a; Miernik, 2023):

- Kinect for Xbox 360 – the gaming device to track body movement – introduced by Microsoft in 2010;
- Siri – a voice assistant to understand natural language – introduced by Apple in 2011;
- Alexa – a virtual assistant – introduced by Amazon in 2014;
- Language Model for Dialogue Applications (LaMDA) – a model to generate human-like responses in conversation – introduced by Google in 2021;
- DALL-E – a model that generates images from text – developed by OpenAI in 2021; and
- ChatGPT (Generative Pretrained Transformer) – a generative AI tool that generates human-like responses based on text inputs – developed by OpenAI in November 2022.

Generative AI creates text, images, video, audio and other content in response to natural language prompts. Generative AI is unlike traditional AI for two reasons. First, it generates new content, rather than just training computers to make predictions. Second, it allows for communication with a computer in natural human language, rather than in programming languages. ChatGPT is the fastest application to have reached 200 million monthly active users across the globe (Goldman Sachs, 2023a). The most noteworthy examples of generative AI tools are presented in Table 1.

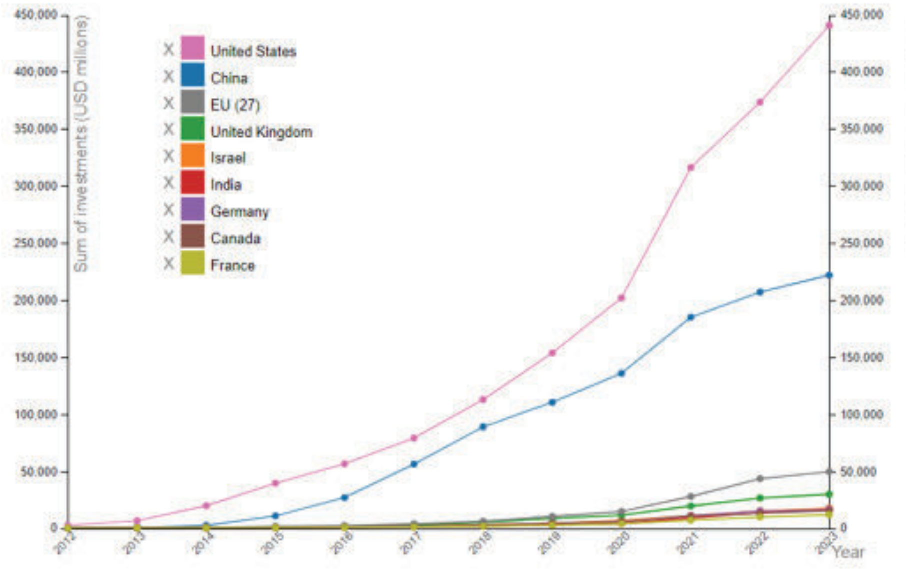
Table 1: **Generative AI tools**

Category of generative AI	Example	Information
Chatbots	ChatGPT	Developed by OpenAI
	Bard	Developed by Google
	Bing Chat	Developed by Microsoft
Text generators	Copy.ai	Generates social media posts and emails
	Frase.io	Generates slogans, summaries, introductions, articles, titles, and product descriptions
Image generators	DALL-E	Generates images, offering the commercial rights to created content
	starryai	Generates artwork with different style options, offering ownership of produced content
	Midjourney.com	Generates images based on text
Video generators	Elai.io	Converts text to video
	Flexclip	Video creation and editing tools
Voice generators	Lovo.ai	Text-to-speech conversion and generates realistic AI voiceovers
	Music generators	Generates royalty-free music based on preferences, offering a perpetual licence

Source: Own compilation based on Goldman Sachs, 2023a; Rzeźnik, 2023.

AI technology is expected to affect the global economy significantly, estimated to add value of USD 17 trillion up to USD 26 trillion to it (Business Insider, 2023). Generative AI itself is anticipated to increase global GDP by as much as 7%. Such forecasts are driving huge capital inflows into the AI industry. The United States is by far the largest venture capital investor in AI technology with close to USD 400 billion already spent by the end of 2022, followed by China with over USD 200 billion and the European Union with USD 50 billion (Figure 1). Investment in AI is forecast to reach USD 200 billion per annum globally by 2025 (Goldman Sachs, 2023b, 2023c).

Figure 1: **Cumulative venture capital investment in AI between 2012 and 2023 (USD million)**



Source: OECD.AI, 2023, visualisations powered by JSI using data from Preqin, accessed on 27/8/2023, www.oecd.ai.

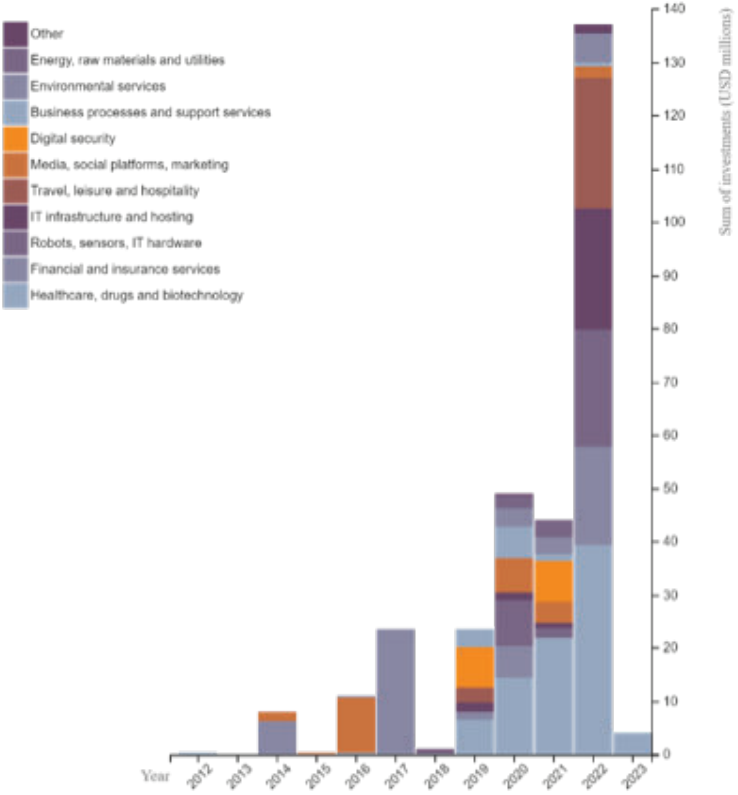
Note: The chart displays venture capital investments in AI in USD millions by country from 2012 onwards; cumulative value for the end of the year; estimated value for 2023.

Entities that are engaged with AI development include the world's biggest companies like Alphabet (Google), Amazon, NVIDIA or Microsoft. Their individual market capitalisations exceed USD 1 trillion (Jareno & Yousaf, 2023). The main areas of potential applications of AI are automation of services and products, improvement of labour productivity, marketing, and detection of anomalies (Wilczyńska-Baraniak, Rentflejsz & Dzięciołowski, 2022). It is worth stressing, however, that benefits expected from AI application might be generated directly not only by businesses but by the government sector as well, following the digitalisation trend seen in public administration in recent years (Aristovnik et al., 2021). In general, entities around the world are using AI in multiple applications such as (Ramaswamy, 2017; Rzeźnik, 2023):

- in the IT area: to detect and deter security intrusions, to resolve user's technology problems (e.g., the Polish companies Nethone and Allegro employ machine learning algorithms and behaviour analysis to analyse risk and prevent fraud online);
- in the marketing area: to anticipate future customer purchases and present offers accordingly, to monitor social media comments and tailor promotions;
- in finance and accounting: to perform high-frequency trading;
- in production: to automate processes (e.g., General Electric shortened the planning process for engine production through AI models); and
- in customer service: to automate call distribution, to develop product recommendations based on analysis of customer data (e.g., Netflix employs AI algorithms to recommend movies to users; Polish fashion brand NAOKO uses AI to adjust designs based on data provided by female customers).

Poland is considered a good prospect when it comes to the application of AI. It has globally respected developers (as stressed by the CEO of OpenAI) and the first companies that are implementing AI-based solutions on a wider scale such as SentiOne (offering an alternative chatbot to ChatGPT). More and more companies plan to invest in AI driven by the 'FOMO' effect (fear of missing out, or fear of missing a market opportunity) (Duszczyk, 2023). Although the level of venture capital investment in AI in Poland is not significant by international standards, it has been rising dynamically, especially since 2020. In 2022, it reached over USD 130 million (Figure 2). Most of the capital has been invested in healthcare, drugs and biotechnology, media, social platforms and marketing. Experts indicate that the further development of the Polish AI market will be determined by the availability of capital and planned regulations. The Polish market is expected to gain strongly from the EU's regulations requiring the local processing of data used in AI models. The European Union wants to prohibit user data from 'moving outside' the member states. This might prevent large companies from the USA and China from entering the European market, thereby creating an opportunity for Polish and EU technology companies (Duszczyk, 2023).

Figure 2: **Total venture capital investments in AI by industry in Poland between 2012 and 2023 (USD million)**



Source: OECD.AI, 2023, visualisations powered by JSI using data from Preqin, accessed on 27/8/2023, www.oecd.ai.

The application of AI in companies in Poland is growing rapidly. According to a study by KPMG and Microsoft entitled Business Digital Transformation Monitor (KPMG, 2023), 15% of businesses in Poland are employing AI technologies (compared to 35%–37% across the globe), while 13% plan to implement them by the end of 2023. AI is used most often in marketing (50% of companies surveyed by KPMG), manufacturing (46%) and supply chain planning (42%). In the coming years, KPMG expects an acceleration of AI adoption in the customer service and HR areas, as a result of language models like ChatGPT. AI tools are most often applied by businesses operating in the following sectors: information technology, media and communications (25%) and life sciences (21%). The most frequently implemented AI technologies include mobile solutions (73% of companies surveyed), computer-aided decision-making (used in 70% of companies surveyed), cloud services (68% of companies surveyed), solutions based on automation and robotics (implemented in 58% of companies surveyed) and machine-to-machine communication (39% of companies surveyed). In Poland, surprisingly, as many as 62% of companies that use AI do not monitor the effectiveness of its implementation (vs. the global average of 68%). However, all those companies that measure efficiency indicate that AI brings multiple financial benefits: it increases productivity, improves the quality of products/services, improves financial results and strengthens competitiveness. At the same time, AI helps companies achieve their sustainability goals by reducing energy consumption, lowering greenhouse gas emissions and improving efficiency in the use of natural resources (KPMG, 2023).

The growing engagement with the development of AI in Poland is also observed in the area of research. The leading research units are the Polish Academy of Sciences and Warsaw University of Technology, as measured by the number of AI publications and their citations between 2020 and 2023 (Table 2). The Polish Academy of Sciences generated 11.3% of all citations of Polish publications on AI in the period 2020–2023. The Warsaw University of Technology produced 8.2% of all AI publications during this period. The research activity is very concentrated considering that the top nine institutions are responsible for over 50% of publications in Poland. The Polish researchers on AI mostly cooperate with scientists from the USA, the UK, Canada, France, Germany, Italy and Spain (Figure 3).

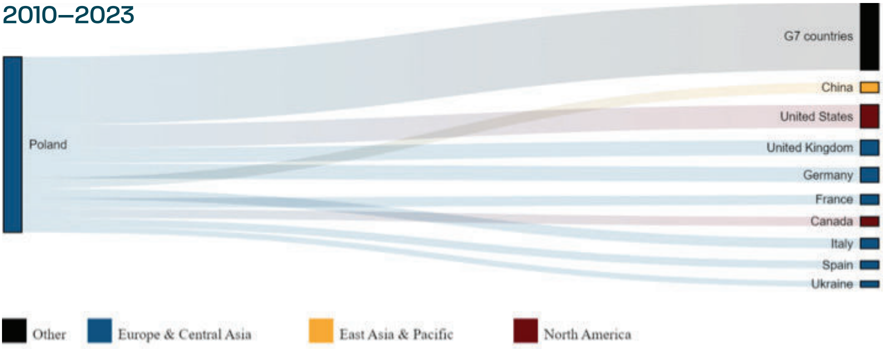
The application of AI in companies in Poland is growing rapidly - 15% of businesses in Poland are employing AI technologies (compared to 35%–37% across the globe).

Table 2: **Top institutions preparing AI research publications and citations in Poland (cumulative number of publications and citations in the period 2020–2023).**

Institution	No. of citations	No. of publications
Polish Academy of Sciences	87,110	6,239
Warsaw University of Technology	63,321	8,026
University of Warsaw	58,287	5,089
Wroclaw University of Science and Technology	56,316	6,454
Poznan University of Technology	42,001	4,365
AGH University of Science and Technology	40,978	6,118
Jagiellonian University	38,868	3,820
Silesian University of Technology	37,703	5,799
Gdansk University of Technology	26,378	3,710
Poland	**773,661**	**98,337**
Top institution as % of Poland	**11.3%**	**8.2%**
Median for top institutions	**42,001**	**5,799**
Ranking leader as % of the median for top institutions	**207.4%**	**138.4%**

Source: Own compilation based on OECD.AI, 2023; OECD.AI, 2023, visualisations powered by JSI using data from OpenAlex, accessed on 28/8/2023, www.oecd.ai.
Note: Forecast for 2023.

Figure 3: **Collaboration in AI research publications in Poland in the period 2010–2023**



Source: OECD.AI, 2023, visualisations powered by JSI using data from OpenAlex, accessed on 28/8/2023, www.oecd.ai.
Note: The diagram presents collaboration on AI scientific publications. Each publication is assigned to one or more countries depending on the affiliation(s) of its author(s). The thickness of a connection represents the number of AI publications co-written by authors affiliated with institutions in another country; cumulative data since 2010.

3 Current state of AI policy and examples of AI companies in Poland

The Polish government has adopted a policy on AI. "The policy for the development of artificial intelligence in Poland from 2020" (AI Policy, 2020), approved by the Council of Ministers in Poland on 28 December 2020, sets the goal of raising the number of state-owned companies fostering AI projects in Poland. More broadly, the AI Policy was formulated to support the public, companies, researchers and government in taking advantage of the opportunities associated with AI, while protecting human rights and fair competition. The AI Policy incorporates objectives defined previously in Poland and the EU in documents such as the "Dynamic Poland 2020" Strategy for Innovation and Efficiency of the Economy, Public Data Opening Programme, Position of the Visegrád Group on Artificial Intelligence, and Recommendations of the High-Level Expert Group on Artificial Intelligence to the European Commission "Ethics Guidelines for Trustworthy AI" (AI Policy, 2023; GovTech, 2023).

The Polish government has described industries and activity areas of enterprises and research entities that will be supported by public programmes as their potential has been rated the highest in terms of adding value to the economy and their competitiveness in foreign markets. These areas are called national smart specialisations (NSS). The list of NSS currently contains 13 industries and technological areas (Ministry of Development and Technology, 2023).

NSS areas directly related to AI development are NSS 10 "Smart networks and information, communication and geo-information technologies" and NSS 11 "Automation and robotics" (Rzeźnik, 2023). According to the description in NSS 10, "smart networks" refer to ICT technologies and systems applied to infrastructure to ensure resource savings and environmental protection. "Information and communications technology" encompasses technologies that collect and transmit information in electronic form. "Geo-information technologies" cover technologies related to acquiring, analysing, sharing and visualising geo-information (NSS, 2023). In turn, NSS 11 includes process design and optimisation, process automation and robotisation technologies, diagnostics and monitoring, control systems, automation machinery and equipment (NSS, 2023). The importance of NSS 10 and NSS 11 in Poland's international trade and government support for R&D is presented in Table 3.

Table 3: **Importance metrics for National Smart Specialisations Nos. 10 and 11**

Importance metric	NSS 10	NSS 11
Share of national exports (%)	3.5	1.7
Exports (in PLN billion)	30.8	14.8
Share of national imports (%)	3.7	2.2
Imports (in PLN billion)	32.6	19.0
Government financing for R&D projects between 2014 and 2020 (in PLN billion)	6.2	4.9

Source: Own compilation based on NSS, 2023.
Note: International trade metrics provided for 2017.

There are multiple examples of AI businesses in Poland offering products in the areas of NSS 10 and 11, including (Rzeźnik, 2023):

- Synerise S.A. – which provides a platform for large organisations to undertake automated marketing activities and innovative data processing solutions. It combines the collection and analysis of real-time data with decisions automated by AI.
- Nomagic Sp. z o.o. – a startup designing software and building blocks for robots that perform repetitive tasks in logistics centres (offering products such as a robotic arm, an AI-based system for recognising objects, and a cloud-based system for managing robots). The technology allows robots to recognise objects, pick them up and place them in the desired location.
- Neptune.ai (Neptune Labs Sp. z o.o.) – a startup that develops a platform supporting the AI model building. The company specialises in solutions for developers and engineers of information systems.
- AGICortex Sp. z o.o. – a startup that has developed AI software for robots, drones and analytics systems. The technology used is characterised by 'unsupervised learning', which allows automatic modelling of the space surrounding the robot, permitting it to be used in a wide range of different applications.
- MIM Solutions – a spin-off operating at the Faculty of Mathematics, Informatics and Mechanics of the University of Warsaw. The team of the technology spin-off advises organisations on how they can use AI in their business. It has developed AI algorithms to solve problems faced by organisations in the healthcare, e-commerce, automotive, security and public sectors.

Other NSS areas prescribed forth by the Polish government relate to AI more indirectly (AI technology can be applied by entities operating in these areas). NSS 1 "Healthy society" covers drugs, diagnosis and therapy of diseases. In the field of medical technology, AI is increasingly used in the processes of developing new products, disease diagnosis and therapy through diagnostic imaging (MRI and CT scans) or image recognition. Below are some examples of Polish companies that offer AI-based products in the area of NSS 1 (Rzeźnik, 2023):

- NaturalAntibody S.A. – a startup that has developed a method to reduce the time and cost of production of antibody-based drugs. The method is based on machine learning and data collected by researchers over the last 40 years and has been applied by large pharmaceutical companies such as AstraZeneca.
- Radiato.ai (Medical Image Dataset Annotation Service Sp. z o.o.) – a startup created by researchers from the Gdansk University of Technology. The company specialises in developing AI-based systems to support the diagnosis of kidney tumours based on abdominal CT images.

NSS 2 "Modern agriculture, forestry and food" amongst others covers innovative machinery, equipment and technologies used in agri-food, measures to reduce agriculture's negative impact on the environment, biological progress in plant and animal production, or innovative fertilisers and plant protection products. NSS 3 "Sustainable (bio)products, (bio)processes and environmental development" focuses on issues like the development of biological systems (including genetic and metabolic engineering and bioinformatics), bioproducts, specialty chemical products, and modern technologies in environmental protection (NSS, 2023). SmokeD Sp. z o.o. is an example of the application of AI in NSS 2. It has developed a system, applied by more than 80 forest districts in Poland, for automatic fire detection in which cameras monitor the area looking for smoke (Rzeźnik, 2023).

NSS 4 "Sustainable energy" encompasses energy generation issues, power supply reliability, smart solutions in power grids, smart metering, energy storage methods, renewable energy sources and fuels, prosumer energy, alternative fuels and environmental protection (NSS, 2023). Examples of Polish companies applying AI in this area include (Rzeźnik, 2023; Klekowski, 2023):

- S-Labs Sp. z o.o. – a startup that has developed energy-use sensors powered by cloud computing analytics. Installed in apartments and accompanied with software, the measuring devices allow analysis of energy consumption behaviour.
- Aigoritmics offering the AigoML platform that is used to predict energy

production, such as from photovoltaic and wind farms, forecast electricity demand and energy prices, balance of multi-source generation systems (e.g., solar farms, wind farms).

- Connectpoint Sp. z o.o. – a startup that has developed the application Smartvee to collect data from electricity meters. The application allows for human errors to be eliminated and irregularities in consumption to be reported.

NSS 5 "Smart zero-carbon construction" refers, among others, to innovative building materials, energy and environmental auditing, technologies in the field of materials processing and reuse (NSS, 2023). Below are some examples of Polish companies that offer AI-based products in the area covered by NSS 5 (Rzeźnik, 2023):

- AMS AI sp. z o.o. – a startup which has developed an AI-based service Sunmetric to perform automated analysis of the performance of a potential photovoltaic installation at a given site.
- Quantifier Sp. z o.o. – a startup which has developed the digital platform Envirly to help companies manage their carbon footprint. The platform allows the measuring, analysing and running of simulations to reduce carbon emissions. The company is cooperating with the Polish Development Fund and the National Energy Conservation Agency.

NSS 6 "Environmentally friendly transportation" concentrates on environmentally friendly transport systems (NSS, 2023). AI is applied by the following Polish companies in this area (Rzeźnik, 2023):

- Nevomo Sp. z o.o., which has developed the high-speed magnetic rail concept MagRail allowing rail vehicle to travel at speeds up to 550 km/h on existing railroad tracks using magnetic levitation. The company wants to develop a vacuum railroad reaching speeds of up to 1,200 km/h using a new rail infrastructure. At the end of 2022, the company was awarded funding of EUR 5 million to further develop its technology.
- Blees Sp. z o.o., which is developing, together with the Silesian University of Technology and the Cracow University of Technology, an 'on-demand' minibus that will move fully autonomously (with the option of the driver taking control). The company also offers the solution "City Eye" for active monitoring based on AI intelligent image analysis for the automatic detection of dangerous events.

NSS 7 "Closed loop economy" focuses on models where the added value of resources is maximised or the amount of waste generated is minimised, while the waste generated is managed in accordance with the waste hierarchy (waste prevention, recycling, disposal) (NSS, 2023). Examples of Polish companies applying AI in this area include (Rzeźnik, 2023):

- Bin-e Sp. z o.o., a startup recognised by the industry website The Recursive.com as a leader in the Polish climate-tech market. The startup has designed a device that uses AI to automatically sort and compress waste, control fill levels and analyse data to optimise logistics processes.
- Four Point Sp. z o.o., developing technologies to reduce the environmental impact of open-pit mines. It offers several applications. TerraEye helps mining professionals make decisions through satellite imagery and analytics using AI algorithms. The Autonomous Transport Platform service fosters autonomous machine operation in open-pit mines.

NSS 8 "Advanced materials and nanotechnology" encompasses, amongst others, nanostructured materials, biomimetic, bionic and biodegradable materials, advanced materials in renewable energy, ultra-lightweight materials, and ultra-strong materials with radically improved heat resistance (NSS, 2023). In this area, QSAR Lab, a spin-off of the University of Gdansk in Poland, is using AI and developing the application nanoQSAR Toolbox to predict the toxicity of metal oxide nanoparticles and select those with the best properties that are safe for health and the environment. The project has received funding from the European Regional Development Fund in the amount of PLN 1.7 million (Rzeźnik, 2023).

NSS 9 "Electronics and photonics" concentrates on technologies for sensors and detectors, photovoltaics and fibre optics, telecommunications systems and networks, innovative circuits and systems for electronics, and integrated photonics (NSS, 2023). Under this heading, In-Lab Sp. z o.o. manufactures technical equipment and software such as intelligent battery-operated radio sensors used to measure temperature and other parameters, including in public transportation vehicles in Poland.

The NSS 12 "Creative industries" covers design, including tools to support the design process, games and multimedia (NSS, 2023). Below are some examples of Polish companies that offer AI-based products in the area covered by NSS 12 (Rzeźnik, 2023):

- Esports Lab Sp. z o.o. – a startup developing AI-based tools for e-sports such as a platform to analyse individual performances of players. The project has received funding from the National Centre for Research and Development and from the European funds.

- SentiOne Sp. z o.o., providing AI-based applications for monitoring social media such as SentiOne Listen to monitor online discussions about a selected brand, and SentiOne Automate – a conversation bot for brand communication.
- GGPREDICT Sp. z o.o., offering the AI-based system GGPredict to adjust training to players.

NSS 13 "Marine technologies" includes the design, construction and conversion of specialised vessels, marine and coastal structures, processes and equipment used for logistics based on maritime and inland transportation (NSS, 2023). Navdec Sp. z o.o. is an example of a company offering products in this area. It offers a decision support system developed by a team from the Maritime Academy in Szczecin to help ships avoid collisions at sea, and for automatic route planning. The system integrates with existing systems on the ship, forecast the situation around the ship, and generates an alert for the ship's crew (Rzeźnik, 2023).

4 Current state of AI regulations in Poland and the EU

In Poland, no legal regulations explicitly refer to the protection of AI as such, or define it. One can nevertheless obtain protection for its individual components based on regulations already found in the Polish legal order:

- Software, constituting an essential part of AI, can be legally protected as a computer application under the Act of 4 February 1994 on Copyright and Related Rights. Articles 1(1) and (2)(2) of the Act state that applications are subject to copyright protection if they are a manifestation of creative activity of an individual character. As a result, only the holder of the copyright to the AI software can use it and decide to what extent others may do so (similarly to OpenAI that holds copyright to the source code of its ChatGPT). What is more, if certain elements of AI can be considered an invention subject to patent protection for 20 years under the Industrial Property Law of 30 June 2000, the creator will also be able to obtain the right of protection for the invention, i.e., a patent (similarly to Google that holds a patent to a machine learning system to prepare the best recommendations for users). Although computer programmes are not presently recognised in Poland as inventions, computer-assisted inventions can be subject to patenting (Rze⊠nik, 2023, pp. 39–40, 42–43).
- In turn, a database constituting a component of AI is eligible for protection under the Polish Law of 27 July 2001 on the Protection of Databases. The

producer of such a database, which has made a significant investment in its creation, is entitled to an exclusive sui generis-right even if the database is not of a 'creative' nature and not subject to copyright protection (Rzeźnik, 2023, p. 43).

- If an AI tool contains the know-how of its creator or confidential information, such company secrets can be protected based on relevant contractual obligations to maintain confidentiality (non-disclosure agreements, NDAs) and under the provisions of the Act of 16 April 1993 on Combating Unfair Competition (Rzeźnik, 2023, p. 44).

On the other hand, in Poland there is no legal protection of AI creations, i.e., works created solely by artificial intelligence. They are not subject to copyright protection, even if highly original, since a work subject to protection under the copyright law can only be created by a human being. In the case of AI-generated inventions, it is also not possible to grant them patent protection since no individual can be designated as the inventor. Currently, AI creations are becoming part of the public domain and their use is not subject to any restrictions (Rzeźnik, 2023, pp. 44–45). However, when a person has given a certain property of the work, or rather, the characteristics referred to in Article 1 of the copyright law, copyright protection might be afforded to this person. When the user of AI provides a pre-existing input (e.g., photograph) for processing by AI without human involvement, the result of such work is not a work, and especially not a related work. Still, the consent of the owner of the copyright of the original work may be required (Wilczyńska-Baraniak, 2022).

In Poland, regulators additionally try to govern areas connected with AI. For instance, in the Polish Cloud Communiqué the Polish Financial Supervision Authority sets out recommendations for the management of cloud services. Entities using cloud computing need to ensure the adequate competency of employees and technical standards (e.g., ISO standards), have a plan for processing information, use cryptographic methods, and monitor the processing environment for insurance secrets or agency secrets (Wilczyńska-Baraniak, Rentflejsz & Dzięciołowski, 2022).

From the perspective of intellectual property protection regulations, there is also no legal definition of artificial intelligence on the international level, and in most countries a work created solely by AI is not subject to copyright protection at all given that only the result of human work can be considered a work. The ownership of works created by AI is thus an issue that virtually every country is pondering. The United Kingdom is one of the few countries to have regulated this aspect. UK law defines "computer-generated work" as work without human intervention. Such work is protected by copyright for 50 years (instead of 70 years for human-made work) (Wilczyńska-Baraniak, 2022).

The European Commission in turn proposed its own regulations in 2021 (the AI

Act). The aim of the draft Act is to establish uniform rules for classifying AI systems according to the level of risk. The main foundations of the new regulations are the overseeing role of human beings, technical security, privacy protection, non-discrimination and fairness, along with social and environmental well-being. The draft prohibits practices such as manipulation that causes physical or psychological harm, real-time biometric identification in public places for law enforcement purposes (with some material exceptions) or locating suspects of certain crimes. Also covered by the draft are AI systems of high risk, including those used in the biometric identification of individuals, critical infrastructure management, education and training, labour management, prosecution of crimes, management of migration, and administration of justice. Entities using high-risk AI systems will be required to run detailed data management and security systems, and ensure human supervision, or face sizeable penalties of up to EUR 30 million or 6% of annual global turnover. On 14 June 2023, the European Parliament adopted a position on the AI Act. Its priority is to ensure that AI systems used in the EU are safe, transparent, non-discriminatory and environmentally friendly. The EU also published, on 15 September 2022, a draft of the Cyber Resilience Act (CRA) to regulate digital resilience issues and strengthen the security of 'products with digital elements' (PDEs). The CRA imposes obligations to report any incident that affects the security of PDEs within 24 hours. This obligation applies for the lifetime of the PDE or 5 years after it was placed on the market, whichever is shorter (Wilczyńska-Baraniak, Rentflejsz & Dzięciołowski, 2022; European Parliament, 2023).

5 Challenges and opportunities for AI policy in Poland

Researchers predict the rapid application of AI technologies in many industries in Poland and around the world, bringing both great opportunities and threats for society. In this context, several aspects should be considered.

First, AI can significantly impact the labour market, employment, and human productivity. There are concerns that AI will eventually take over the duties of the worker and change the structure of employment (European Parliament, 2020). AI can lead to the automation of some occupations, especially less value-adding ones, and due to generative AI also challenge creative occupations (Rzeźnik, 2023). Orchard & Tasiemski (2023) claim that AI will affect the economy and job market in a revolutionary manner comparable with introduction of the Internet. They foresee that generative AI such as large language models might play the role of an analytical tool, assisting white-collar workers in business and life-critical decisions. General-purpose models can provide a quality service (such as copywriting) to customers accepting less creative content, leaving human writers as the premium service for other, more demanding clients. In turn, specialised models with access to specialist knowledge could provide a higher-quality service over that provided by human experts. Orchard & Tasiemski (2023) expect that while some jobs will be replaced by AI, new workplaces should also emerge – both highly and less specialised. Similarly, The Economist (2023) and Business Insider (2023) see AI as either a replacement of workforce across a wide range of industries and occupations or as a booster of worker productivity, akin to Schumpeter's 'creative destruction' by steam engines or computing. Quantitative forecasts in this respect are very meaningful. For instance, Goldman Sachs expects that at least 300 million jobs across the globe will be disrupted by the new technology. McKinsey forecasts that more than 12 million workers in the USA will be forced to switch to new occupations by 2030. Still, researchers assure that no AI has such an extensive memory as the human brain, and thus is unable to replace it completely (Miernik, 2023).

The European Parliament estimates that 14% of jobs in OECD countries can be highly automated, and another 32% may face major changes as a result of AI. At the same time, the European Parliament estimates that labour productivity will rise by between 11% and 37% due to the development of AI (Rzeźnik, 2023, pp. 7–8). In the Polish market, a survey by EY indicates that 59% of HR managers anticipate no layoffs as a result of AI, 13% are planning to increase employment, and only 3% think the workforce will be slightly reduced. Those in the customer service (37%) and industrial manufacturing (32%) industries face a higher risk of being replaced by AI-based tools (Miernik, 2023; Olak, 2023).

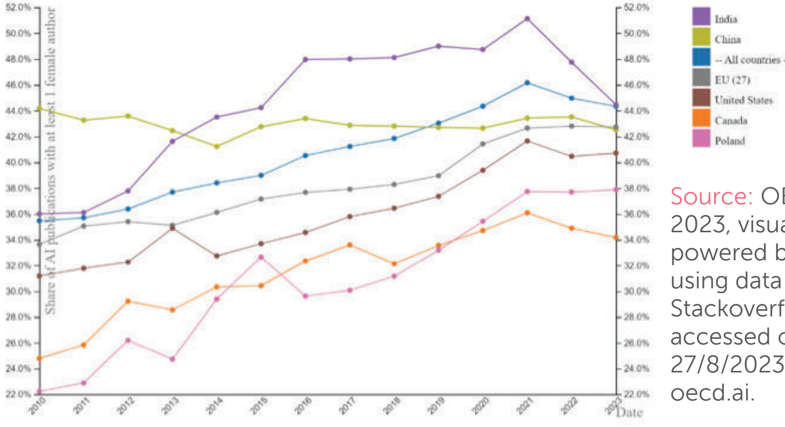
Considering the workforce engaged in developing AI in Poland, there are currently several challenges related to its structure and characteristics:

- The concentration of AI talent in Poland (understood as individuals who have both statistical modeling and big data computing skills) is relatively low by European standards. Only 3.1% of Europe's AI workers are located in Poland. This represents a huge disadvantage compared to Western Europe, where most AI talent is concentrated, especially the clear leaders in this regard: the UK with a share of 23.9% and Germany with a share of 14.1%. The situation seems even worse if the number of AI workers is considered relative to the active population (i.e. when measuring the ratio of the share in EU's AI talent to the share in EU's active population). In this ranking, Poland, with a ratio of 0.44, ranks only ahead of Slovakia, Austria, the Czech Republic and Hungary. The ratio for Poland is significantly lower than the level for leading European countries such as Ireland (3.5) and Finland (2.2) (Linkedin, 2019).
- The remuneration for AI professionals in Poland is relatively lower than the world average (only 2.4% of AI professionals earn more than USD 160 k in Poland compared to the world average of 15.6%), although people

working on the development of AI in Poland are more educated than the world average (70.7% of AI professionals in Poland have completed an advanced degree compared to only 57.5% as the world average) (Figure 4). This may reflect the fact that AI professionals are mostly young people (more than 50% are below 34 years of age in Poland; around the world, this age group constitutes only 43.9% of AI professionals).

- Almost the entire community of AI developers in Poland are men (with a share of 98%), similarly to the global perspective (on the global level 94.2% of AI professionals are male). Policy should address this aspect and promote the engagement of female IT specialists in the field of AI technology. The low female participation is also related to Polish researchers. The share of women in AI scientific publications, although on the rise, still does not exceed 38% and is below the world average of more than 44% – see Figure 5).
- The performance of AI researchers in Poland is relatively low. The number of publications, especially in high-quality journals, and the number of citations, is lower than for the top EU members in this respect. For instance, the cumulative number of publications of Polish researchers in the period 2020–2023 is only 24.2% of the figure for Germany (Table 4). In the case of high-quality publications and citations, the relative performance vis-à-vis Germany is even worse (16.4% and 10.8%, respectively).

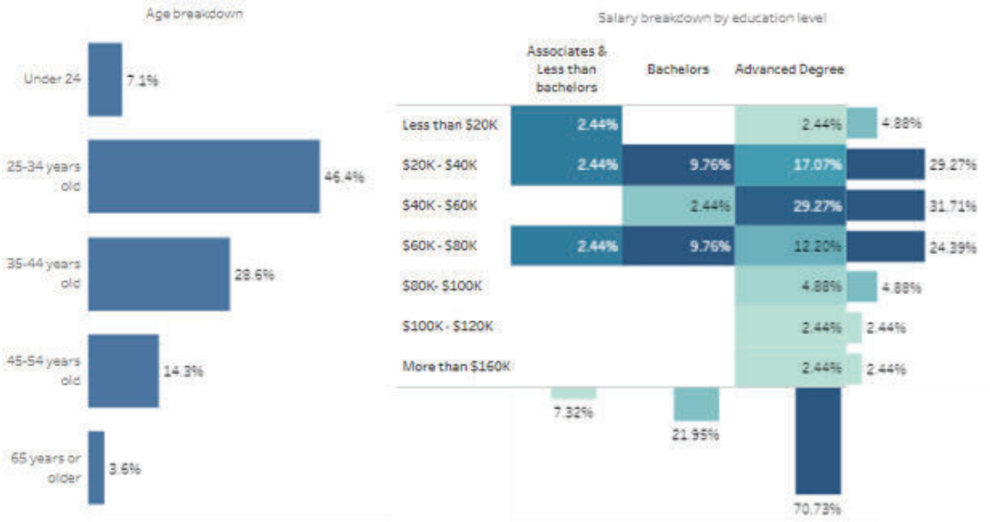
Figure 4: **Demographics of AI professionals in Poland by age in 2022**



Source: OECD.AI, 2023, visualisations powered by Tableau using data from Stackoverflow, accessed on 27/8/2023, www.oecd.ai.

Note: The figure presents the age, education and income breakdown of Stack Overflow Survey respondents. Bar size and colour correspond to the distribution of respondents in each category.

Figure 5: Share of women in AI scientific publications in Poland and other countries between 2010 and 2023



Source: OECD.AI, 2023, visualisations powered by JSI using data from Elsevier (Scopus), accessed on 29/8/2023, www.oecd.ai.

Note: The chart shows the share of female authors in AI over gendered authors. "Gendered author" refers to the subset of authors for which a gender could be assigned with a high level of confidence.

Table 4: AI research publications in top countries and Poland (cumulative number of publications between 2020 and 2023)

Country	No. of publications	Country	No. of high-impact publications	Country	No. of low-impact publications	Country	No. of citations
CHN	2,049,660	USA	619,383	CHN	1,561,801	USA	39,894,621
USA	1,742,545	CHN	471,233	USA	1,092,191	CHN	14,791,180
GBR	515,982	GBR	202,902	JPN	306,51	GBR	11,605,584
DEU	405,594	DEU	141,778	GBR	303,86	DEU	7,142,658
IND	395,012	CAN	109,732	IND	285,945	CAN	5,896,626
JPN	378,977	IND	103,849	DEU	256,247	AUS	4,587,359
FRA	315,882	ITA	93,967	FRA	222,619	FRA	4,545,792

Country	No. of publications	Country	No. of high-impact publications	Country	No. of low-impact publications	Country	No. of citations
CAN	290,894	AUS	93,742	CAN	175,786	ITA	3,927,032
ITA	261,003	FRA	88,145	ITA	162,289	JPN	3,096,348
AUS	233,568	JPN	67,108	AUS	135,748	IND	2,900,885
POL	**98,337**	**POL**	**23,23**	**POL**	**73,553**	**POL**	**773,661**
EU-27	**1,882,170**	**EU-27**	**599,077**	**EU-27**	**1,249,890**	**EU-27**	**27,819,140**
Average for top countries	**658,912**	-	**199,184**	-	**450,3**	-	**9,838,809**
Median for top countries	**386,994**	-	**106,791**	-	**271,096**	-	**5,241,993**
Poland as % of EU-27	**5.2%**	-	**3.9%**	-	**5.9%**	-	**2.8%**
Poland as % of DEU	**24.2%**		**16.4%**		**28.7%**		**10.8%**
Poland as % of the top country	**4.8%**	-	**3.8%**	-	**4.7%**	-	**1.9%**

Source: Own compilation based on OECD.AI, 2023; OECD.AI (2023), visualisations powered by JSI using data from OpenAlex, accessed on 28/8/2023, www.oecd.ai.
Note: Forecast for 2023.

Second, the development of AI requires significant investment in infrastructure such as communication networks, data centres and computer hardware, and in education, notably in fields related to computer science, mathematics, or natural sciences (Rzeźnik, 2023). Only large players will be able to meet these needs. As a result, AI might lead to a further deepening of discrepancies across the globe and global shifts in economic sectors. Global players like the USA and China are investing heavily in AI development, which could trigger shifts in the global economic balance. When it comes to AI investment in general, the USA is outperforming other nations. In 2022, the USD 47.4 billion spent in the USA was around 3.5 times the amount spent in China of USD 13.4 billion. Therefore, advocating for equal access to AI and fostering global cooperation in this field will be required (Rzeźnik, 2023, pp. 8–9, 13).

Third, ethical implications arise from AI and its use in society. AI poses many ethical dilemmas such as accountability for the decisions made by AI systems and potential discrimination resulting from algorithms. This technology might be used for unethical purposes such as 'deepfakes' which can be used to spread disinformation. This might be in the form of e.g., image manipulation and the creation of fake videos. Undoubtedly, content built in this way can influence public opinion. While deepfake was originally used to create funny videos of celebrities, it is an easy way to create manipulative messages and evidence in court cases or affect public affairs. AI can also lead to fraud by manipulating the voice of people, like the case when AI falsified the voice of a company's CEO and gave the CFO an order to transfer funds out of the company's account. Consequently, scientists stress that tools will have to be developed to identify fake videos, pictures, voices etc. to prevent unethical actions based on AI. Ethical guidelines on the use of AI are currently being developed to ensure its fair use (Maras & Alexandrou, 2019; Miernik, 2023; Wilczyńska-Baraniak & Walarus, 2022; Rzeźnik, 2023).

Fourth, the use of AI models might lead to issues with protection of the data used to train those models (the use of Big Data). Adequate regulations and safeguards are needed to protect personal data from unauthorised use. In Poland, the protection of personal data, also in the context of AI, is regulated by the General Data Protection Regulation on the Protection of Personal Data (RODO) and the Act of 10 May 2018 on the Protection of Personal Data. Polish companies are required to apply appropriate data protection measures so as to guarantee the privacy and security of users, both in Poland and while conducting business in EU countries (similarly, Google is obliged to comply with the regulations applicable in the EU, including the RODO, when offering European customers products that use AI, such as Google's voice assistant). One may expect that the legal framework related to AI will be evolving in both Poland and other countries and that those frameworks will vary significantly depending on a country's approach to privacy rights. Orchard & Tasiemski (2023) predict that in countries with less restricted use of personal data or greater state engagement in the processing of privacy-infringing data, AI technology will be increasingly used to control populations.

Last but not least, the environmental impact of AI technology should be analysed and monitored on an ongoing basis. The environmental costs of AI (stemming for example from the energy intensity of large models) should be compared to the benefits. This is a sizeable issue considering that the carbon dioxide emissions created by training a language processing model amount to the emissions needed to build and maintain five gasoline cars for 20 years (Rzeźnik, 2023). It seems natural that the engagement approach will need to be applied in this respect. In general, engagement as one of the two major ESG investment approaches, is a more rational strategy because it takes less time to implement and can bring more impactful results (Buks & Sobański, 2023).

6 Policy recommendations for Poland in the digital transformation era of AI

As AI is quickly becoming popular, there is a rapid need to develop detailed regulations to govern the use of AI and safeguard society. Consideration should be paid to multiple aspects, from accountability for decisions made by AI systems, data management by AI systems and its legal status, legal protection of AI models, to copyright of work created by AI systems. Regulating the legal status of AI creations is undoubtedly a major legislative challenge for regulators around the world, including Poland. Currently, very intensive work is underway in this regard, at the EU level, on the Artificial Intelligence regulation (the AI Act) and the AI liability directive, which will certainly also impact the Polish regulatory framework on AI. It is clear that specific regulations are required considering the huge potential for developing AI tools around the world and its deep impact across societies. Policymakers in Poland and the EU must consider several options in this respect, such as (Wilczyńska-Baraniak, 2022):

- establishing a separate law for AI-created works, with, e.g., a compensation requirement for authors of works that inspire AI-created works;
- defining a computer-created work, similarly to the UK approach; and
- making the formation of copyright dependent on the level of human participation.

An issue of considerable importance is to clearly define the responsibility for the actions taken by the AI algorithm. AI technologies can operate and make decisions autonomously, without human involvement. This means it is necessary to determine who is responsible for those decisions should damage be caused by AI: the AI developer, the data provider, the user, or another entity. There are presently no specific regulations in this area.

Ethics-based regulation is also needed to curb threats arising from AI. AI is meant to serve humans and hence any violations of basic human rights must be detected and eliminated. For example, AI algorithms must not lead to discrimination against certain groups of society. At the same time, regulations should be flexible to keep pace with ongoing changes in AI technology. This is the case with the draft EU regulation where the definition of AI is provided in an annex to allow it to be amended in a shorter procedure than the entire legal act. The golden rule for technology regulation is guidelines and recommendations to be based on generally accepted standards. The next important aspect is ensuring coordination across the world to develop common regulations for the field of AI. An undertaking along these lines was started by G7 countries in 2018 when they announced the Global Partnership on Artificial Intelligence, which in 2020 was joined by some members of the Organization for Security and Co-operation in Europe and UNESCO (Wilczyńska-Baraniak & Walarus, 2022).

Based on the current state of affairs, the recommendations below can also be formulated for AI users in Poland (Rzeźnik, 2023):

- Entities creating original AI software that is covered by copyright protection should place appropriate copyright markings on the software (e.g., in the source code) and apply appropriate technical safeguards (e.g., encryption or authentication) to restrict unauthorised parties from accessing your intellectual property.
- Apply for patent protection of your unique AI technologies at the Patent Office of the Republic of Poland in order to obtain exclusive rights to use your AI-based invention for 20 years from the date of application and license it to others. If the know-how related to the AI is not eligible for patenting, other security measures like non-disclosure and confidentiality agreements concerning company secrets might be used. Confidentiality agreements for employees developing AI-based tools should also be used.
- Entities that have AI-based patents should search patent databases and the market to determine their illegal use by other market participants and take appropriate legal actions if an infringement is noticed.

7 Conclusion

Artificial intelligence, aimed at automating decision-making processes, appears to be one of the most disruptive technologies of the current era of digital transformation. The technology is expected to bring huge economic benefits and an increase in living standards, adding over USD 15 trillion in value to the world economy and driving global investment of USD 200 billion a year by 2025. Today, the European Union is one of the largest venture capitalists in AI technology, surpassed only by the United States and China. Poland is considered a good prospect when it comes to the application of AI. It is home to globally respected developers and the first companies implementing AI-based solutions on a wider scale, such as SentiOne offering a chatbot alternative to the well-known ChatGPT. Despite the size of venture capital investment in AI in Poland not being significant by international standards and the share of companies using AI being half that for the world, investment in AI has been rising dynamically, especially since 2020. The Polish market is expected to benefit strongly from EU regulations requiring the local processing of the data used in AI models, which could prevent large players from the USA and China from entering the European market.

The policy on AI adopted by the Polish government in 2020 aims to support the public, companies, researchers and the government in taking advantage of AI, while preserving human rights and fair competition. Its specific goal is to increase the number of state-owned companies in Poland implementing AI projects. Support by the public programmes focuses on 13 national smart specialisations, i.e., industries rated the highest in terms of adding value to the economy and their competitiveness in foreign markets. Specialisations directly related to the development of AI are concentrated on "Smart networks and information, communication and geo-information technologies" and "Automation and robotics". In Poland, one can already find many examples of companies successfully offering products based on AI technology.

While AI technology is generally seen as benefitting humans, it poses serious challenges for policymakers in Poland and across the world. First, AI holds the potential to affect the labour market in a revolutionary manner and change the employment structure. Experts expect that one-third of employees will be disrupted by the new technology. Second, AI could further deepen economic disparities around the globe because only large global players like the USA and China will be able to meet the huge investment requirements of AI technology. Third, AI creates ethical dilemmas, such as accountability for the decisions made by AI systems, and unethical use of AI leading to discrimination or disinformation. Fourth, AI models might create problems with protecting personal data. Finally, the environmental impact of AI technology is not yet clear and must still be estimated. This is an enormous issue given that the carbon footprint of training a language processing model is equal to the emissions needed to build and maintain five gasoline cars for 20 years.

In Poland, no legal regulations directly refer to or define AI protection as such. Nevertheless, it is possible to obtain protection for its individual elements based on the regulations already found in the Polish legal order. For instance, software or databases, constituting an essential part of AI, can be legally protected under the Act of 4 February 1994 on Copyright and Related Rights and the Act of 27 July 2001 on the Protection of Databases, respectively. However, in Poland there is no legal protection of AI creations since a work subject to protection under the copyright law can only be created by a human. In the case of AI-generated inventions, it is also not possible to grant them patent protection as no individual can be designated the inventor. Currently, AI creations are becoming part of the public domain, while their use is not subject to any restrictions.

The fact that AI is rapidly gaining in popularity establishes a need for detailed and specific regulations to govern its use and protect the public. Policymakers need to think over several aspects, from accountability for the decisions made by AI systems, data management by AI systems and its legal status, legal protection of AI models, to copyright of the work created by AI systems. Very intensive work is presently underway in this respect on the EU level, which is certain to also affect the Polish regulatory framework on AI. In general, policymakers in Poland and the EU will have to decide on several aspects such as: establishing a separate law for AI-created works, defining a computer-created work, or making the formation of copyright dependent on the level of human participation. Ethics-based regulation is also in need to curb threats arising from AI to human rights. In the meantime, AI users in Poland are forced to apply the legal framework in place on copyright, patents, databases and know-how, and seek legal protection where ever possible.

REFERENCES

- AI Policy. (2020). Polityka dla rozwoju sztucznej inteligencji w Polsce od roku 2020 [Policy for the development of artificial intelligence in Poland from 2020]. Annex to Resolution No. 196 of the Council of Ministers of December 28, 2020. (Item 23). Retrieved 10 August 2023 from: https://www.gov.pl/web/ai/polityka-dla-rozwoju-sztucznej-inteligencji-w-polsce-od-roku-2020.
- Aristovnik, A.; Kovač, P., Murko, E., Ravšelj, D., Umek, L., Bohatá, M., Hirsch, B., Schäfer, F.-S., & Tomaževič, N. (2021). The Use of ICT by Local General Administrative Authorities during COVID-19 for a Sustainable Future: Comparing Five European Countries. Sustainability, 13, 11765. https://doi.org/10.3390/su132111765.
- Buks, A. G., & Sobański., K. (2023). Divest or Engage? Effective paths to Net Zero from the U.S. perspective". Economics and Business Review, 9(1), pp. 65-93, https://doi.org/10.18559/ebr.2023.1.3.
- Business Insider. (2023). AI is going to eliminate way more jobs than anyone realizes. 14 August. Retrieved 1 September 2023 from: https://www.businessinsider.com/ai-radically-reshape-job-market-global-economy-employee-labor-innovation-2023-8?IR=T.
- Duszczyk, M. (2023). Sztuczna inteligencja w Polsce rośnie w siłę [Artificial intelligence is going from strength to strength in Poland]. Rzeczpospolita, 31 May 2023. Retrieved 2 September 2023 from: https://cyfrowa.rp.pl/biznes-ludzie-startupy/art38547381-sztuczna-inteligencja-w-polsce-rosnie-w-sile-oto-firmy-ktore-sie-licza.
- Economist. (2023). The AI boom: Lessons from history. https://www.economist.com/finance-and-economics/2023/02/02/the-ai-boom-lessons-from-history.
- European Parliament. (2020). Sztuczna inteligencja: szanse i zagrożenia [Artificial intelligence: opportunities and threats]. Retrieved 24 August 2023 from: https://www.europarl.europa.eu/news/pl/headlines/society/20200918STO87404/sztuczna-inteligencja-szanse-i-zagrozenia.
- European Parliament. (2023). Regulacje ws. sztucznej inteligencji: oczekiwania Parlamentu [Artificial intelligence regulation: Parliament's expectations]. Retrieved 25 August 2023 from: https://www.europarl.europa.eu/news/pl/headlines/society/20201015STO89417/regulacje-ws-sztucznej-inteligencji-oczekiwania-parlamentu.
- Goldman Sachs. (2023a). Global Macro Research. Generative AI: Hype, or truly transformative? Issue 120, 5 July 2023. Retrieved 1 September 2023 from: https://www.businessinsider.com/ai-radically-reshape-job-market-global-economy-employee-labor-innovation-2023-8?IR=T.

- Goldman Sachs. (2023b). AI investment forecast to approach $200 billion globally by 2025. Published 1 August 2023. Retrieved 1 September 2023 from: https://www.goldmansachs.com/intelligence/pages/ai-investment-forecast-to-approach-200-billion-globally-by-2025.html.
- Goldman Sachs. (2023c). Generative AI could raise global GDP by 7%. Published 5 April 2023. Retrieved 1 September 2023 from: https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html.
- GovTech. (2023). Polityka rozwoju AI w Polsce przyjęta przez Radę Ministrów – co dalej? [AI development policy in Poland adopted by the Council of Ministers – what's next?]. Retrieved 10 August 2023 from: https://www.gov.pl/web/govtech/polityka-rozwoju-ai-w-polsce-przyjeta-przez-rade-ministrow--co-dalej.
- Jareno, F., & Yousaf, I. (2023). Artificial intelligence-based tokens: Fresh evidence of connectedness with artificial intelligence-based equities. International Review of Financial Analysis, 89, 102826. https://doi.org/10.1016/j.irfa.2023.102826.
- Klekowski, T. (2023). Jak sztuczna inteligencja może przyspieszyć transformację sektora energetycznego [How artificial intelligence can accelerate transformation of the energy sector]. Report of the Digital Transformation Observatory THINKTANK. Retrieved 2 September 2023 from: https://think-tank.pl/wp-content/uploads/2023/02/raport-eng.pdf.
- KPMG. (2023). Sztuczna inteligencja w firmach w Polsce: potencjał do wykorzystania [Artificial intelligence in companies in Poland: Potential to exploit]. 26 July 2023. Retrieved 2 September 2023 from: https://kpmg.com/pl/pl/home/media/press-releases/2023/07/media-press-sztuczna-inteligencja-w-firmach-w-polsce-potencjal-do-wykorzystania.html#:~:text=Dynamicznie%20ro%C5%9Bnie%20znaczenie%20i%20popularno%C5%9B%C4%87,korzysta%20z%20tego%20typu%20rozwi%C4%85za%C5%84.
- Linkedin (2019). AI Talent in the European Labour Market. Linkedin Economic Graph. Retrieved 21 November 2023 from: https://economicgraph.linkedin.com/content/dam/me/economicgraph/en-us/PDF/AI-TAlent-in-the-European-Labour-Market.pdf
- Maras, M.-H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. The International Journal of Evidence & Proof, 23(3), 255–262. https://doi.org/10.1177/1365712718807226
- Miernik, A. (2023). Sztuczna inteligencja - czym jest? Czy AI wpłynie na sytuację na rynku pracy w Polsce? [Artificial intelligence - what is it? Will AI affect the labor market situation in Poland?]. EY Poland. 7 June 2023. Retrieved 2 September 2023 from: https://www.ey.com/pl_pl/workforce/sztuczna-inteligencja-ai-i-rynek-pracy-w-polsce.

- Ministry of Development and Technology. (2023). National Smart Specializations. List of national smart specializations (effective as of February 13, 2023). Retrieved 2 August 2023 from: https://www.gov.pl/web/rozwoj-technologia/krajowe-inteligentne-specjalizacje#:~:text=Lista%20krajowych%20inteligentnych%20specjalizacji,od%2013%20lutego%20 2023%20r.)&text=KIS%20polega%20na%20okre%C5%9Bleniu%20 priorytet%C3%B3w,jej%20konkurencyjno%C5%9Bci%20na%20rynkach%20 zagranicznych.
- NSS. (2023). National Smart Specializations. Retrieved 2 August 2023 from: https://smart.gov.pl/pl/.
- OECD. (2021). Venture Capital Investments in Artificial Intelligence: Analysing trends in VC in AI companies from 2012 through 2020. OECD Publishing, Paris. Retrieved 2 August 2023 from: https://www.oecd-ilibrary.org/science-and-technology/venture-capital-investments-in-artificial-intelligence_f97beae7-en.
- OECD.AI. (2023). OECD.AI Policy Observatory. Live data. OECD. Retrieved 2 August 2023 from: https://oecd.ai/en/trends-and-data.
- Olak, R. (2023). Badanie EY: rozwój sztucznej inteligencji nie wpływa na plany pracownicze polskich firm [EY study: development of artificial intelligence does not affect employee plans of Polish companies]. EY Poland. 15 May 2023. Retrieved 1 September 2023 from: https://www.ey.com/pl_pl/news/2023/05/rozwoj-si-nie-wplywa-na-plany-pracownicze-polskich-firm.
- Orchard, T., & Tasiemski, L. (2023). The Rise of Generative AI and Possible Effects on the Economy. Economics and Business Review, 9(2). https://doi.org/10.18559/ebr.2023.2.732.
- Ramaswamy, S. (2023). How companies are already using AI. Harvard Business Review, 14 April 2017. Retrieved 28 August 2023 from: https://hbr.org/2017/04/how-companies-are-already-using-ai.
- Rzeźnik, G. (Ed.). (2023). Zastosowanie sztucznej inteligencji w gospodarce. Przegląd wybranych inicjatyw i technologii z rekomendacjami dla przedsiębiorców [Application of artificial intelligence in the economy. Review of selected initiatives and technologies with recommendations for entrepreneurs]. Thematic report No. 3. Polska Agencja Rozwoju Przedsiębiorczości. Retrieved 2 September 2023 from: https://www.parp.gov.pl/storage/publications/pdf/Raport-tematyczny_zastosowania_sztucznej_inteligencji_w_gospodarce_20230616.pdf.
- Wilczyńska-Baraniak, J. (2022). Do kogo należy „dzieło” stworzone przez sztuczną inteligencję? [Who does the ‘work’ created by artificial intelligence belong to?]. EY Poland. 17 October 2023. Retrieved 1 September 2023 from: https://www.ey.com/pl_pl/law/do-kogo-nalezy-dzielo-stworzone-przez-sztuczna-inteligencje.
- Wilczyńska-Baraniak, J., & Walarus, J. (2022). Sztuczna inteligencja – zagrożenie

czy zbawienie? Podsumowanie debaty oksfordzkiej: Kto się boi sztucznej inteligencji [Artificial intelligence - threat or salvation? Summary of the Oxford debate: Who is afraid of artificial intelligence]. EY Poland. 2 June 2023. Retrieved 1 September 2023 from: https://www.ey.com/pl_pl/law/sztuczna-inteligencja-podsumowanie-debaty-oksfordzkiej.

- Wilczyńska-Baraniak, J., Rentflejsz, O., & Dzięciołowski, M. (2022). Jakie szanse i ryzyka niesie za sobą wdrożenie sztucznej inteligencji? [What are the opportunities and risks of implementing artificial intelligence?]. EY Poland. 5 October 2022. Retrieved 1 September 2023 from: https://www.ey.com/pl_pl/biuletyn-ryzyka/jakie-szanse-i-ryzyka-niesie-za-soba-wdrozenie-sztucznej-inteligencji.

Chapter 9

The current state of AI Policy in Poland and comparison with the EU

Konrad Maj

Paulina Grzybowicz

1 Introduction

Artificial Intelligence (AI) lies at the heart of the contemporary digital revolution, significantly influencing the future of work, education, healthcare and numerous other facets of daily life. Within both Poland and the broader European Union (EU), AI is recognised as a pivotal instrument for innovation, competitiveness and sustainable development, reflecting a proactive response to the dynamic shifts and global technological trends presently underway.

In comparison to its EU counterparts, Poland is distinguished as a leading investor in AI. However, it must also expedite efforts to bridge existing gaps, particularly in the realms of private investment and international research funding.

It is worth mentioning that many global companies such as Aptiv, byteDance, Capgemini, Intel, TomTom, IBM, Google, nVidia, Roche, Ringier Axel Springer, Samsung, T Mobile and TCL, have opened research and development efforts in Poland (Aipoland, 2022). This shows Poland's great development potential in the digital area.

This chapter seeks to delineate the current status and trends pertaining to AI in Poland, juxtaposed with activities and strategies observed across the EU. Through analysis of Poland's AI policy, the article identifies principal challenges and opportunities before offering recommendations in different areas and for different stakeholders.

2 Key data for AI trends in Poland

The introduction of artificial intelligence solutions can positively impact a range of socio-economic processes. It can enable the analysis of large data sets, the automation of many processes and personalisation of services in many sectors of the economy.

Poland plays a leading role in the field of AI within the Central and Eastern European region, being ranked 7th in the EU in terms of the number of experts engaged in AI. Implementation of AI solutions in the Polish economy could annually add 2.65% to the GDP growth rate.

Poland has invested considerable resources in the development of AI, with a strong emphasis on research, development and innovation. The government has introduced various financial initiatives and support programmes for technology startups and AI-related research (Wieczorek, 2023).

Attached to Resolution No. 196 of the Council of Ministers dated 28 December 2020 (item 23) was a document entitled "Policy for the Development of Artificial Intelligence in Poland from 2020" (gov.pl, 2021). The policy describes how Poland is emerging as a significant player in AI, aiming to become an international leader with achievements grounded in its citizens' intellectual capabilities. Polish 15-year-olds scored 516 points in mathematical reasoning in the PISA study, outperforming the OECD average by 27 points, with only Estonia and the Netherlands scoring higher among European countries. Such educational prowess provides a robust foundation for AI development, with Poland producing over 110,000 STEM graduates annually, ranking fourth in the EU.

The country has contributed to key AI projects globally, with Polish experts having participated in developing renowned AI solutions like OpenAI, PyTorch, FastText, Flo, Inception-v3, and AlphaStar. These solutions are acknowledged and utilised around the world, underscoring Poland's presence in the global AI landscape.

Poland's AI potential is further accentuated by its economy's strong reliance on electronic data, contributing to 46% of its GDP. Further, over 33% of the population belongs to the creative class, higher than in the USA, Spain, Japan, and comparable to Italy and Israel. This creative potential is crucial for innovative AI research and applications.

AI's introduction is poised to have a substantial impact on various sectors in Poland, with priority given to public administration, construction (especially smart buildings), cybersecurity, energy, retail, healthcare, industry, agriculture, transport, and logistics. The benefits of implementing AI in these sectors contribute approximately 2.65% to Poland's total GDP. It is predicted that this AI economic growth will continue, both in Poland and other EU countries. The AI market is forecasted to grow by 18.44% in Poland by 2030 (Statista Search Department, 2023), and should the European Union continue its technological trajectory, it could add around 2.7 trillion euro, or 20%, to its combined economic output, by the same year (Bughin et al., 2019).

AI and automation integration is anticipated to considerably influence the Polish job market, creating 130 new jobs for every 100 existing ones and potentially automating up to 49% of work by 2030. This brings both opportunities and challenges, calling for adjustments in the education system and the development of strategies to address possible technological unemployment while improving overall job quality.

Global AI investments are led by countries like the USA, China, France and the UK. In contrast, Poland's public sector and major state-owned companies play a pivotal role due to the limited number of large private enterprises in the country. Poland is actively engaged in discussions to redesign AI initiatives and tools to enhance human capital investment, facilitate hardware and software acquisition, support R&D, translate research into production, invest in strategic infrastructure, and promote open data sharing.

On the international stage, Poland calls for substantial AI funding, suggesting that it be included in the EU's Multiannual Financial Framework for 2021–2027. It supports AI development funding through various EU funds and programmes, pushing for proportional fund distribution based on each country's economic size to maximise beneficiaries and ensure sustainable development across the EU.

In 2019, the Polish government launched the GovTech Poland project. It deals with technological innovations in the public sector, helping to find and solve technological problems, including the use of AI solutions. GovTech deals with public procurement, participating in technical dialogue between contractors and offices.

GovTech Poland will be supported in the area of AI as part of the Digital Sandbox project. The aim of the project is to create a test environment and obtain information for GovTech. Thanks to the knowledge acquired in Digital Sandbox, it will be possible to implement new or improved public administration services using API interfaces.

GovTech Poland is aligned with the EU: representatives of GovTech participate in EU work, For instance, the EU White Paper outlining the advancement of artificial intelligence (AI) in Europe, a strategy concerning European data, or the continuous public discussions regarding the Digital Services Act (DSA). In 2020 GovTech Poland submitted an application to the European Parliament to scale the project and start international competitions with administrations and innovators from other European Union Member States (gov.pl, n.d.).

The Polish Agency for Enterprise Development (Polska Agencja Rozwoju Przedsiębiorczości – PARP) also supports companies in adopting AI. This year, in cooperation with experts from SWPS University, PARP has prepared a compendium of knowledge for entrepreneurs on the use of AI in companies entitled "Applications of artificial intelligence in the economy. Review of selected initiatives and technologies with recommendations for entrepreneurs". The report discusses basics aspects like what artificial intelligence is, the current trends and challenges of AI in the economy, the legal regulations regarding the protection of intellectual property, and provides practical information on AI applications in individual areas of National Smart Specialisations (Krajowe Inteligentne Specjalizacje – KIS) through examples of companies and startups using this technology for research, production and testing of products or services.

It is worth explaining that National Smart Specialisations are industries recognised as priorities for creating innovative socioeconomic solutions, increasing the added value of the Polish economy and making it more competitive in the international arena.

The PARP report also contains legal recommendations for companies regarding the protection of AI intellectual property. Experts suggest securing copyrights to AI software, including documenting the creative process, as well as using technical safeguards and appropriate contractual clauses with employees. It further recommends reporting innovations to patent offices and protecting know-how as a trade secret, which can be supported by confidentiality agreements and technological security measures, along with preparing licence agreements regulating the licensing of AI technologies.

Science is an important element of the AI ecosystem. According to Digital Poland, Polish universities published over 12,000 articles on artificial intelligence between 2013–2018 and approximately 20,000 Polish students begin studies in computer science each year (Aipoland, 2022). In addition, around 28,000 graduates graduate in various technical fields annually, while 4,000 graduate in mathematics. The main centres of artificial intelligence research in Poland are concentrated in large cities, in regions such as Masovia (especially Warsaw), Lesser Poland (especially Krakow), Silesia (especially Katowice and Częstochowa).

AI is used in many sectors in Poland. In the financial sector, AI helps in risk analysis and portfolio management. In healthcare, it is used to diagnose and personalise treatment plans. The education and manufacturing sectors also rely on AI to optimise processes and deliver personalised solutions. These technologies are also used by local governments and state institutions. For example, one city in the Lublin Voivodeship – Świdnik – uses AI in waste management and the National Health Fund uses AI to find inaccuracies in hospital bills for medical services (Aipoland, 2022).

Currently, an important project is the national cloud, which aims to accelerate the digital transformation of Polish enterprises and public institutions, along with the development of AI technology in Poland. The project is being implemented by the largest Polish bank – PKO BP, together with the Polish Development Fund, as well as Microsoft and Google. Microsoft itself contributed USD 1 billion to the project. Increasing the supply of computing power and storage space will help accelerate the project's implementation. This aligns with the European Union's progression toward providing access to reliable, sustainable, and compatible cloud infrastructures and services (European Commission, 2023b)

Some Polish companies using AI are doing well in international markets. An example is Cosmos AI. This company connects the best aspects of offline and online shopping to create a seamless experience for shoppers and increase sales for retailers. Cosmos AI employs world-class talents in its offices in Paris, Warsaw, Singapore, Hong Kong, Shanghai and Tokyo, including winners of the international programming competitions ACM ICPC and IOI. Cosmos AI empowers some of the world's most prominent companies like LVMH, Richemont, L'Oréal, and Estée Lauder, whilst also offering AI-driven recommendations to its users encouraging them to shop in nearby stores, saving time, money and the environment.

Poland is a leader in video games, a tech field aligned with AI development skills, including programming, adaptability and creativity. The globally acclaimed game The Witcher exemplifies this, having won 250+ awards and reaching over 150 million users through games, books and a TV series, also serving as a tool for Polish diplomacy.

When it comes to social attitudes, Polish society generally holds a positive stance on technology. For example, a study conducted this year on behalf of ING Bank Śląski showed that 93% of Poles believe that technologies make life easier, and more than half are interested in super applications (Kijowski and Borycka, 2023). Further, 71% of respondents feel that technology gives them greater control, and 60% believe that it expands their knowledge and possibilities of action. According to the respondents, technology saves time (64%), facilitates various activities (44%), simplifies procedures (34%) and is available at any time (53%). In addition, 72% of respondents want technology to give them choices, while 68% believe that humans should supervise the use of AI in finance. Interestingly, most are willing to pay more for replacement services. The greatest concerns are associated with personal data, including the lack of control over data and the consequences of it being leaked. Only necessary data consents are provided by 89% of respondents. The results also show that when it comes to data security Poles have the greatest trust in banks and medical facilities.

In general, these follow and even exceed EU trends, reflected in the most recent Eurobarometer survey on technologies in which 75% of respondents thought most recent digital technologies have a positive impact on the economy, 67% on their quality of life, and 64% on society at large. 86% of EU citizens think that the overall influence of science and technology is positive, and 61% believe that technologies like artificial intelligence will have a positive effect in the future (McDonnell et al., 2022) .

When it comes to digital skills and AI readiness, there is a notable disparity between Western/Northern and Eastern/Southern Europe. For instance, countries such as Finland and the Netherlands have 79% of their populations with at least basic digital skills, whilst countries like Italy and Hungary are at 45.6% and 49.1%, respectively. According to this study, 43% of people in Poland have at least basic digital skills, with 3.6% of employees being digital experts (European Commission, 2023a).

In order to help EU member states increase digital skills in Europe, the Digital Skills and Jobs Coalition was launched in 2016. The goal of the coalition is to increase digital education for citizens, the labour force, and ICT specialists (European Union, n.d).

Poland has also placed a strong emphasis on education and training in the field of technology and AI, offering a variety of programmes and courses at universities and online platforms, as the Polish labour market is starting to notice growing demand for AI experts and data analysts. The Digital Competence Development Programme specifically aims to introduce issues related to digitisation and AI into school curricula and focus on digital skill development for ICT sector specialists and employees of small and medium-sized enterprises. The programme operates within the framework of the EU Digital Skills and Job Coalition. Cybersecurity is also to be a key element of these programmes (Jākobsone, 2021)

Private initiatives are also gaining momentum. The CAMPUS platform was launched in September 2023. The generative platform CAMPUS AI is an educational and research initiative focused on artificial intelligence (Campus AI, 2023). Its aim is to integrate various fields of science, technology and business to collaboratively work on the development and application of AI. The platform may offer a diverse range of courses, workshops, research projects and networking opportunities for students, researchers and professionals. It serves as a hub where theory meets practice, and innovations are nurtured and developed. Expanding on this, the CAMPUS AI platform may be seen as a multidisciplinary ecosystem that brings experts from different sectors together. By so doing, it seeks to accelerate the pace of AI research and its practical applications. The platform could serve as a catalyst for cutting-edge AI projects, providing the resources and expertise needed to take ideas from conception to implementation. It might also act as a bridge between academia and industry, facilitating the translation of research findings into real-world solutions. In this way, CAMPUS AI could play a crucial role in shaping the future landscape of AI in terms of both technological advancements and ethical considerations.

3 Current state of AI policy in Poland

The European Union has defined its own AI strategy aimed at increasing public and private investment in AI, preparing for changes in the labour market, and creating ethical guidelines. This strategy serves as a framework for member states' national strategies.

However, many countries have developed their own national AI strategies. For instance, in March 2018 France presented its artificial intelligence strategy called "AI for Humanity". With an allocated budget of EUR 1.5 billion, the strategy intends to strengthen France's position as a global pioneer in the field of AI. The initiative's four main pillars are: public health, environmental protection, transport, and defence issues.

Analysis of EU strategies and regulations as well as national strategies provide valuable tips for Poland. In general, Poland's AI strategy is consistent with the EU's main goals and priorities EU, although there are differences in financing, regulation and educational initiatives.

Poland's AI strategy is outlined in the previously mentioned AI Policy, adopted in December 2020. This foundational document seeks to assist various societal sectors in capitalising on AI, stressing human dignity and fair competition protection. It lays out around 200 goals for development of the AI sector, addressing obstacles like the inadequate collaboration between academia and business. Solutions include strengthening academia–business ties and fostering knowledge, innovation and productivity.

This document defines short-, medium- and long-term actions and goals in the area of AI, concentrating on the development of Poland's society, economy and science. The policy is divided into six key areas: AI and society, AI and innovative companies, AI and science, AI and education, AI and international cooperation, and AI and the public sector (see Table 1). In the AI Policy, additional tools have been added to each goal to achieve individual goals.

The policy aims to support diverse groups, including society, businesses, scientists and public administrations, in taking advantage of the opportunities offered by AI, while ensuring the protection of human dignity and fair competition in the global arena. Account is taken in the policy of international, legal, ethical and technical-organisational aspects, specifying the requirements and conditions for the effective use of AI throughout the technology lifecycle.

The AI Policy suggests law amendments for AI ecosystem functionality, proposing regulation on a legal definition of AI, the denial of AI legal personality, personal data ownership and portability, AI manufacturer liability based on diligence, and AI operator liability based on risk.

Table 1: **Short-, medium- and long-term goals for AI in specific areas**

Area	Short-term Objectives (by 2023)	Medium-term Objectives (by 2027)	Long-term Objectives
AI & Society	• Effectively preventing and mitigating the negative consequences of the development of AI for the labour market. Starting dialogue with the market in order to introduce protective measures, preceded by socioeconomic analysis. • Analysing the ethical ramifications of AI implementation and the impact of AI systems on the sphere of human rights • Ensuring security and building public trust and willingness to use AI-based solutions combined with democratising access to AI • Launching campaigns to prepare society for changes related to the adoption of a data-driven economy model (algorithmic economy) • Making Poland an attractive country for highly skilled AI experts and workforce	• Analysing and eliminating legislative barriers and administrative burdens for artificial intelligence startups • Taking action in specific areas linked to the development of AI, in particular for efficient and easy access to data and its use by all economic actors, regardless of size • Supporting programmes preparing society for the changes brought by the development of an algorithmic economy in Poland • Preventing unemployment and flexible job creation in the labour market for disadvantaged groups • Defining regular programmes for supporting artistic and creative activities in the area of AI	• Poland is one of the biggest beneficiaries of the data-driven (algorithmic) economy • Poles are aware of the opportunities and threats brought by the development of modern technologies and make career choices based on them, using a wide range of educational materials and dedicated curricula • Poland is among the top-10 countries in the AI Readiness Index • Poles foster a culture of lifelong learning and the ability to quickly re-skill, while the government policy curbs technological unemployment • Poles are prepared to consciously and critically use AI-based systems • Poles exposed to AI-based systems, especially in the public sphere, are aware of their rights and have access to mechanisms that protect them from system errors or other violations of their rights and freedoms

Area	Short-term Objectives (by 2023)	Medium-term Objectives (by 2027)	Long-term Objectives
AI & innovative Companies	• Increasing the demand for AI-based solutions • Increasing the supply of AI-based solutions developed in Poland • Increasing the number of Polish state-owned companies implementing AI projects • Increasing the use of new AI-based technologies by companies operating in Poland • Identifying talents, especially teams that develop innovative AI-based solutions • Creating knowledge bases and developing good practices for implementing and using AI-based solutions	• Increasing the number of companies providing AI-based solutions, including those listed on the Warsaw Stock Exchange • Ensuring Poland is perceived as a leader in implementation projects and scientific research in the agri-food sector • Ensuring Poland is perceived internationally as a developer of AI systems • Increasing the competencies of Polish managers in the field of AI	• Poland has at least one globally recognised Polish company operating in the field of AI • There are Polish technology companies listed simultaneously on the Warsaw Stock Exchange and one of the world's largest stock indices • The Polish economy has a significant level of venture capital investment from both private and public funds, covering all stages of small business growth • Poland is among the top 25% of economies producing innovative AI-based solutions
AI & Science	• Disseminating practical knowledge concerning AI at the undergraduate and graduate level in teaching activities and research. • Developing projects tailored to Polish problems and challenges, such as machine processing of the Polish language and its translation into foreign languages, through research cooperation between Slavic language-speaking countries and involvement of Polish speakers at foreign universities	• Establishing ties between academia and business. • Increasing the attractiveness of Polish universities for the most talented students and academic staff through, among other things, a more flexible course plan and openness towards interdisciplinary classes in AI and new technologies • Internationalising higher education and doctoral training with two-way movement of doctoral students between countries • Increasing visibility of research in international markets.	• Polish universities are internationally competitive in terms of AI-related educational offering. • Polish scientists are often nominated for the most important industry awards, including the Turing Prize. The number of publications in leading journals and conferences (e.g., Conference on Neural Information Processing Systems, Conference on Computer Vision and Pattern Recognition, Association for Computational Linguistics conferences) in the field of AI exceeds the OECD average. • The number of patents for AI obtained by Polish inventors exceeds the OECD average

Area	Short-term Objectives (by 2023)	Medium-term Objectives (by 2027)	Long-term Objectives
AI & Education	• Disseminating practical knowledge of artificial intelligence in all stages of education • Supporting the development of the most talented school and university students in Poland • Developing educational materials about AI for different age and professional groups • Using the National Educational Network in interactive education on AI-based solutions and techniques	• Implementing a comprehensive educational curriculum on AI in primary and secondary schools, with support for customised education • Boosting the image of Poland as an attractive place to acquire qualifications and develop skills in the AI field thanks to competitions on the national and international levels	• Poland is the European leader in education in AI and other digital technologies on the secondary school level • Polish students are at the forefront of educational research in Europe (as measured by PISA and others) • Poland co-organises mathematics and AI competitions on the European and global levels
AI & International Cooperation	• Creating an environment that fosters international investment in innovative ventures implemented in Poland • Strengthening cooperation within the EU, NATO, the Three Seas Initiative, Visegrád 4, the Weimar Triangle, and the UK, Switzerland and Norway • Increasing the international visibility of Polish research teams • Identifying priority areas where Poland has a chance of being internationally competitive	• Actively cooperating with other countries on innovation, and the development of new technologies and AI • Coordinating Polish action plans with broader European and international initiatives • Promoting EU international funding programmes for AI development • Building the international image of Poland as an innovative country, open to new technologies	• Poland is positioned as a country with a significant role in the creation and broad application of AI-based solutions internationally • Poland has innovative AI centres of excellence and testing that collaborate internationally with public and private sectors • Poland has a long-term strategy for AI development, taking the situation and European and global regulations in this field into account

Area	Short-term Objectives (by 2023)	Medium-term Objectives (by 2027)	Long-term Objectives
AI & the Public Sector	• Effectively coordinating all work and activities related to development of the Polish AI ecosystem • Developing rules ensuring transparency, auditing and accountability concerning the use of algorithms by public administration • Developing regulations aimed at obtaining public APIs from public and municipal enterprises with access to the widest possible catalogue of up-to-date data, respecting the principles of personal data protection and the priority of improving the quality of public services • Enhancing the ability of the state to use AI in emergency situations to forecast threats and support decision-making, as well as in situations requiring intervention or support from various government bodies on different levels • Using AI-based solutions for continuous monitoring and improvement of the environment in Poland • Maximising the research potential of medical data to improve citizens' health, taking account of the protection of privacy and personal data with or without the use of privacy protection techniques (e.g., anonymisation or pseudonymisation) if the explicit consent of the data subject is present	• Ensuring public data is available and widely used • Ensuring Poland is one of the most active countries in developing the ethical use of data according to the concept of trustworthy AI.4	• Polish public data is accessible and easy to use for citizens, researchers and industry. The data are adapted to machine analysis and accessible via modern APIs. The release of public data respects the laws on the protection of classified information, business secrets, the protection of personal data, and the free movement of and access to non-personal data. • The rights and interests of Polish citizens whose data may be used by researchers or industry are secured with appropriate guarantees (including but not limited to the protection of their privacy) • Transparent mechanisms for sharing non-public data are developed • Polish diplomacy promotes Polish AI businesses and scientific centres • Thanks to the work of Polish specialists and Polish international activities, Poles are among the leading authors cited in AI publications

AI & the Public Sector	• Increasing the number of AI procurements in the public sector, including central and local government, as well as state-owned companies and municipal companies run by local government bodies, thanks to the development of tools developed by the GovTech Polska Programme • Taking advantage of Poland's role as host of the 2020 UN Internet Governance Forum organised in Katowice to exchange experiences and promote Poland in the area of modern technologies and artificial intelligence		
	• Effectively preventing and mitigating the negative consequences of the development of AI for the labour market. Starting dialogue with the market in order to introduce protective measures, preceded by socioeconomic analysis. • Analysing the ethical ramifications of AI implementation and the impact of AI systems on the sphere of human rights • Ensuring security and building public trust and willingness to use AI-based solutions combined with democratising access to AI • Launching campaigns to prepare society for changes related to the adoption of a data-driven economy model (algorithmic economy) • Making Poland an attractive country for highly skilled AI experts and workforce	• Analysing and eliminating legislative barriers and administrative burdens for AI startups • Taking action in specific areas related to the development of AI, in particular for efficient and easy access to data and its use by all economic actors, regardless of size • Supporting programmes preparing society for changes brought by the development of an algorithmic economy in Poland • Preventing unemployment and flexible job creation in the labour market for disadvantaged groups • Defining regular programmes for supporting artistic and creative activities in the area of AI	• Poland is one of the biggest beneficiaries of the data-driven (algorithmic) economy • Poles are aware of the opportunities and threats brought by the development of modern technologies and make career choices based on them, using a wide range of educational materials and dedicated curricula • Poland is among the top-10 countries in the AI Readiness Index • Poles foster a culture of lifelong learning and the ability to quickly re-skill, while the government policy curbs technological unemployment • Poles are prepared to consciously and critically use AI-based systems • Poles exposed to AI-based systems, especially in the public sphere, are aware of their rights and have access to mechanisms that protect them from system errors or other violations of their rights and freedoms

Source: Based on the Policy for the Development of Artificial Intelligence in Poland from 2020 (gov.pl, 2021).

Many universities and scientific institutions are opening various research centres and studies in AI. In 2019, the top-10 Polish universities founded the AI Tech Scientific Consortium with the aim to educate specialists in the field of ICT, viewing education in the field of AI, machine learning and cybersecurity as the most important.

At the beginning of 2023, the Information Processing Centre – National Research Institute presented a report commissioned by the Ministry of Education and Science entitled: "Artificial Intelligence: publication achievements in the field of science and technology in 2010–2021".

The report reveals that with respect to the European Union most publications on AI in science and technology came from Germany (36.9 thousand), France (28.6 thousand), Italy (27.9 thousand) and Spain (26.6 thousand). These publications had an average citation level in the world (MNCS = 1), with 11 EU-27 countries distinguishing themselves with a bigger influence. Publications from Denmark, the Netherlands and Germany were cited 32%, 27% and 26% more often, respectively. The least cited works were from Latvia and Bulgaria. A total of 18% of publications from the EU-27 were among the 10% most cited papers globally, with Germany and Italy in the lead, also noting that 39% of EU-27 publications were produced in international cooperation. Although this model unfortunately usually increases the impact of works, countries such as Latvia, Poland and Romania are the least likely to publish in international cooperation, while the leaders in this respect include: Luxembourg, Denmark, Belgium and the Netherlands.

International cooperation was responsible for 27% of the 3.8 thousand publications by Poles on AI in science and technology, mainly involving researchers from the USA, Canada, Great Britain and China, with a high level of impact (MNCS = 2.37 from the USA and 2.04 from China). Among the EU-27 countries, Poland's biggest partners are Germany and Italy, with publications being cited 21%–23% more often. Poland also cooperates effectively with Ukraine, India, Saudi Arabia and Australia, achieving an above-average level of influence.

Between 2010 and 2021, with 13,959 scientific works in the field of AI in science and technology Poland took 5th place in the EU and 19th in the world. These publications accounted for 7.2% of EU and 1.2% of global works in this area. In 2015, the biggest increase in the number of papers was recorded (+24%), exceeding 1,000 publications per year for the first time. During the period under study, the works of Polish scientists accounted for 6% to 9% of the EU's publication output in this field. In 2021, there were 1,851 publications from Poland, representing a 2.5-fold increase over 2010.

Acknowledging global AI trends, Poland's government has initiated legislation supporting the development of AI. However, no binding provisions for AI, blockchain, or big data currently exist in Poland. Issues like AI liability and copyright, with AI not recognised as the creator or owner, are governed by general legal frameworks, with AI producers or operators typically held responsible (Wieczorek, 2023). Most European Union countries face similar issues, but the EU is well into the process of working on the AI Act, a proposal for regulating AI in member states by addressing safety, transparency, trustworthiness, data governance, and human oversight. The act is due to be adopted in 2024 (European Parliament, 2023).

At the same time, the Polish government's establishment of the Digitisation Committee is intended to help coordinate the implementation of IT projects and the coherence of IT projects with the state's strategic activities, including compliance with the Integrated State Computerisation Programme, the assumptions of the state's information architecture and the National Interoperability Framework.

4 Challenges and opportunities for AI policy in Poland

Poland and other European Union countries must also further develop and adapt their regulatory framework to be consistent with the EU's AI requirements and standards. International cooperation, notably with EU partners, can help Poland achieve its AI goals by exchanging knowledge, experiences and best practices.

Both in Poland and throughout the EU, considerable funds are being allocated to support enterprises in the field of AI. Still, the fact is that about 40% of startups in Europe, despite claiming to deal with AI, have little to do with artificial intelligence (Kelnar, 2019). The term is used for promotional purposes or to acquire customers and investors (Przegalińska & Jemielniak, 2023). This creates a problem in reliably assessing the actual level of development and sophistication of artificial intelligence.

Using AI establishes technological and legal challenges, including determining liability for AI errors – whether the creator, user, data provider, or other entity should be held liable. There are no legal regulations on this matter. Further, it is necessary to design AI in a way that avoids bias and discrimination, assuring that decisions are based on fairness and equality. This also raises the question about the status of AI as a creator of works such as images, music or texts. These types of challenges are shared across the entire European Union.

Ethics and privacy are also at the forefront of AI challenges in Poland. AI implementation must comply with ethical principles and protect user privacy. Poland has defined its own position on EU regulations regarding artificial intelligence, including in the document entitled "Poland's position in the consultations on the White Paper on Artificial Intelligence – a European approach to excellence and trust", presented on 12 July 2020 (Europe

Commission, 2020). The paper explains that Poland supports the European Commission's EUR 100 million pilot scheme for funding the development of AI , alongside the expected increased support through the InvestEU Programme starting in 2021. The country emphasises the importance of investment strategies that ensure geographical cohesion amongst EU regions and calls for an increase in public funding. The aim is to attract over EUR 20 billion in total AI investment annually across the EU in the upcoming decade. Poland is also calling for the creation of an EU platform (EU GovTech) which would aggregate demand for modern technologies from public institutions. This platform would facilitate contracts for microenterprises while promoting the GovTech sector's development.

In terms of private sector partnerships, Poland endorses the formation of public–private collaborations in the domains of AI, data and robotics. This endorsement extends to cooperation with research centres and innovation hubs. Poland suggests that priority areas should be expanded to include agriculture, transport and logistics. It also advocates for the development and implementation of AI solutions in public services that are transparent and effective.

Regarding Digital Innovation Hubs (DIH), the European Commission plans to establish specialised AI hubs in each member state. Yet, Poland opposes the idea of a flagship research centre, proposing instead a network of various research centres across the EU. This approach aims to prevent the neglect of smaller institutes, particularly those located in Central and Eastern Europe.

On the matter of data access and infrastructure, Poland underscores the importance of data quality and supports the establishment of a decentralised data space. The country also highlights the need to have high-capacity network infrastructure, emphasising the crucial roles of 5G networks and high-performance and quantum computers in the development of AI.

A clear and coherent regulatory framework is deemed necessary, serving to build confidence in AI, promote its widespread use, and address the risks associated with its application. Poland advises that the investment approach should be prioritised over regulation. Regulatory efforts should focus on minimising potential damages and risks without stifling innovation. The framework should also incentivise voluntary controls and certifications, ensuring human oversight throughout the AI system lifecycle.

In relation to high-risk AI, Poland agrees with implementing a risk-based approach to regulation and insists on clear criteria to define "high-risk" AI applications. The country supports the inclusion of additional sectors and applications in the high-risk category, calling for transparency and traceability in AI systems, especially those used in state security and law enforcement.

The described document also underscores the importance of establishing a mutually recognised liability framework within the EU, covering the design and application of AI. It supports a model where responsibilities are distributed among AI creators, developers and operators based on their ability to control risks and comply with ethical AI guidelines.

For the development of AI, cooperation in fundamental research, education, and societal development is essential. Poland is in favour of international collaborations with non-EU countries that share the values of the EU. The country also seeks the removal of trade restrictions, promoting data access in trusted spaces.

Finally, the proposal for a European AI governance structure is well received provided that it complements, without duplicating, the competencies of existing bodies. Overall, Poland supports the development of AI that is in line with EU values, promoting cooperation among various stakeholders both domestically and internationally. The focus is on developing human-centred, ethical AI that fosters trust in public services and facilitates international collaboration.

In Poland, there is currently widespread concern about the use of AI by various services related to state security and the police. While boosting their operations' effectiveness, concerns arise regarding their impact on civil rights and freedoms. The dilemma has two sides: while powerful surveillance tools like electronic correspondence monitoring and citizen databases aid in crime detection, careless use of them may violate constitutional rights (Dworzecki & Nowicka, 2021).

Changes in the labour market and the need for education and training are detected in both Poland and other EU countries. Strategies to support workers affected by automation and investments in education and retraining are required.

Other countries, similar to Poland, are developing educational programmes related to AI and struggling with a shortage of qualified specialists. This is a common problem that probably requires coordinated action on the EU level.

The AI Policy outlines the problems that AI may cause in the labour market. While the use of solutions based on artificial intelligence will lead to a decrease in employment in many sectors of the economy, in the long term it will bring an increase in overall employment as well as an increase in labour productivity. It is said that, by 2030, as much as 49% of working time in Poland may be automated using existing technologies; instead of 100 existing jobs, 130 new ones will appear.

Governments bear the responsibility for preparing and retraining workers who risk becoming unemployed following the implementation of AI, for aligning their new skills with market needs. This effort requires adjustments to legal and educational frameworks, as outlined in Poland's in-development Integrated Skills Strategy 2030. The strategy aims to respond flexibly to technological advancements by fostering a legal environment conducive to research and development, creating economic models, eliminating barriers, and enhancing the legal system's readiness for market shifts.

However, a survey conducted by EY Poland in May 2023 shows that the majority of Polish companies (59%) do not plan changes in employment linked to the development of AI in the next 2 years, while 13% are still in the analysis phase (Olak, 2023). Despite the growing importance of AI, 43% of companies do not anticipate changes in their operating models. Innovations such as shortening working time or delegating tasks to AI are still distant. Only 7% of organisations had not encountered technological, legal or organisational barriers. People working in customer services (37%) and industrial production (32%) are most at risk of being replaced by AI. Due to intense discussions about AI, we can expect that ever more companies will begin to analyse possible changes in their employment plans and business strategy related to technology.

In turn, the National Reconstruction Plan (Krajowy Plan Odbudowy – KPO), which was approved by the European Commission, talks about plans to introduce the 'digital default', i.e., the primacy of electronic documents over paper ones, which will force the digitisation of a number of processes in companies and institutions.

By using AI, Poland can improve the quality and efficiency of public services, from education to healthcare, which in turn will benefit its citizens.

In 2018, the Coalition for AI in Health was established, which brings together several dozen different organisations dealing with health issues. The organisation seeks to shape policy regarding AI in the Polish healthcare system. The group aims to create an environment that facilitates the implementation of AI innovations in healthcare while maintaining the central role and trust in medical professionals. The coalition is involved in activities popularising modern medical solutions, in both regulatory and technological aspects, and also conducts research on the implementation of AI in medicine. As part of its activities, it published the White Paper on AI in clinical practice.

The "Business Digital Transformation Monitor" study conducted in Poland by KPMG in partnership with Microsoft shows that as many as 6 out of 10 companies that use AI in their activities do not monitor the effectiveness of its implementation.

The KPMG survey data show that only 15% of surveyed organisations confirmed having made investments in AI, whereas the global average is around 35%–37%. Still, just like in other parts of the world, we do not sufficiently measure the benefits arising from AI; in Poland, 62% of organisations do not do it, and the global average is 68%). This is presently the biggest global challenge facing AI adoption.

AI is particularly used in business areas such as marketing and production. However, KPMG anticipates stronger adoption in the coming years in the area of customer and employee service. This is expected to result from the wider use of multi-tasking language models (e.g., ChatGPT). The metaverse, edge computing and blockchain are relatively new and not well understood by enterprises in Poland with respect to their impact on market competitiveness.

The study showed that cloud solutions are used by approximately 70% of organisations, with the most frequently used AI solutions having the following nature:

- Mobile: Deployed in 73% of organisations, 23% plan to do so within 1 year
- Computer Decision Support: implemented in 70% of organisations, 35% plan to implement it within 1 year
- Automation and Robotisation: Implemented in 58% of companies, planned in 14%
- Machine-to-Machine Communication: Implemented in 39%, planned in 17%.

When it comes to new/unknown technologies, the data are as follows:

- Metaverse: In 7% of companies, planned in 1%
- Edge Computing: 13%, planned in 3%
- Blockchain: 11%, planned in 5%.

Mobile solutions are most often implemented by companies from the information technology, media and communication sectors (90%) and the financial sector (85%). The financial sector is also a leader in the implementation of processes based on automation and robotisation (63%), communication between machines (63% of implementations) and the Internet of Things (53%).

It also turns out that Polish companies hold different priorities when it comes to the digital transformation.

- For 32% of companies, the priority is Sales and Marketing
- For 30% of companies, it is Customer Services
- For 24% of companies, it is Internal Operations
- For 23% of companies, it is Operational Management and Production.

There are also differences between sectors. In the Finance and Life Sciences sector: 50% of companies plan intensive digitalisation in the area of customer services. In turn, digitisation of the Sales and Marketing area is of great importance for the Finance (45%) and Life Sciences (43%) sectors.

The Life Sciences sector anticipates dynamic digitisation in the areas of Internal Operations (46% of companies plan digitisation) and Operational Management (seen as a priority by 43% of companies), overtaking other sectors in these fields.

It is apparent that the development of AI also poses huge challenges to the IT industry. Even considerable experience in this sector does not guarantee employment. According to a survey of managers conducted by the Sectoral Council for IT Competencies and Antal, just 20% of representatives of IT companies consider their current competencies to be sufficient (Paluszyński, 2023). Even though AI can automate a number of tedious tasks, it requires specialists to manage data and processes. The study showed that 85% of respondents believe that the development of AI and machine learning will significantly increase the demand for new competencies. They indicate the need for skills related to machine learning (68%), Python (61%) and experience of working with data science and AI libraries (56%). Despite the growing popularity of the 'low-code solutions' (without requiring extensive coding skills), the demand for programmers with soft skills, such as the ability to think analytically and solve problems, remains high. The dynamic development of technology, new competency requirements in the IT sector, constant changes and the need for additional training may cause many psychological difficulties among specialists working in the field of AI, and perhaps even reluctance to work in this industry and to leave it.

The above data from the company survey perfectly complement the report on the attitudes held among Polish society to AI. In October 2023, the Digital Poland Foundation presented the next edition of the report "Technology in the service of society. Will Poles become society 5.0?".

The Digital Poland study indicates that 88% of Poles know the term "AI", but only 56%, after being presented with the OECD definition, declare that they understand or use the technology. Despite some knowledge gaps, AI is present in the everyday lives of many respondents.

A positive attitude to AI is held by 85% of Poles. Society is divided in its assessment of the benefits and risks of AI; 24% see more benefits, 27% see more risks, while 35% believe they are balanced. A significant share of respondents (64%) admit that they lack knowledge about AI, and their biggest concern is data privacy (61%).

Trust in AI and willingness to share data are divided, with 33% trusting, 30% not trusting, and the rest undecided. Respondents emphasise the need for AI supervision, privacy protection and cybersecurity, and the younger generation and educated people are more aware of ethical and legal issues.

Opinions are also divided over the pace of AI development in Poland. When it comes to institutions developing AI, Poles most trust public universities (30%), international research organisations (29%) and large technology companies (22%). Polish companies, the government and local governments enjoy much less trust.

Society is divided on the impact of AI on the labour market. Although 42% fear that AI may cause layoffs, others see the potential to create new jobs. Young people and people familiar with AI are more open to using this technology.

With respect to women in the field of AI, Poles are undecided. While 32% believe that AI can support women in combining careers with motherhood, opinions are divided on whether the lack of women in the technology sector is harmful. In this regard, men and the younger generation are more positive.

In turn, in terms of regulation, 40% of Poles believe that the current rules regarding AI are not sufficient for its safe development, with greater concern being seen among the residents of large cities and educated people. Self-regulation in this field is supported by 22% and 46% want stricter regulations by the European Union, even at the expense of being dependent on technology from the USA and China. While the main areas for regulation are cybersecurity, surveillance and data privacy, there is low awareness of AI-related copyright issues. After clarification, 38% are against the free use of online content to train AI without respecting property rights. Respondents are open to the use of AI by the public sector to a limited extent, yet only 6% are against any role of AI in public administration.

5 Policy recommendations for different stakeholders

Based on the data collected and analysed, we present 10 key recommendations for the sustainable development of AI in Poland in relation to various areas and stakeholders.

To maximise the potential of AI, Poland must retain and attract AI specialists and experts while safeguarding human rights, following a European approach focused on ethical, human-centric AI. This approach seeks to promote the ethical use of AI, provide access to AI technologies and their benefits while minimising associated social, economic and political risks. Active participation in international organisations like the EU and the UN is crucial for Poland to contribute to the development of ethical and regulatory frameworks for AI, addressing issues concerning human dignity, rights, and the practical implementation of these values in AI evaluation criteria for trustworthiness and responsibility.

AI education and workforce training. Education is the bedrock of the development of AI. Developing comprehensive AI curriculums at both universities and vocational centres is paramount. The aim should not only be to nurture new talents but also to offer reskilling programmes for the existing workforce. Moreover, promoting AI literacy among the general public will foster an environment in which AI can thrive.

Fostering innovation and supporting startups. Support for startups and small businesses is vital. Implementing tax incentives and providing grants will act as catalysts for innovation. The creation of innovation hubs and incubators will encourage collaboration, while ensuring that data access for AI training strikes a balance between development needs and privacy concerns.

Ethical and regulatory framework. Establishing an ethics committee will guide the nation's approach to AI, ensuring that it is used responsibly and ethically. This committee would oversee the development of guidelines and standards, providing a clear regulatory framework that supports innovation while protecting individuals and society at large.

EU collaboration and support for member states. Aligning AI strategies among the EU member states will foster a collaborative approach to the development and implementation of AI. Financial and technical support for member states will facilitate this collaborative approach, aiding in the development and execution of AI policies across the continent.

Responsible and ethical AI development. The commitment to developing AI ethically and responsibly is non-negotiable. Having EU-wide ethical standards and guidelines will ensure a uniform approach to the development and use of AI, promoting ethical considerations across all member states.

Supporting national AI initiatives. Promotion and support of AI solutions developed within Poland are essential. This involves facilitating cooperation between different sectors and implementing mechanisms that enhance transparency and accountability in AI systems. Protecting citizens' rights and freedoms is of the utmost importance, with special emphasis on supporting AI in healthcare.

Cybersecurity and disinformation. Backing initiatives that focus on cybersecurity and counteracting disinformation is essential in the age of digital misinformation. Active support for projects in this realm will foster a safer and more trustworthy digital environment for all.

Continuous improvement and adaptation. With the rapid pace of AI development, policies need to be reviewed and updated regularly. Poland should actively participate in international AI organisations, contributing to and learning from the global dialogue on AI ethics and regulations.

Promoting "invented in Poland". The "invented in Poland" brand should be actively promoted, highlighting the country's innovative AI solutions. This not only fosters national pride but also positions Poland as a significant player in the global AI landscape.

Public and private sector collaboration. An additional key recommendation is to foster deeper collaboration between the public and private sectors. This partnership is crucial for driving innovation and ensuring the effective implementation of AI technologies. Joint initiatives, shared resources, and a collaborative approach to tackling challenges will accelerate the development and adoption of AI in Poland, benefiting the entire nation.

6 Conclusion

Presented in December 2020, Poland's AI Policy is evidence of the country's proactive approach, containing short- and long-term initiatives to guide the integration of AI into society, the economy, science and other fields, while paying attention to ethical and legal standards. Despite the clear lack of binding legal provisions regarding AI in Poland, domestic investment in education and research on this technology is significant, as shown by initiatives like the AI Tech Scientific Consortium Tech and the considerable number of scientific publications in the field of AI coming from Poland.

Even though international cooperation, as well as financial support from companies, is extremely important for the development of applications and research and development in the field of AI, it does not resolve the problem of the excessive and unfair promotion of startups and companies operating in this area. Due to civil rights issues, Poland is also challenged by social concerns with the use of AI by various security services and the police. Another problem is anticipating changes in the labour market caused by automation and robotisation and preparing appropriate government responses, such as in the form of investments in education and training.

An important task of the authorities is to constantly monitor the changes caused by AI and the introduction of other new technologies, while supporting public-private partnerships and promoting national solutions based on AI. Without doubt, Poland holds enormous potential to not simply adapt its policy to the visions developed within the EU, but also to mark its presence as a significant player in the field of AI in the global market.

REFERENCES

- Aipoland. (2022). Public Policy. Retrieved 1 5 September 2023 from: https://aipoland.org/public-policy/
- Aipoland. (2022). R&D and Universities. Retrieved 1 5 September 2023 from: https://aipoland.org/rd-and-universities/
- Bughin, J., Seong, J., Manyika, J., Hämäläinen, L., Windhagen, E., & Hazan, E. (2019, February 7). Tackling Europe's gap in digital and ai. McKinsey & Company. https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-europes-gap-in-digital-and-ai
- Campus AI (2023). CampusAI. Retrieved 15 September 2023 from https://www.campusai.pl
- Dworzecki, J., Nowicka, I. (2021). Artificial Intelligence (AI) and ICT-Enhanced Solutions in the Activities of Police Formations in Poland. In: Visvizi, A., Bodziany, M. (eds) Artificial Intelligence and Its Contexts. Advanced Sciences and Technologies for Security Applications. Springer, Cham. https://doi.org/10.1007/978-3-030-88972-2_11
- European Commission. (2023b). Cloud computing. Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/policies/cloud-computing#:~:text=The%20European%20Commission%20aims%20to,integral%20part%20of%20the%20goal
- Europe Commission. (2020). White Paper on Artificial Intelligence – a European approach to excellence and trust. Retrieved 7 July 2023 from: https://digital-strategy.ec.europa.eu/en/consultations/white-paper-artificial-intelligence-european-approach-excellence-and-trust
- European Parliament. (2023, June 14). EU AI act: First regulation on artificial intelligence: News: European parliament. EU AI Act: first regulation on artificial intelligence. News | European Parliament. https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence
- European Union . (n.d.). Digital Skills and Jobs Coalition. Digital Skills and Jobs Platform. https://digital-skills-jobs.europa.eu/en/about/digital-skills-and-jobs-coalition
- gov.pl. (2021). The Policy for the Development of Artificial Intelligence in Poland from 2020. Retrieved 14 August 2023 from: https://www.gov.pl/web/ai/polityka-dla-rozwoju-sztucznej-inteligencji-w-polsce-od-roku-2020
- gov.pl. (n.d.). GovTech Poland: International cooperation . Website of the Republic of Poland. https://www.gov.pl/web/govtech-en/wspolpraca-miedzynarodowa

- Kelnar, D. (2019). The State of AI 2019: Divergence. Retrieved 7 July 2023 from: https://www.stateofai2019.com/introduction/#:~:text=As%20Artificial%20Intelligence%20(AI)%20proliferates,for%20entrepreneurs%20is%20also%20changing.
- Kijowski, S., Borycka, M. (2023). Polacy entuzjastami w podejściu do nowych technologii. Pod względem bezpieczeństwa danych najbardziej ufają bankom. Newseria Biznes. Retrieved 23 September 2023 from: https://biznes.newseria.pl/news/polacy-entuzjastami-w,p546611149
- Jākobsone , M. (2021, May 19). Poland - Digital Competence Development Programme (2020-2030). Digital Skills and Jobs Platform. https://digital-skills-jobs.europa.eu/en/actions/national-initiatives/national-strategies/poland-digital-competence-development-programme
- McDonnell, A., Verdin, R., O`Reilly, J. (2022). EU Citizens' attitudes to digitalisation and the use of digital public services: Evidence from Eurobarometers and eGovernment Benchmark EUROSHIP Working Paper No. 12. Oslo: Oslo Metropolitan University. DOI: 10.6084/m9.figshare.19188227. Available at: https://euroship-research.eu/publications
- Olak, R. (2023). Badanie EY: rozwój sztucznej inteligencji nie wpływa na plany pracownicze polskich firm. EY. Retrieved 18 August 2023 from: https://www.ey.com/pl_pl/news/2023/05/rozwoj-si-nie-wplywa-na-plany-pracownicze-polskich-firm
- Paluszyński, W. (2023). Sztuczna inteligencja wymusza zmianę i podnoszenie kompetencji. Potrzebne będą nowe kategorie specjalistów IT. Newseria Biznes. Retrieved 29 September 2023 from: https://biznes.newseria.pl/news/sztuczna-inteligencja,p1152689602
- Przegalińska, A., & Jemielniak, D. (2023). Strategizing AI in Business and Education: Emerging Technologies and Business Strategy (Elements in Business Strategy). Cambridge: Cambridge University Press. doi:10.1017/9781009243520
- Statista Search Department (2023). Artificial Intelligence - Poland. Statista. https://www.statista.com/outlook/tmo/artificial-intelligence/poland
- Wieczorek, L. (2023). Poland: Artificial Intelligence – Current Landscape and Developments. CEE Legal Matters Magazine, 9(10).

Chapter 10

AI Policy and Governance in the European Union

Eva Murko

Dejan Ravšelj

1 Introduction

Digital transformation is one of the core initiatives of the European Union (EU) given that digital innovations hold the potential to generate cascading impacts across economic sectors. The Digital Decade policy programme is among the six focal priorities the European Commission has established for the EU's future (European Commission, 2023a). This year's report on the state of the Digital Decade for 2022 showed the EU's digital transformation journey was notably shaped by several escalating trends like heightened climate concerns and related social and economic worries, an increasing need for high-speed connectivity, rising threats to democracy and EU principles and, lastly, rapid advancements in fields like artificial intelligence (AI) (European Commission, 2023b).

Simultaneously, emerging geopolitical challenges, particularly Russia's aggression towards Ukraine, have led to economic uncertainties. These geopolitical dynamics, accentuated by strategic divergences and values, have resulted in increased living costs, more cyberattacks in Europe, and disrupted supply chains. Central to these challenges is the role of digital technologies. Their rapid evolution, intertwined with a fierce global technological race, has the potential to establish global digital leaders and reshape the EU's competitive stance, growth and sovereignty (European Commission, 2023b). The landscape of AI development, its policy and regulation is an important factor in the global technological race and it is continuously evolving. Each region is attempting to balance the promotion of AI innovation with the need to address ethical, social, and economic concerns. The differences in approaches reflect broader geopolitical and economic strategies and could potentially lead to divergent AI development paths with global implications.

In light of the increasing importance of AI, the primary rationale of this year's ELF project is to examine the current AI landscape and initiatives in the EU and explore their role within the context of the human-centric society. It is crucial to understand the governance and regulation of AI, the challenges and opportunities for the EU to leverage the digital evolution and AI, identify good practices and scalable solutions to support a future-oriented Europe and maintain a stable and healthy economy.

In April 2018 (European Commission, 2018a), the European Commission announced a EUR 1.5 billion investment in AI research through 2020, with a larger goal for the EU – both public and private sectors – to invest at least EUR 20 billion in AI R&D by 2020, and the same amount annually over the next decade. The Commission aims to make Europe a key player in AI by supporting research, innovation, and adopting AI technologies, particularly among small and medium-sized enterprises. It also emphasised the importance of modernising education to adapt to the digital transformation and job shifts caused by AI. In addition, the Commission stressed that new technologies should align with EU values like human dignity, democracy, and respect for human rights. The Commission is committed to creating a framework for AI that encourages innovation while upholding these values and aims to be a global leader in setting ethical AI standards.

With its emphasis on ethical AI and regulatory clarity, the EU seeks to position itself as a global leader in "trustworthy AI". The AI landscape in the European Union EU is multi-faceted, reflecting the EU's approach of balancing rapid technological advancement with ethical considerations, data privacy, and the protection of its citizens.

Hence, this paper presents a short overview of AI in the second section before turning in the third section to the key AI policy in the EU and the AI Act. The fourth section briefly describes the challenges of AI governance, the fifth section addresses the impacts of AI on society and perceptions of EU citizens while the final section brings everything together and provides concluding remarks.

2 Briefly about AI

Artificial Intelligence has a long history, dating back to the 1950s when the British mathematician Alan Turing first considered the possibility of machines able to think (Turing & Haugeland, 1950). The term "artificial intelligence" was then introduced in 1955 by Dartmouth maths professor John McCarthy as a neutral term to describe this emerging field (Siebel, 2019). Despite early efforts in AI, the lack of computing power and underdeveloped mathematical concepts and techniques saw a decline in interest and funding for AI research during the 1970s. This period is referred to as the "AI winter" (OECD, 2019; Siebel, 2019). However, in the 2000s the field of AI experienced a resurgence due to a confluence of factors, including the rapid growth of computational power, the rise of the Internet and the massive amount of data it provided, as well as significant advancements in the mathematical foundations of AI, notably in the area of machine learning (Siebel, 2019). With these developments, AI has become a focus of policy attention as governments seek to invest in and regulate its development and use (Kuziemski & Misuraca, 2020).

For the first time, self-adapting algorithms are being applied in various contexts such as industrial processes, data analytics, and everyday activities like advanced mobile devices and autonomous vehicles. Economically and socially, AI amplifies both industrial capabilities and technological innovation, leading to elevated productivity, better public services, and improved quality of life. However, its long-term development and societal impact warrant scrutiny and the mitigation of associated risks (Righi et al., 2022). The European Union intends to be at the forefront of ethical, secure and cutting-edge AI development, advocating a human-centred approach on a global scale (Misuraca & Van Noordt, 2020).

What is AI? There is no single, universally accepted definition of AI (yet), and instead several different ones. Some are formulated based on the disciplines for which AI systems are used and others on lifecycle phases (Berryhill et al., 2019). Wirtz et al. (2019) studied different definitions of AI and proposed an integrative definition for AI as the ability of a computer system to perform human-like intelligent behaviour and problem-solving with the help of certain core competencies, including perception, understanding, action and learning. In line with this, the authors' understanding of an AI application refers to integrating AI technology into a computer application field with human-computer interaction and data interaction (Wirtz et al., 2019). Artificial intelligence encompasses various technologies such as machine learning, neural networks, natural language processing etc. and can be defined as a technology for advanced prediction (Agrawal et al., 2017). AI technology identifies patterns in large amounts of data to predict outcomes for similar instances (Dwivedi et al., 2019).

AI is already having a transformative impact on our lives, mostly positively, such as boosting productivity, enhancing safety, and improving healthcare (Stone et al., 2016). Its potential is far-reaching, offering the ability to make more cost-effective and accurate predictions, decisions and recommendations. Interestingly, some of the most notable advancements in AI are outside computer science in fields like biology, medicine, finance and healthcare. Economically, AI is transitioning into a general-purpose technology like computers did in the 1990s, extending its influence beyond specialised industries to the broader economy and society (OECD, 2019).

3 Key policy documents and the AI act

Since 2016, there has been a global discussion among various stakeholders about how to develop and regulate AI in a socially beneficial way while mitigating the risks. The global state of AI policy and regulation is a dynamic and rapidly evolving field, particularly in major regions like the European Union (EU), the United States (USA), and China. Each of these regions has its own approach to AI governance, reflecting their different political, economic, and social priorities.

In the United States, Biden's executive order on AI, issued on October 30, establishes guidelines for security and privacy, expanding upon the voluntary pledges made by over a dozen companies. As part of the Biden-Harris Administration's comprehensive strategy for responsible innovation, the Executive Order builds on previous actions the President has taken, including work that led to voluntary commitments from 15 leading companies to drive safe, secure, and trustworthy development of AI (The White House, 2023).

In China, companies cannot develop AI services without obtaining the necessary permissions. Effective from August 15, a series of 24 guidelines issued by the government specifically addresses generative AI services like ChatGPT, which produce content including images, videos, and text. According to these guidelines, content created by AI must be clearly marked and comply with regulations concerning data privacy and intellectual property rights (Library of Congress, 2023).

Within this context, the European Union has been actively considering its role in the development of AI, initially driven by concerns about falling behind North America and Asia in technological innovation. Various EU entities have shaped the region's AI policy, covering areas like the digital market, internal affairs, and research. This discourse has been marked by numerous key policy documents and reports, fuelling an ongoing debate on the EU's approach to AI (Ulnicane, 2022).

3.1 AI strategy for Europe

In October 2017, the European Council invited the Commission to determine a pan-European approach to artificial intelligence (European Council, 2017). Heeding this call and aligning with the Parliament's resolution, the Commission inaugurated the European Union's strategic AI blueprint in April 2018: "Artificial intelligence for Europe" (European Commission, 2018a).

It is imperative to contextualise this initiative by noting that during this period a multitude of organisations and nations globally had either already formalised or were on the cusp of introducing their AI strategies (Ulnicane et al., 2021a, 2021b; Ulnicane, 2022). The document articulately captures this prevailing sentiment by emphasising that "…AI… (is) one of the most strategic technologies of the 21st century. The stakes could not be higher. The way we approach AI will define the world we live in. Amid fierce global competition, a solid European framework is needed" (European Commission, 2018a). The document also underscores its rationale by referencing the recent AI strategic initiatives and financial commitments undertaken by notable global players, namely the United States, China, Japan and Canada (European Commission, 2018a).

To fully harness the potential of AI and effectively address the emerging related challenges, the AI strategy states that the EU must adopt a unified and coordinated approach. By leveraging its intrinsic values and strengths, the EU is poised to champion the ethical and inclusive development and application of AI for the collective benefit of all (European Commission, 2018a).

According to the document, several factors place the EU in an enviable position to pioneer this endeavour:

- The EU boasts a premier cadre of researchers, cutting-edge laboratories, and innovative startups. Moreover, the region has a strong foothold in the field of robotics and stands as a global leader in pivotal industries, including but not limited to transport, healthcare and manufacturing. Such industries are pivotal and should spearhead the integration of AI technologies.
- The Digital Single Market within the EU acts as a bedrock for digital innovation and expansion. Establishing unified regulations,

encompassing areas like data protection, seamless data flow within EU borders, cybersecurity, and enhanced connectivity, not only facilitates businesses' trans-border operations but also adds to investor confidence.

- In addition, the EU possesses extensive industrial, research and public sector data. If effectively harnessed, this data may serve as the foundation upon which advanced AI systems can be built. Recognising the intrinsic value of this data, the Commission has concurrently initiated measures to simplify data-sharing protocols and make data more accessible for reuse. This is particularly evident in the efforts to disseminate data related to public utilities, environmental statistics, research findings, and health data.

Thus, by capitalising on these assets and fostering a harmonised approach, the EU can lead the global discourse on the responsible and equitable evolution and deployment of AI.

Three objectives are outlined in this strategy, having grown to become the cornerstones of the EU's AI policy.

(1) Enhancement of Technological and Industrial Competency: The initiative seeks to strengthen the EU's technological and industrial capability in AI. It aspires to expedite the integration of AI across various economic sectors, encompassing both private and public sectors. It involves championing investments in research and innovation. Further, it emphasises the need to facilitate improved data accessibility.

(2) Socioeconomic Adaptation to AI-induced Changes: Recognising the transformative potential of AI, the objective underscores the need to contemporise educational and training systems. The strategy also calls for the nurturing of talent, foresight in anticipating labour market shifts, and robust support mechanisms to assure seamless transitions within the labour market and the adaptation of social protection systems in line with AI-induced changes.

(3) Establishment of an Ethical and Legal Framework: The initiative stresses the importance of relying on the EU's core values and aligning with the Charter of Fundamental Rights of the EU. The strategy outlines forthcoming guidance on current product liability laws, the thorough investigation of emerging problems, and collaboration with stakeholders through a European AI Alliance to create AI ethics guidelines.

Various actions have followed this strategy. To realise the third objective and engage a broad range of stakeholders, the European Commission set up the European AI Alliance in June 2018. Since then, the European Commission has been engaged in an open dialogue with citizens, civil society, business and consumer organisations, trade unions, academia, public authorities and experts within the framework of its AI Strategy (Ulnicane, 2022). Starting as an online forum for discussions, the AI Alliance has become a vibrant community that, over the last few years, has engaged around 6,000 stakeholders and contributed to some of the most critical policy initiatives launched in the field of AI (European Commission, 2023c).

Other actions in the strategy followed, like developing a coordinated approach with EU member states, preparing ethics guidelines and international collaboration, and are discussed in the sections below.

3.2 The coordinated plan on AI

When the EU's AI strategy was launched, several member states published or were about to publish their national AI strategies (Ulnicane, 2022). The importance of member states working together to support the EU as a whole in competing globally and to avoid fragmenting the single market is stressed emphasised in the AI strategy. To enable this, the Commission pledged to collaborate with member states on a coordinated plan for AI.

To impede fragmentation in Europe, the Coordinated Plan on Artificial Intelligence has been designed to expedite investment in AI, harmonise AI-related strategies and programmes, and align overarching AI policies. First published in 2018, the Plan serves as a collective pledge among the European Commission, EU member states, Norway and Switzerland to optimise Europe's competitive edge on a global scale in the area of trustworthy AI. The Plan's first iteration outlined activities and financial mechanisms to foster the interest and development of AI across various sectors. Simultaneously, it encouraged member states to formulate and pursue their own national AI strategies (European Commission, 2023d).

The annex to the Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions "Coordinated Plan on Artificial Intelligence" opens with: "The Union aims to develop trusted AI based on ethical and societal values building on its Charter of Fundamental Rights. People should not only trust AI but also benefit from the use of AI for their personal and professional lives. … Overall, the ambition is for Europe to become the world-leading region for developing and deploying cutting-edge, ethical and secure AI, promoting a human-centric approach in the global context" (European Commission, 2018b).

It defines joint actions for closer and more efficient cooperation between member states and the Commission, integrating national- and regional-level actions and measures with the EU-level ones provided for in the strategy (Ulnicane, 2022). In total, the original plan proposes several joint actions in key areas, such as strategic actions and coordination, maximising investments through partnerships, building up research excellence, establishing world-reference testing facilities, accelerating AI take-up through Digital Innovation Hubs, skills and lifelong learning, data, ethics and regulatory frameworks, AI for the public sector and international cooperation (European Commission, 2018b).

The plan is to be monitored and updated regularly. The latest update to the plans was published in 2021 and is closely aligned with the European Commission's digital and green priorities and Europe's response to the COVID-19 pandemic (European Commission, 2023d).

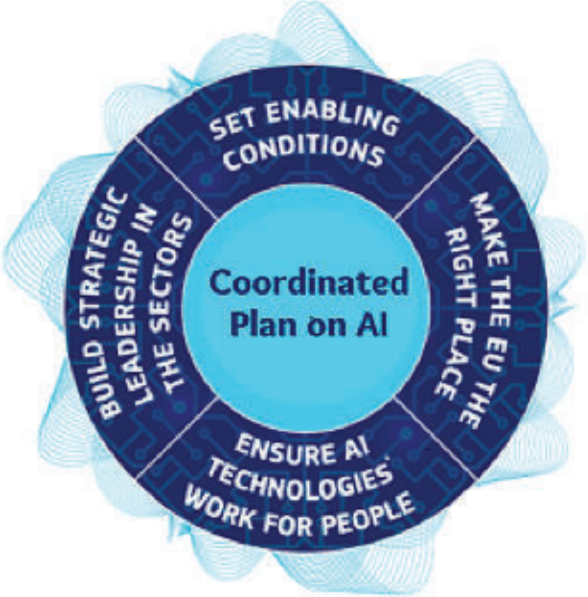
The Coordinated Plan of 2021 aims to turn strategy into action by prompting efforts to (European Commission, 2023d):

- accelerate investments in AI technologies to drive resilient economic and social recovery aided by the uptake of new digital solutions;
- act on AI strategies and programmes by fully and timely implementing them to ensure that the EU fully benefits from first-mover adopter advantages; and
- align AI policy to remove fragmentation and address global challenges.

In order to achieve this, the updated plan sets four key sets of policy objectives, supported by concrete actions and indicating possible funding mechanisms and the timeline to (European Commission, 2023d):

- establish enabling conditions for the development and uptake of AI in the EU;
- make the EU the place where excellence thrives from lab to market;
- ensure that AI technologies work for people; and
- build strategic leadership in high-impact sectors.

Figure 1: **Key sets of policy objectives of the Coordinated Plan on AI**


SET ENABLING CONDITIONS
MAKE THE EU THE RIGHT PLACE
ENSURE AI TECHNOLOGIES WORK FOR PEOPLE
BUILD STRATEGIC LEADERSHIP IN THE SECTORS
Coordinated Plan on AI

Source: European Commission, 2023d.

According to Ulnicane (2022), the European approach to AI synthesises diverse components from previous policy documents of the EU concerning AI, underscoring the intricate interplay of the features of Normative Power Europe and Market Power Europe. Conventionally attributed to Market Power Europe, investment policies and regulatory mechanisms are intimately interconnected with the ethics and values commonly associated with Normative Power Europe. This integration arises from the anticipation that AI investments will target an improvement of societal issues, while regulatory guidelines are intended to safeguard core values and instantiate an ethical infrastructure.

3.3 Ethics Guidelines for Trustworthy AI

Every EU policy document concerning AI points out the imperative to establish an ethical framework for the technology. In 2018, the Commission inaugurated an independent High-Level Expert Group (HLEG) on AI (European Commission, 2018c) with the specific task to formulate ethical guidelines. Comprising 52 specialists drawn from various sectors – industry, academia, civil society – the group was assembled through a transparent and open selection procedure.

One year later, after the first draft of ethics guidelines had been made public and received over 500 comments through an open consultation, the "Ethics Guidelines for Trustworthy AI" were published (European Commission, 2019a).

According to the Guidelines, trustworthy AI should be (European Commission, 2019b):

(1) lawful – respecting all applicable laws and regulations;

(2) ethical – respecting ethical principles and values; and

(3) robust – from a technical perspective and while taking its social environment into account.

A 'human-centric' approach to AI is emphasised in the Guidelines. In this approach, "AI is not an end in itself, but rather a promising means to increase human flourishing, thereby enhancing individual and societal well-being and the common good, as well as bringing progress and innovation" (European Commission, 2019a).

The Guidelines propose seven fundamental criteria to certify the trustworthiness of AI systems, along with a targeted assessment checklist for their verification:

- **Human Agency and Oversight:** AI should augment human capabilities and protect fundamental rights while implementing effective oversight through human-centric models like human-in-the-loop, human-on-the-loop, or human-in-command.
- **Technical Robustness and Safety:** Resilience, security and reliability are paramount. Systems should have contingency plans and minimise unintentional harm through accuracy and reproducibility.
- **Privacy and Data Governance:** AI must respect privacy and data protection while ensuring sound data governance, focusing on data quality, integrity, and authorised access.
- **Transparency:** Clarity in data, systems, and business models is essential. Traceability mechanisms and context-specific explanations should be provided. Users should know they are interacting with AI and understand its capabilities and limits.
- **Diversity, Fairness and Non-Discrimination:** Systems should avoid unfair biases to prevent discrimination and societal harm, while promoting inclusivity and stakeholder engagement throughout their lifecycle.

- **Societal and Environmental Well-Being:** AI should be sustainable and eco-friendly, carefully assessing social and environmental impacts.
- **Accountability:** Mechanisms for responsibility, including auditability for critical applications and accessible redress options, should be established.

The Trustworthy AI framework is rooted in the foundational human rights prescribed in EU Treaties and the Charter for Fundamental Rights. The guidelines strive to optimise the advantages of AI while reducing the associated risks. The document cites various domains where AI can offer significant benefits, such as climate action, sustainable infrastructure, public health, educational quality, and digital transformation. Conversely, it also identifies areas of critical concern, including the use of AI for individual identification and tracking, as well as the deployment of lethal autonomous weapons systems (European Commission, 2019a).

3.4 White Paper on AI

On 19 February 2020, the Commission issued a "White Paper on Artificial Intelligence: A European approach to excellence and trust" (European Commission, 2020a). This policy document discusses strategies for promoting AI adoption while managing the risks.

"Against a background of fierce global competition, a solid European approach is needed, building on the European strategy for AI presented in April 2018. To address the opportunities and challenges of AI, the EU must act as one and define its own way, based on European values, to promote the development and deployment of AI."

The Commission believes AI can greatly benefit Europe's society and economy, propelling the EU to a leadership position in the data economy. Yet, there are concerns about AI's potential threats to fundamental EU rights, like non-discrimination. The development of AI must thus honour EU citizens' values and rights, such as privacy.

The White Paper emphasises two main components:

- "An ecosystem of excellence" – A strategy to gather resources, especially for research, innovation, and for supporting small to medium enterprises; and
- "An ecosystem of trust" – essential components listed for Europe's prospective AI regulatory framework that should align with EU regulations.

The second component is the biggest novelty of this document (Ulnicane, 2022) as it proposes several possibilities for an AI regulatory framework, with the main emphasis on building trust among consumers and businesses (European Commission, 2020a). To promote excellence, it provides several previously introduced investment suggestions in AI. It additionally references the European Green Deal, highlighting AI as a vital tool for achieving its objectives.

The White Paper suggests that the systems should be transparent with human supervision for AI uses in high-risk areas like health and transport. For instance, algorithms in cosmetics or cars should be testable by the authorities. The Commission sought a Europe-wide discussion about biometric data usage for remote identification, like facial recognition, also stressing exceptions to general prohibitions and establishing common safety measures in line with the EU data protection rules and the Charter of Fundamental Rights. Moreover, the Commission was/is evaluating whether the present EU liability laws adequately protect victims of AI mishaps. Although a total overhaul is not seen as necessary, the focus is on maintaining safety standards and ensuring victim compensation.

Alongside the White Paper, the European data strategy (European Commission, 2020b) was unveiled as part of a new digital strategy in response to Europe's digital transformation (European Commission, 2020c). This strategy, "Shaping Europe's digital future", is in harmony with Commission President Ursula von der Leyen's vision and stresses three pillars:

- Technology that works for the people
- A fair and competitive digital economy
- An open, democratic and sustainable society

A public consultation linked with the White Paper's release was initiated, inviting feedback from European citizens and stakeholders (academia, industry, civil society) by 31 May 2020. Following a public consultation on the White Paper, the Commission published its proposal for regulation in 2021, as presented below.

3.5 AI Act

In the European Union, AI will be regulated by the AI Act, which once confirmed will become the world's first comprehensive AI law (European Commission, 2021a). As described above, the White Paper was the first step in meeting the EU's objective to regulate AI with harmonised rules, which is part of its digital strategy (European Commission, 2020c), to ensure the conditions for developing and using AI systems. They can be used in various applications, to create many benefits (e.g., more efficient manufacturing, more sustainable energy, cleaner and safer transport…) and are classified in the AI Act according to the risk they might pose to users.

The proposed AI Act forms part of a broader comprehensive package of measures that address problems created by the development and use of AI (Machinery Directive, General Product Safety Directive, Data Governance Act, Open Data Directive (also other initiatives under the EU strategy for data), Digital Services Act).

The proposal aligns well with the Commission's overarching digital strategy (European Commission, 2020c), specifically to advance technology that benefits individuals. This is one of the three core pillars in the policy direction and goals outlined in the document "Shaping Europe's Digital Future". The proposal provides a well-structured, impactful and balanced approach to ensure that AI is designed to uphold human rights and gain public confidence, thereby preparing Europe for the digital era and setting the stage for the forthcoming Digital Decade (European Commission, 2023e).

The Commission has put forward the specific objectives, as written in the proposed regulatory framework on Artificial Intelligence (European Commission, 2021a):

- to ensure that AI systems placed on the Union market and used are safe and respect existing laws on fundamental rights and Union values;
- to assure legal certainty to facilitate investment and innovation in AI;
- to enhance governance and the effective enforcement of the existing law on fundamental rights and safety requirements applicable to AI systems; and
- to facilitate the development of a single market for lawful, safe and trustworthy AI applications and to prevent market fragmentation.

The Commission distinguishes several risk levels regarding AI practices.

Unacceptable risks

The Commission's proposal outlines four categories of prohibitions related to AI systems. Three categories are entirely banned:

Two of these three focus on manipulation:

i) AI systems using subliminal techniques that go beyond a person's awareness to significantly alter their behaviour, resulting in physical or psychological harm to them or others.

ii) AI systems that target the vulnerabilities of certain groups (e.g., age, physical or mental disability) to significantly change their behaviour, causing or likely causing physical or psychological harm.

The third focuses on social scoring:

The Draft AI Act prohibits AI systems that are:

i) used by or for public authorities;

ii) designed to produce 'trustworthiness' scores; and

iii) result in the unfair or disproportionate treatment of individuals or groups or detrimental treatment that is justifiable but occurs in a context unrelated to the input data.

The fourth category imposes a conditional ban:

AI systems that use "real-time" and "remote" biometric identification in publicly accessible spaces by law enforcement are forbidden. An example of such a system would be a large-scale CCTV network coupled with facial recognition software. This is prohibited unless used for specific law enforcement objectives and accompanied by an independent authorisation regime.

High risk

The Commission's proposal identifies AI systems that pose a risk to safety or fundamental human rights at the level of "high risk" and classifies them in two main categories:

(1) AI systems integrated into products governed by EU product safety legislation, such as toys, aviation, automobiles, medical devices, and elevators.

(2) AI systems in eight specific domains that must be registered in an EU database:

- the biometric identification and classification of individuals;
- the management and functioning of critical infrastructure;
- education and vocational training;
- employment and labour management, including access to self-employment;
- access to essential private and public services and benefits;
- law enforcement;
- migration, asylum, and border control procedures; and
- legal interpretation and application of the law.

The proposal stipulates that these high-risk AI systems must undergo evaluation before entering the market and continuously throughout their lifecycle.

Limited risk

The Commission's proposal also addresses AI systems that represent a "limited risk", emphasising that these systems should meet basic transparency standards. The idea is to empower users to make informed choices about whether to continue using such applications. Specifically, users must be notified when they are interacting with an AI system, especially where the AI generates or alters image, audio, or video content, such as with deepfakes.

On 14 June 2023, members of the European Parliament adopted the Parliament's negotiating position on the AI Act. Talks will now begin with EU countries on the Council regarding the law's final form, and the aim is to reach an agreement by the end of this year (European Commission, 2023f).

4 Challenges in AI governance

In its broader definition, AI governance represents a legal framework aimed at ensuring that AI technologies are developed with the primary objective of helping humanity navigate the adoption of these technologies in ethical and responsible ways. Recently, the adoption of AI technology has experienced rapid growth across almost every industry sector, including education, healthcare, financial services, retail, transportation and public safety. AI governance has therefore attracted much attention from policymakers. Namely, AI systems could raise several concerns and challenges without proper governance, such as biased decision-making, privacy violations, and data misuse. These threaten transparency and compliance with regulations like the General Data Protection Regulation (Barney, 2023). AI governance is faced with several challenges, covering four broader areas (UK Parliament, 2023):

1. Bias and fairness challenges: First, the bias challenge relates to the potential of AI to either introduce or sustain biases that are considered unacceptable by society. Moreover, the misinterpretation challenge is associated with the capacity of AI to create content that intentionally distorts an individual's behaviour, opinions or personality. Finally, the intellectual property and copyright challenge is related to the AI models and tools that can utilise the content created by others, requiring appropriate policies to define and uphold the rights of the creators of that content.

2. Data and privacy challenges: First, the privacy challenge is linked to the capacity of AI to identify individuals and utilise their personal information in ways that exceed public consent. Moreover, the access to data challenge is associated with the requirement of powerful AI systems for extensive datasets, which are typically owned by a limited number of organisations. Finally, the access to computing challenge is associated with the development of powerful AI, which requires significant computing power, access to which is limited to a few organisations.
3. Regulation and transparency challenges: First, the open-source challenge is related to the debate on whether the public availability of code could foster transparency and innovation, whereas allowing it to remain proprietary may result in a concentration of market power but potentially facilitate more robust regulation against possible harms. Further, the liability challenge is associated with the need for suitable policies to determine whether developers or providers of the technology should be held accountable for any harm caused when third parties employ AI models and tools for malicious purposes. Lastly, the black-box challenge is linked to the incapability of certain AI models and tools to explain their specific outcomes, posing a transparency challenge.
4. Global governance and employment challenges: First, the international coordination challenge is related to the global nature of AI technology, requiring the establishment of international governance frameworks to regulate its applications. Moreover, the existential challenge is associated with the perspective held by some individuals that AI poses a significant threat to human existence, and if such a possibility exists, governance structures must be in place to protect national security. Finally, the employment challenge is linked to the potential of AI to disrupt both the nature and availability of jobs, requiring proactive policymaking to manage this transformation.

In response to these challenges, UNESCO introduced the first-ever global standard on AI ethics known as the Recommendation on the Ethics of Artificial Intelligence (Recommendation), which all 193 member states adopted in November 2021. The Recommendation emphasises human rights and dignity, grounded in promoting fundamental principles like transparency and fairness, always highlighting the importance of human oversight of AI systems. At the heart of the Recommendation lie four fundamental core values, serving as the foundations for AI systems that work for the good of individuals, society and the environment. These core values are: 1) respect, protection and promotion of human rights and fundamental freedoms and human dignity; 2) living in peaceful, just and interconnected societies; 3) ensuring diversity and inclusiveness; and 4) a flourishing environment and ecosystem. The core values are further operationalised through ten core principles which provide a human-rights-centred approach to the ethics of AI (UNESCO, 2021):

1. Proportionality and do no harm: AI systems should be employed only to the extent needed to attain legitimate aims, and the utilisation of risk assessment should be employed to avert potential harms stemming from these applications.
2. Safety and security: AI stakeholders should take appropriate measures to prevent and eliminate undesirable consequences, including safety hazards and susceptibility to attacks, arising from AI systems.
3. Right to privacy and data protection: Privacy should be protected and fostered during every stage of the AI development process, while adequate data protection frameworks should also be implemented.
4. Multi-stakeholder and adaptive governance and collaboration: Data utilisation should follow international law and respect national sovereignty, allowing countries to regulate all data while recognising that involving diverse stakeholders is crucial for inclusive AI governance approaches.
5. Responsibility and accountability: AI systems should offer auditability and traceability, with established mechanisms for oversight, impact assessment, auditing, and due diligence to prevent conflicts with human rights standards and risks to environmental well-being.
6. Transparency and explainability: The ethical deployment of AI systems relies on ensuring their transparency and explainability, including the disclosure of AI-driven decisions, with recognition that the appropriate level of transparency and explainability must be contextually balanced, taking potential conflicts with principles such as privacy, safety and security into account.
7. Human oversight and determination: Countries should ensure that AI systems do not replace the ultimate human responsibility and accountability.
8. Sustainability: AI technologies should be evaluated based on their effects on sustainability, encompassing a dynamic set of objectives, including those outlined in the Sustainable Development Goals adopted by the United Nations.
9. Awareness and literacy: Public awareness and literacy regarding AI and data should be fostered through accessible education, civic involvement, the development of digital skills and AI ethics, as well as media and information literacy.

10. Fairness and non-discrimination: AI stakeholders should promote equity, fairness and non-discrimination while adopting an inclusive approach to ensure that the advantages of AI are accessible to everyone.

5 Impacts of AI on society and perceptions of EU citizens

The emergence of data-driven technologies has propelled the advancement of AI, leading to increased automation of tasks traditionally performed by humans. The COVID-19 pandemic accelerated the adoption of AI and data sharing, creating new opportunities but also posing challenges and threats to human and fundamental rights. It is not surprising that recent advancements in AI have garnered widespread attention from the media, civil society, academia, human rights bodies, and policymakers. While much of that attention has been directed to its potential to support economic growth, the impact of AI on fundamental rights has received less attention. Table 1 presents the potential benefits and possible errors that could occur in the context of AI in four core areas: social benefits, predictive policing, health services, and targeted advertising (European Union Agency for Fundamental Rights, 2020).

Table 1: **Potential benefits and possible errors in the context of AI**

Area	Potential benefits	Potential errors	
		AI wrongly declares a result (false positive)	**AI fails to declare a match (false negative)**
Social benefits	- Better access to social welfare	- A person receives benefits they are not entitled to	- A person does not receive their benefits
	- Improved public administration		
Predictive policing	- More crimes detected	- An innocent person is suspected	- Crimes not identified
	- Less crime	- Less trust	- A rise in crime
Medical diagnosis	- Better healthcare	- Wrong treatment	- Disease not diagnosed
		- Less trust	- Healthcare does not improve
Targeted advertising	- Better consumer information	- A person receives irrelevant adverts/ offensive content	- Adverts miss their target
	- Greater profits for companies		- Inefficient adverts

Source: European Union Agency for Fundamental Rights, 2020

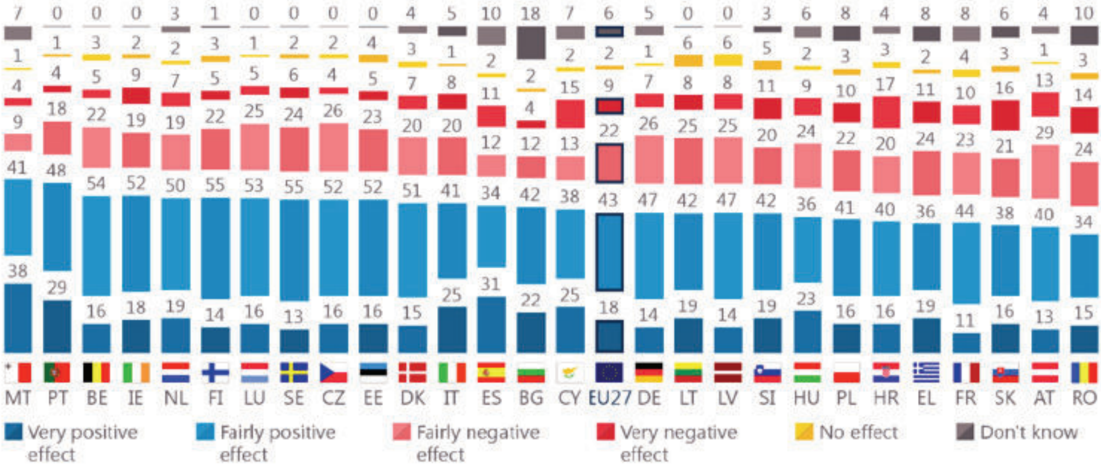
To effectively address the challenges associated with AI, including the potential errors, policymakers should promote collaboration and knowledge-sharing between individuals and organisations dedicated to human rights protection and those focused on AI. This collaboration should encompass both technological expertise and an understanding of fundamental rights. Despite the growing trend of AI adoption in the EU, it remains in its infancy. However, given that the technology is advancing faster than the regulation of it, policymakers should seize the opportunity today to ensure that the future EU regulatory framework for AI and its governance is firmly rooted in respecting human and fundamental rights while fostering trust in AI technology within society (European Union Agency for Fundamental Rights, 2020).

Based on recent data, 61% of EU respondents believe AI will bring about a positive transformation in our way of life over the next two decades. Still, there are some interesting differences in respondents' perceptions that depend on their sociodemographic characteristics (European Commission, 2021b):

- Men tend to hold more favourable opinions regarding the future impact of AI on life in the next two decades, with 66% expressing positivity, compared to 57% of women.
- Younger respondents are more likely than their older counterparts to think AI will positively impact society in the next two decades. Namely, about two-thirds of those aged between 15 and 54 think AI will have a positive impact compared to 54% of those aged 55 and older.
- More educated respondents are likely to think that AI will positively impact life in the next two decades. Specifically, more than two-thirds of respondents who stayed in education the longest think AI will have a positive effect, compared to 35% who completed their education when aged 15 or younger.
- Respondents who live in towns are more likely to be positive about the effect of AI on society in the next two decades. For instance, 64% of respondents living in large towns hold favourable opinions about AI, compared to 55% living in rural villages.

Moreover, the comparison between EU countries reveals some differences (Figure 2). The biggest shares of respondents with a positive belief in AI are observed in Malta (79%), Portugal (77%), Belgium and Ireland (both 70%). In comparison, Romania (49%), Austria (53%) and Slovakia (54%) are identified as the countries with the smallest shares of respondents holding a positive belief in AI.

Figure 2: **European citizens' perception of the future impact of AI on life in the next two decades**



Source: European Commission, 2021b.

It seems that, apart from the sociodemographic characteristics of EU citizens, the geographical perspective is also important when it comes to views on the future impact of AI on life in the next two decades. Therefore, policymakers should promote awareness and understanding of AI technology among marginalised groups of EU citizens, including women, older and less educated citizens, and those residing in rural areas, as well as establish appropriate regulatory frameworks that ensure that AI technologies are developed with the primary objective of helping humanity navigate the adopting of these technologies in ethical and responsible ways.

6 Conclusion

As AI increasingly permeates various aspects of daily life, it is crucial to consider how EU citizens perceive this transformative technology. Variations in perceptions based on sociodemographics and geographical locations signify the important need to foster awareness among marginalised groups. This makes it imperative for policymakers to not only drive understanding of AI but also to formulate regulatory frameworks that are human-centric, ethical and inclusive.

Simultaneously, the EU's approach to AI has been evolving against the backdrop of global efforts to harness the potential held by this technology. While the EU aims to lead in ethically-driven AI, it is also concerned about the international competition, notably from the USA and China. However, an excessive focus on competitiveness could divert attention and resources from other crucial policy areas and even hinder international collaborations. The EU's commitment to human-centric and value-based AI development has earned it a unique position in global forums like the OECD. Nevertheless, this approach is not without its challenges, including establishing a balance of stakeholder interests and the ongoing dialogue on the type of regulation needed for civilian and military AI applications.

As we move forward, public trust in AI will play a pivotal role in its successful integration into society. Ensuring this trust will require an emphasis on transparent governance and an inclusive approach that considers the views and needs of all population segments. This is especially crucial given the EU's broad ambitions, which range from enhancing its digital and technological framework to meeting its commitments, such as the Green Deal. A human-centric and ethical approach to AI is not just a question of policy but a prerequisite for achieving these broader objectives. It will involve harmonisation across EU institutions, financial commitments from various sources, and a genuine attempt to balance the interests of everyone, from large enterprises to the most vulnerable social groups.

In summary, the EU's approach to AI is a balancing act between fostering innovation and ensuring social and ethical responsibility. The ultimate success in AI policy, regulation, governance and development will depend on how well the EU can maintain this balance, keep public trust, and align AI policies with its broader social and economic agendas.

REFERENCES

- Agarwal, P. K. (2018). Public administration challenges in the world of AI and bots. Public Administration Review, 78(6), 917–921.
- Barney, N. (2023). Definition: Artificial intelligence (AI) governance. Available at: https://www.techtarget.com/searchenterpriseai/definition/AI-governance.
- Berryhill, J., Heang, K. K., Clogher, R., & McBride, K. (2019). Hello, World: Artificial intelligence and its use in the public sector. OECD library.
- Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... & Williams, M. D. (2019). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994.
- European Commission. (2018a) Artificial intelligence for Europe. Communication. COM(2018) 237 final. Brussels 25.4.2018.
- European Commission. (2018b). Annex to the coordinated plan on artificial intelligence. Communication COM (2018) 795 final ANNEX. Brussels 7.12.2018.
- European Commission. (2018c). Commission appoints expert group on AI and launches the European AI Alliance. Retrieved 24 August 2023 from https://digital-strategy.ec.europa.eu/en/news/commission-appoints-expert-group-ai-and-launches-european-ai-alliance.
- European Commission. (2019a). Ethics guidelines for trustworthy AI. Independent High-Level Expert Group on Artificial Intelligence set up by the European Commission. Brussels. 8.4.2019.
- European Commission. (2019b). Ethics guidelines for trustworthy AI. Retrieved 25 August 2023 from https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.
- European Commission. (2020a). White paper on artificial intelligence – A European approach to excellence and trust. White Paper. COM(2020) 65 final. Brussels 19.2.2020.
- European Commission. (2020b). A European strategy for data. COM(2020) 66 final. Brussels 19.2.2020.
- European Commission. (2020c). Shaping Europe's digital future. Retrieved 26 August 2023 from https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066.
- European Commission. (2021a). Proposal for a Regulation of the European parliament and of the council. Laying down harmonised rules on artificial intelligence (Artificial intelligence act) and amending certain union legislative acts. COM(2021) 206 final. 2021/0106(COD). Brussels, 21. 04. 2021.

- European Commission. (2021b). European citizens' knowledge and attitudes towards science and technology. Brussels: European Commission.
- European Commission. (2023a). The European Commission's priorities. Retrieved 20 August 2023 from https://commission.europa.eu/strategy-and-policy/priorities-2019-2024_en.
- European Commission. (2023b). 2023 Report on the state of the Digital Decade. Retrieved 20 August 2023 from https://digital-strategy.ec.europa.eu/en/library/2023-report-state-digital-decade.
- European Commission. (2023c). The European AI Alliance. Retrieved 21 August 2023 from https://digital-strategy.ec.europa.eu/en/policies/european-ai-alliance.
- European Commission. (2023d). Coordinated plan on Artificial Intelligence. Retrieved 22 August 2023 from https://digital-strategy.ec.europa.eu/en/policies/plan-ai.
- European Commission. (2023e). Europe's Digital Decade: digital targets for 2030. Retrieved 28 August 2023 from https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/europes-digital-decade-digital-targets-2030_en.
- European Council. (2017). European Council Conclusions. 19.10.2017.
- European Union Agency for Fundamental Rights. (2020). Getting the future right: Artificial intelligence and fundamental rights. Available at: https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-artificial-intelligence_en.pdf.
- Kuziemski, M., & Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. Telecommunications policy, 44(6), 101976.
- Library of Congress. (2023, July 18). China: Generative AI Measures Finalized. Retrieved November 21, 2023, from https://www.loc.gov/item/global-legal-monitor/2023-07-18/china-generative-ai-measures-finalized/.
- Misuraca, G., & Van Noordt, C. (2020). AI Watch-Artificial Intelligence in public services: Overview of the use and impact of AI in public services in the EU. JRC Working Papers, (JRC120399).
- OECD. (2019). Artificial Intelligence in Society. OECD Publishing, Paris.
- Righi, R., Pineda León, C., Cardona, M., Soler Garrido, J., Papazoglou, M., Samoili, S. & Vázquez Prada Baillet, M. (2022). AI Watch Index 2021. López Cobo, M. and De Prato, G. editor(s), EUR 31039 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-53602-4, doi:10.2760/921564, JRC128744.Siebel, T. M. (2019). Digital transformation: survive and thrive in an era of mass extinction. RosettaBooks, New York.

- Stone, P., Brooks, R., Brynjolfsson, E., Calo, R., Etzioni, O., Hager, G., ... & Teller, A. (2016). Artificial intelligence and life in 2030: the one hundred year study on artificial intelligence. Stanford University.
- Straus, J. (2021). Artificial intelligence–challenges and chances for Europe. European Review, 29(1), 142-158.
- The White House. (2023, October 30). FACT SHEET: President Biden Issues Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence. Retrieved November 21, 2023, from https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/.
- Turing, A. M., & Haugeland, J. (1950). Computing machinery and intelligence. The Turing Test: Verbal Behavior as the Hallmark of Intelligence, 29-56.
- UK Parliament. (2023). AI offers significant opportunities but twelve governance challenges must be addressed says Science, Innovation and Technology Committee. Available at: https://committees.parliament.uk/committee/135/science-innovation-and-technology-committee/news/197236/ai-offers-significant-opportunities-but-twelve-governance-challenges-must-be-addressed-says-science-innovation-and-technology-committee/.
- Ulnicane, I. (2022). Artificial Intelligence in the European Union: Policy, ethics and regulation. In The Routledge handbook of European integrations. Taylor & Francis.
- Ulnicane, I., Eke, D. O., Knight, W., Ogoh, G., & Stahl, B. C. (2021). Good governance as a response to discontents? Déjà vu, or lessons for AI from other emerging technologies. Interdisciplinary Science Reviews, 46(1-2), 71-93.
- Ulnicane, I., Knight, W., Leach, T., Stahl, B. C., & Wanjiku, W. G. (2021). Framing governance for a contested emerging technology: insights from AI policy. Policy and Society, 40(2), 158-177.
- UNESCO. (2021). Key facts: UNESCO’s Recommendation on the Ethics of Artificial Intelligence. Paris: UNESCO.
- Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2019). Artificial intelligence and the public sector—applications and challenges. International Journal of Public Administration, 42(7), 596-615.

Chapter 11

Possible applications of AI and related regulations in Romania

Melania-Gabriela Ciot

1 Introduction

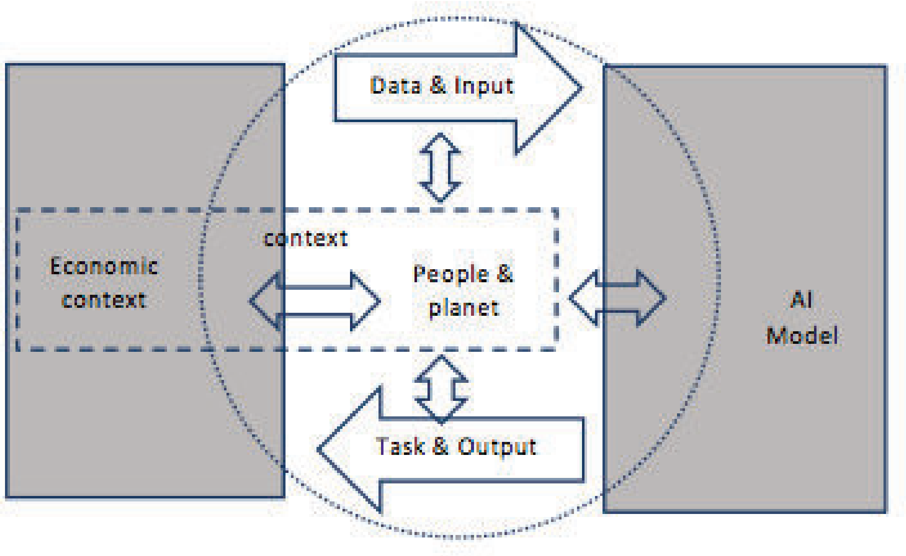
Artificial Intelligence (AI) has attracted considerable attention and use in the last few years due to its value in the decision-making sphere, bringing greater rationality and automation in the process and the scenarios conceived. As presented in reports of the European Cyber Security Agency (ENISA; 2023, 2020), it is influencing the day-to-day lives of people and has an important role in digital transformation. Still, apart from the benefits of its use, there are some concerns regarding manipulation, cyberattacks, privacy, and data protection.

A sustainable EU could be designed and realised in the near future through the digital and green transitions. Better preparing the EU for the socio-economic challenges and threats brought by the pandemic and the war in its Eastern neighbourhood also implies digitalisation. The use of AI will help boost the technological and industrial capacity of the EU, investments in innovation, research, education, support of the labour market, and social protection systems. In line with the Union's values, the ethical and legal framework requires that member states and all stakeholders join forces to support and encourage synergies via cooperation, exchanges of good practices and a clear design of the path forward that EU should take to ensure its competitive global actor role. The integration of AI into the decision-making process is an increasingly widespread trend in a variety of industries and fields, as well as in governance due to its ability to process and analyse massive data and generate real-time results.

2 Framework for addressing AI policies on the European level

The Organisation for Economic Cooperation and Development (OECD) has proposed a framework for classifying AI systems where important elements to be considered are the impact on people and the planet, the economic context, and language resources. The figure below illustrates the interactions and interdependencies existing between different elements that reflect the context in which AI systems/models are operating.

Figure 1: **Framework for classifying AI systems**



Source: OECD, 2022, p. 16.

The framework displayed in the above figure includes factors influencing the AI model such as people, the planet, and the economic context, with inputs and tasks that are generating output. Considering the challenges of the 21st century for the world order and the EU, the framework is a good instrument for the analysis of the usefulness of AI, the difficulties regarding its control and policy developments and regulations in the field, as Clark, Murdick, Perset and Grobelnik (2022) explained. It could be seen as a lifecycle approach to AI able to be used to identify actors in different dimensions of policy and risk management and accountability. An important issue emphasised by the above-mentioned authors is the international dimension of AI, which is claimed to entail the international and European harmonisation of approaches in terms of the collecting and use of data, as well as the use of platforms and better organisation. Policymakers will thereby have support with the design of AI policies regarding the use of AI in the decision-making process, and hence the benefits of its use will be the central interest.

Several principles act as guidelines while using AI systems, including: (1) benefit for people and the planet; (2) human-centred values and fairness; (3) transparency and explainability; (4) robustness, security and safety; (5) accountability; (6) investing in research and development; (7) fostering a digital ecosystem; (8) fostering and enabling the policy environment; (9) building human capacity and preparing for labour transitions; and (10) international, interdisciplinary and multi-stakeholder cooperation (OECD, 2023).

The development of AI has seen policies and regulations taking shape in several states. These approaches must be in line with national and regional policy frameworks (Afina, 2023).

The European Commission's (EC's) initiative from 2018 regarding a Coordinated Plan on Artificial Intelligence (EC, 2018a) was the first important step taken on the EU level for a coordinated approach of the member states (MSs). The declaration emerging from this plan was signed by the MSs and Norway and marked a specific approach to AI, placing the human in the centre of its development. The Communication's objectives were to identify common actions in EU MSs and the EC with a view to assuring an increase in investments, data sharing, the development of talents, building trust, establishing public priority fields such as health, transport and autonomous and interconnected mobility, safety, security, and energy (Universitatea Tehnică din Cluj-Napoca, 2021).

The main legislative pillars for implementing AI on the EU level are: European Strategy for AI (EC, 2018b); Artificial Intelligence for Europa (EC, 2018c); White Paper on Artificial Intelligence – A European approach to excellence and trust (EC, 2020a); European Data Strategy (EC, 2020b); Digital Education Action Plan (2021–2027) (EC, 2020c) and the AI Act (EC, 2021). Each pillar reinforces the foundations for the elaboration of a solid strategic framework for the implementation of AI policies and technologies, also creating instruments flexible enough to cover the future reality and development/evolution of AI.

An important step forward in adopting an AI law on the European level came on June 2023 when the European Parliament (EP) adopted its position on the AI Act, before related discussions involving all the MSs. The rules contained in the AI Act will assure respect for the EU's rights and values, and consider human safety, privacy, transparency, non-discrimination, and social and environmental well-being (European Parliament, 2023).

A useful instrument for monitoring the development and impact of AI in Europe, called the AI Watch, was launched at the end of 2018 by the EC. The AI Watch Index aims to analyse the EU's position on different dimensions of AI relevant to policymakers. Some qualitative indicators are used to make comparisons and establish differences and similarities, especially for the MSs, to facilitate common solutions. The table below indicates the dimensions and indicators used for AI analysis and its impact on policymaking on the level of the EU.

Table 1: **Indicators of the AI Watch Index**

AI Watch Index dimension	AI Watch Index sub-dimension	Indicator name
G – Global view on the AI landscape	AI activity	G1: AI economic players
		G2: AI player intensity
	AI areas of strength	G3: AI areas of specialisation: comparative advantage in AI thematic areas
		G4: AI thematic hotspots
		G5: The EU's comparative advantage in industrial robotics trade
	AI investments	G6: AI investments in the EU
I – Industry	Industry	I1: Profile of AI firms
		I2 Robotics startups in the EU
R – Research and development	R&D activity	R1: AI players in AI R&D
		R2: AI R&D activity score
	Network of collaborations	R3: AI R&D collaborating countries
		R4: Peer-to-peer collaborations
		R5: Strategic position in the network of collaborations
T – Technology	Performance of AI	T1: Performance of AI research
	Standardisation	T2: Standardisation activity engagement

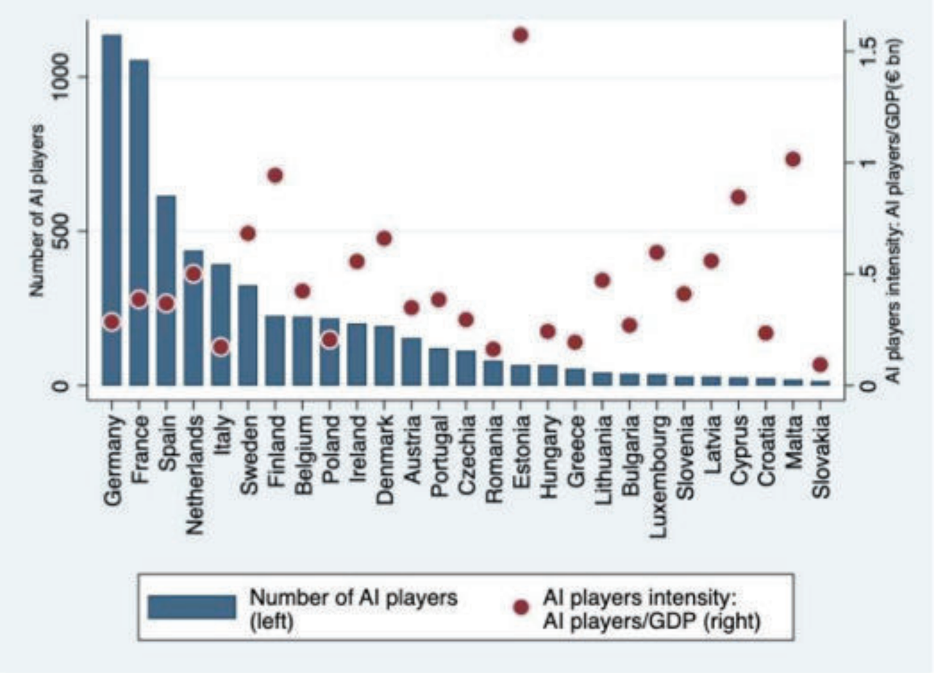
AI Watch Index dimension	AI Watch Index sub-dimension	Indicator name
S – Societal aspects	Diversity in research	S1: Gender diversity index
		S2: Geographic diversity index
		S3: Business diversity index
		S4: Conference diversity index
	Higher education	S5: AI in university programmes in the EU
		S7: AI intensity in university places in the EU

Sources: López Cobo, Montserrat, De Prato, Giuditta (Eds.), 2022, apud. "AI Watch Index. Policy relevant dimensions to assess Europe's performance in artificial intelligence", López Cobo et al., 2021.

As may be seen in the table, five dimensions are used for AI analysis: (a) the global view on the AI landscape; (b) industry; (c) research and development; (d) technology; and (e) societal aspects. The AI Watch Index for 2021 revealed that the USA is the leader on the global scale when it comes to the use of AI, followed by China and the EU. As for the EU, the most significant elements revealed are the use of AI for Services and Robotics and for research and development activities, including software, infrastructures and platforms. The competitive advantage for the EU from the use of AI is the share of economic activities, which is higher than the global average (Lopez Cobo, De Prato, 2022). Further, the EU has an increasing number of patents and research publications and conferences on AI topics, which provide it with a position of influence in the world. The role of projects founded by the European Commission, such as the FP7 and Horizon 2020 frameworks, is important for the research and economic activities associated with AI, creating opportunities to double the economic actors engaging in the technological field. The Index also accentuates the evolution of the AI field, which has consolidated the standardisation of AI activity that is very important on the level of the EU MSs.

As concerns the AI economic actors and their intensity, on the EU level the situation between 2009 and 2020 shows exponential growth for states like Germany, France and less for the Western Balkan and Mediterranean EU states:

Figure 2: AI players and their intensity in the EU between 2009 and 2020

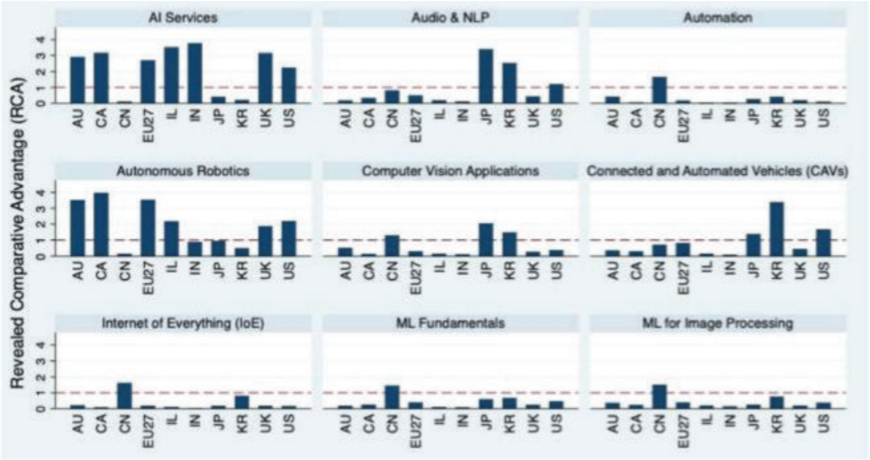


Source: AI Watch Index, 2021.

It is interesting to notice the positions of MSs from Central Europe, found in the middle of the ranking, namely Romania, Poland, Czech Republic etc. For these MSs, the situation has evolved exponentially, especially in 2022 and 2023, due to the specialists and expertise provided for the AI field in terms of the strategic design of its development. The AI topic is present in European and national debates (especially in Romania) because of its effects on the daily lives of European citizens, supporting the automation of different processes, and the advantages for various sectors of activity. In the post-pandemic world, the principles guiding European citizens' lives are changing, with free time and remote jobs leading to more options for employees (EURES, 2020). As mentioned by European Employment Services (2020), four factors have changed the way work is done after COVID-19: the use of technology, the balance between private and professional life, communication, and flexibility. With technology use in first place, the introduction and development of AI has been almost natural.

In the world, the EU has an advantage in AI Services and Autonomous Robotics with effects on sustainability and competitiveness in different sectors such as industry, services, health, manufacturing (Lopz Cobo, De Prato, 2022). When considering different areas of specialisation for AI, such as AI services, audio, and neurolinguistic programming (NLP), automation, autonomous robotics, computer vision applications, connected and automated vehicles, the Internet of Everything, machine learning fundamentals, machine learning for image processing on the global scale, the EU's advantages are obvious. For audio and neurolinguistic programming (NLP), countries from South-East Asia are dominant, such as Japan and South Korea, whereas China is dominating automation, computer vision applications, the Internet of Everything, machine learning fundamentals, and machine learning for image processing. At the moment, the EU must find fast and suitable solutions for the long term if it wishes to be an active player among AI providers and users. The figure below presents the situation on the global level showing areas of AI specialisation:

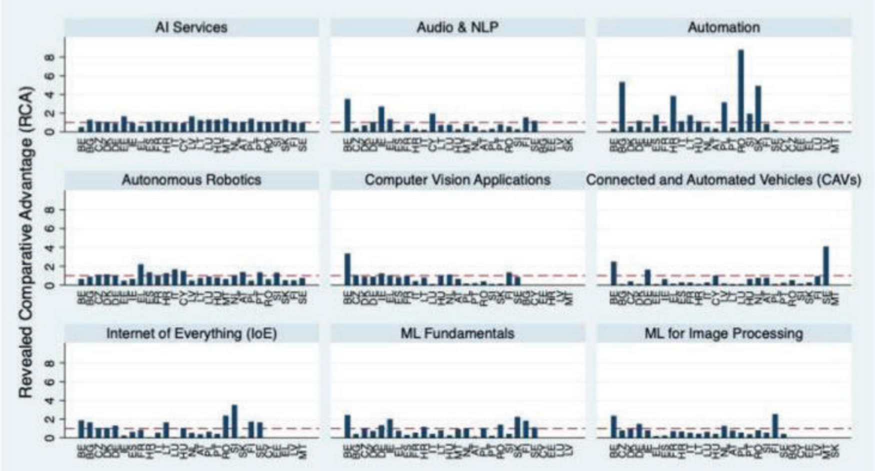
Figure 3: **Areas of specialisation for AI between 2009 and 2020**



Source: AI Watch Index, 2021.

As regards the areas of specialisation, compared with third countries the EU has an interesting distribution, especially in the Central and Eastern MSs (Romania, Bulgaria, Poland, Slovakia), even though these states do not have many AI activities, as the next figure shows:

Figure 4: **Areas of specialisation for AI for EU member states between 2009 and 2020**



Source: AI Watch Index, 2021.

The indicators analysed revealed the active role and dynamic activity of the EU MSs in using AI. In a short analysis of AI services, almost all the MSs are located on the highest level, except for Belgium and Greece. For the audio and NLP components of AI, the scores of Belgium, Ireland and Cyprus are the highest, while those for Slovenia, Austria and Spain are the lowest. The MSs from Central Europe are found at the average level (including Romania). For automation, Romania, Bulgaria and Slovakia occupy the highest position, while six MSs (the Baltic states and Mediterranean states) score the lowest. This AI component reveals the biggest differences in the values for the MSs. The autonomous robotics component of AI technology has very balanced values for all MSs, where the highest place belongs to Greece and the lowest to Estonia. For the computer vision application component, Belgium has the highest position in comparison with the rest of the MSs, and the Central European MSs with the lowest values. With respect to connected and automated vehicles, Sweden, Belgium and Germany have the most significant values, while the remaining MSs have low scores. For the Internet of Everything component, Slovakia, Romania and Belgium have the highest values, whereas the Baltic MSs and Cyprus and Malta do not have a score. When it comes to machine learning fundamentals and image processing, Belgium has the highest scores, while some MSs do not have one, like the Baltic states, Cyprus, Bulgaria or Slovakia. It could be synthesised that on the EU level the AI components have a different distribution and values, meaning that the harmonisation of the values, as well as a suitable AI strategy and policies, are needed in the long term to become more competitive on the global scale.

Overall, there are some interesting characteristics on the EU level and a deeper analysis of what Romania brings by way of assets for the AI component is worth pursuing. In the section below, several indicators for Romania are presented and described.

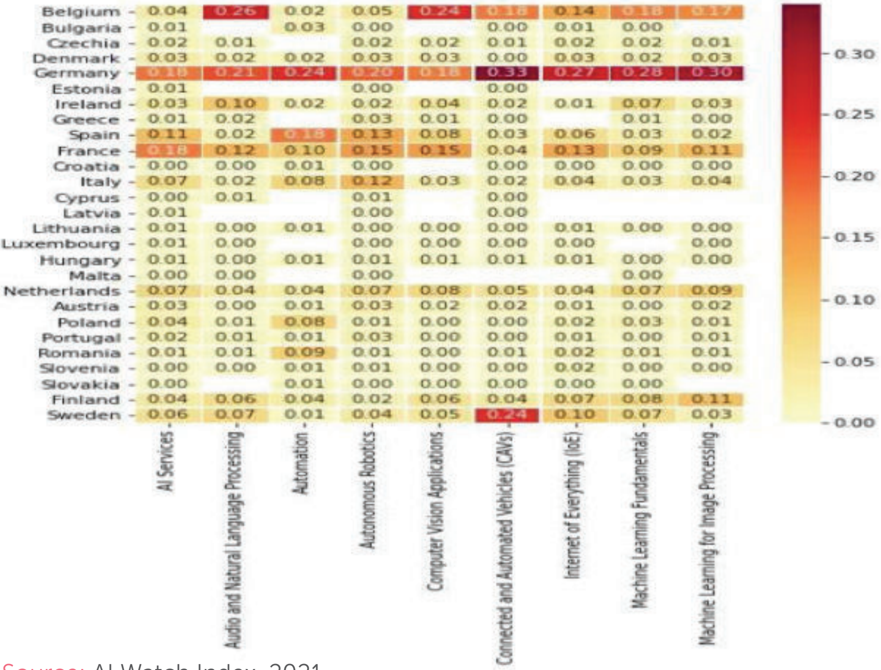
3 Key country data for AI trends in Romania and a comparison with the EU

In Romania, even though artificial intelligence systems and automation technologies are being increasingly applied, they are still below the level of other developed countries in Europe and North America. However, there are several projections of the application of AI segmentation in Romania that could soon lead to a significant rise in the adoption of the technology. In Romania, the use of AI is growing, and this trend is expected to continue in the future. According to a report by the European Cyber Security Agency (ENISA), Romania is among the top-10 countries in Europe in terms of the use of AI in the private sector, especially in the financial, medical and retail fields.

There are several areas of AI application where Romania is well positioned, even though a strategy for AI is missing. As indicated in Figures 2 and 4, Romania, given its economic capacity, scores well for all indicators analysed by the AI Watch Index.

When considering the distribution of AI activity across thematic areas, the situation for Romania is as follows:

Figure 5: **Distribution of AI activities per thematic areas on the EU level between 2009 and 2020**

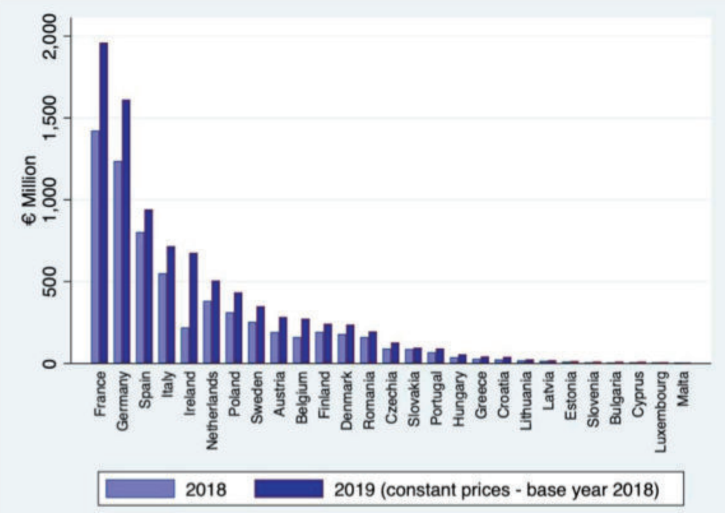


Source: AI Watch Index, 2021.

The AI activities considered for the above figure were: AI services; audio and natural language processing, automation; autonomous robotics, computer vision application; connected and automated vehicles (CAVs); Internet of Everything (IoE); machine learning fundamentals; and machine learning for image processing. As may be seen, Romania scores low, below the EU average, while for one indicator it has no score (computer vision application) and for six indicators a score of 0.01 (AI services; audio and natural language processing; autonomous robotics; connected and automated vehicles (CAVs); machine learning fundamentals; machine learning for image processing). The best indicator for Romania is 0.09 for automation, being among the leading EU MSs, in third position after Belgium and Spain.

Regarding AI Investments in comparison with the average of the EU, the AI Watch Index presents an increasing situation for Romania, similar for the EU (growing increasing from EUR 7.9 billion in 2018 to EUR 9 billion in 2019), which is around the average – 13th place.

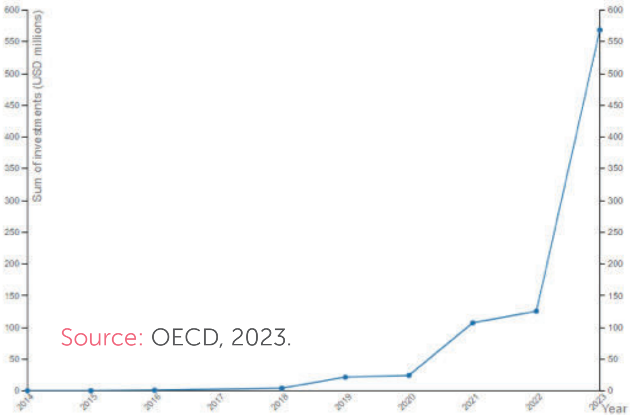
Figure 6: **AI investment in EU 2018–2019**



Source: AI Watch Index, 2021.

In relation to the latest situation concerning AI investment in Romania, the AI Watch Index for 2023 reveals an exponential rise in investment of almost USD 600 billion, largely in the last few years (between 2022 and 2023 from USD 130 billion to almost USD 600 billion), as the figure below shows:

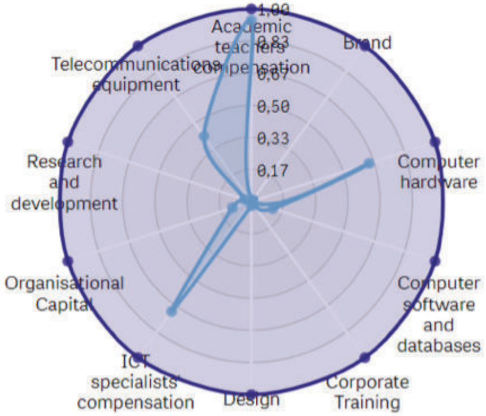
Figure 7: **Investment in AI in Romania between 2014 and 2023**



Source: OECD, 2023.

An interesting graph presents the situation of AI investments in Romania, clearly revealing the huge increase from 2022 to 2023, from almost USD 150 million to almost USD 600 million, as shown by the figure below (referring to AI investments per capita):

Figure 8: **AI Investments in Romania in 2022 and 2023**



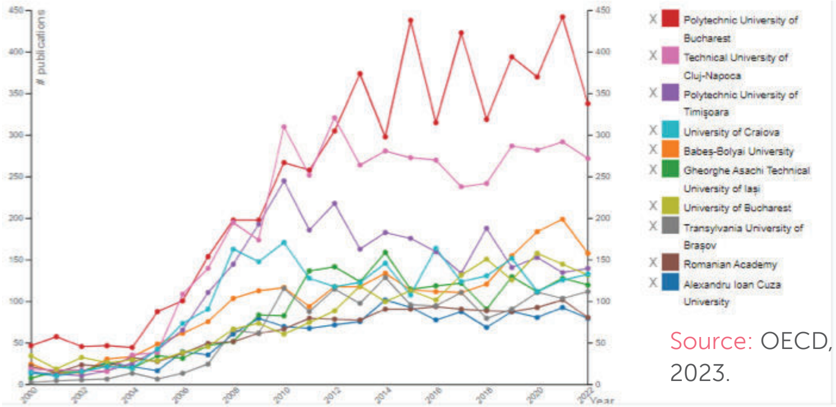
Source: AI Watch Index, 2023.

The dominance of academic research and teaching, telecommunication equipment, ICT specialists' compensation and computer hardware becomes visible. This distribution signals interest in finding new and innovative ways of using AI to benefit society, with an effect on the design and implementation of public policies. Certain components have low scores, such as brand, corporate training, and design. A future strategy for AI development should take this configuration into account.

Research activity in the field of AI is highly valued in Romania. In May 2023, the Scientific and Ethical Council in Artificial Intelligence was established with a consulting role, under the coordination of the Ministry of Research, Innovation and Digitisation (MRID, 2023). The Council is composed of senior researchers and the founders of private companies from the Romanian diaspora who are building the AI environment in Romania and on the global level.

As regards research activity in Romania for AI, it is developed by several universities and the Romanian Academy. The left side of the figure below presents the increasing intense activity over the past 20 years. On the right side of the figure, which displays the impact of AI projects in the public sector, one may notice a decreasing trend, although several good initiatives and collaborations between the authorities and universities are visible, as will be described later in this chapter.

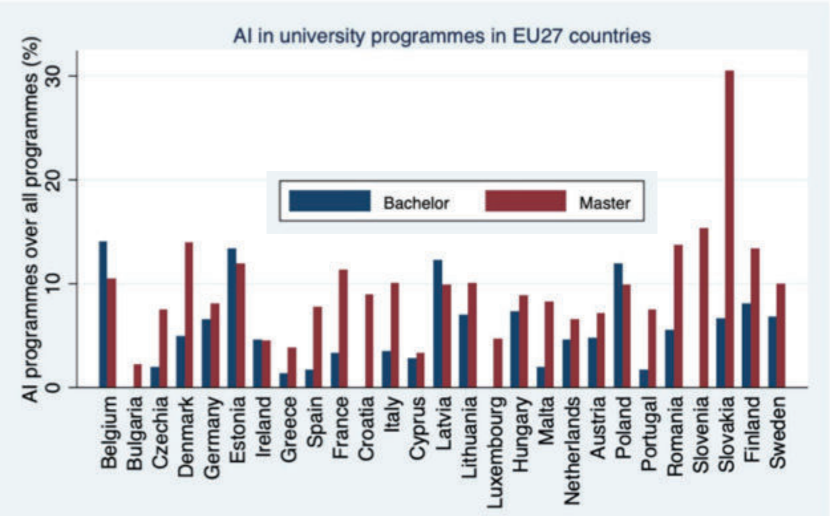
Figure 9: **AI research and software development in Romania**



Source: OECD, 2023.

The figure shows that the leading places when it comes to developing AI research are occupied by the technical universities, namely: Polytechnic University of Bucharest, Technical University of Cluj-Napoca, Babeș-Bolyai University, followed by another four universities and the Romanian Academy. As for the level of studies of AI in universities, compared with other EU MSs the situation in Romania looks like this:

Figure 10: **AI in university programmes**

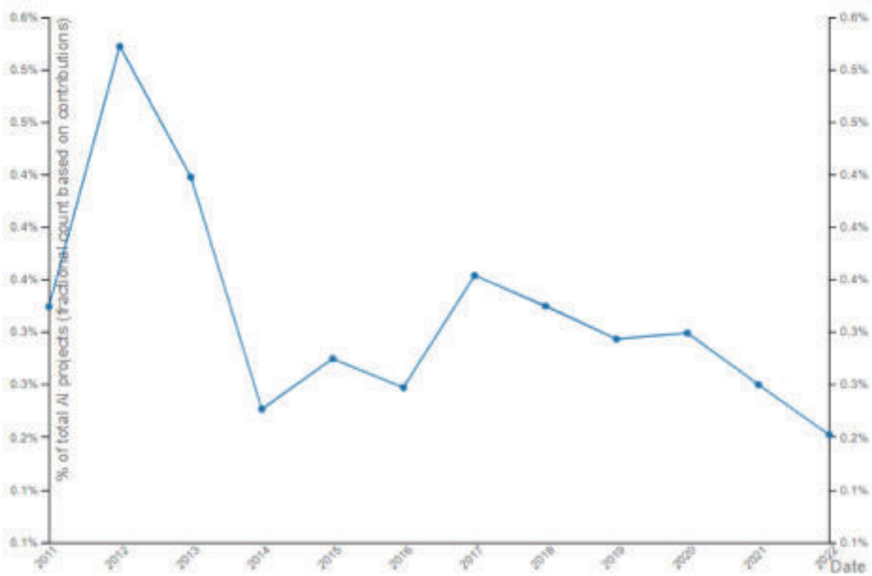


Source: AI Watch Index, 2021.

In Romania, the master's level for AI programmes dominates, being located at the European average. The same goes for the BA level. Between these two levels of education, there are twice as many MA programmes than BA programmes in the AI field. The situation is similar with other MSs, such as Slovakia, Italy or Denmark.

Software development in Romania is supported by public and private environments. The public contribution to AI projects and their impact has been decreasing in the last period, especially from 2017 to 2022, as the below figure shows:

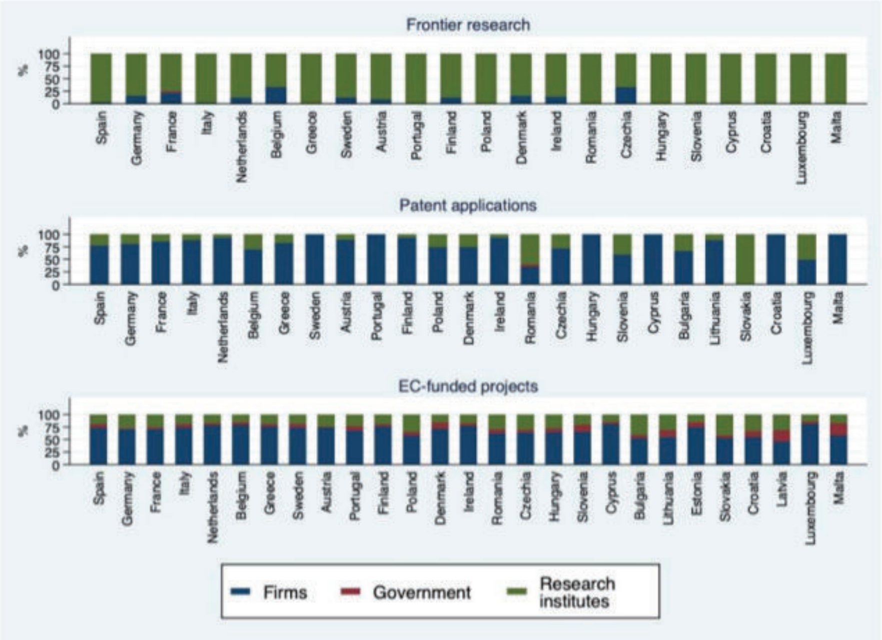
Figure 11: **AI software development in Romania**



Source: OECD, 2023.

In terms of economic players in research and development activities, Romania has average European scores for the elements considered in the analysis, such as:

Figure 12: **Economic players in R&D activities in the EU between 2009 and 2020**

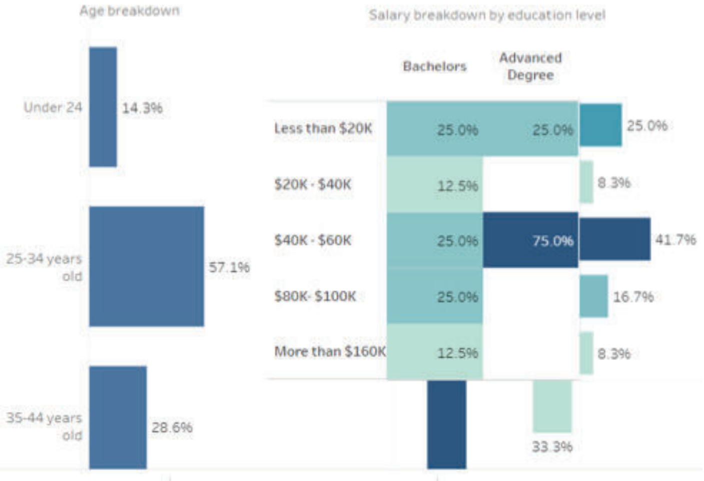


Source: AI Watch Index, 2021.

In terms of frontier research, the score for Romania is low, with the activities being carried out by research institutes. Other MSs, such as Poland or Portugal, have the same scores. For the patent applications, firms and the government are implementing the activities, and the score for Romania is below the European average. The EC-funded projects element for Romania indicates that firms, research institutes and the government are implementing the activities, with the situation being similar to the average of the EUs MSs, such as the Czech Republic, Poland or Slovakia.

Another interesting graph presents the demographics of AI professionals by age in Romania, a very important indicator for the future of AI development:

Figure 13: **Demographics of AI professionals by age in Romania**

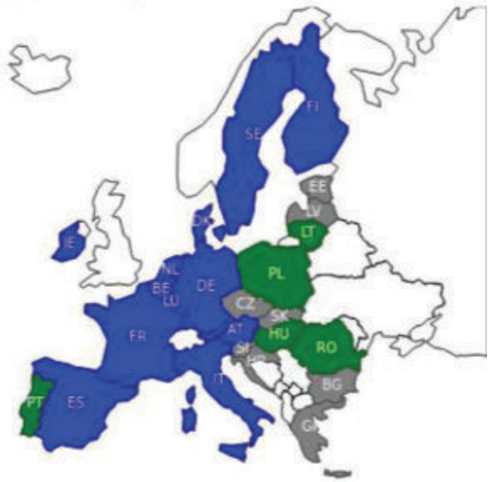


Source: OECD, 2023.

The dominance of the advanced degree for working in the AI field, held by those aged between 25–34 years old, is a good sign reflecting the working capacity and the improvement and innovation capacity.

An interesting indicator to be considered is the participation of the MSs in the standardisation process of AI. Some MSs are active participants, while others are observers or do not have any involvement. Romania is an observer, like other MSs from Central Europe. The figure below presents this situation:

Figure 14: **Participation of the EU MSs in AI standardisation activities in 2021**



Source: AI Watch Index, 2021.

An overview of the key elements regarding the development of AI in Romania indicates that interest and investments in AI activities are increasing similarly to EU trends. The research and development activities are well represented in Romania, with the technical universities being the frontrunners. Moreover, research institutes and private firms are the leading economic players in research activities. The young population is dominant in AI activities, holding an advanced level of education, which holds considerable potential for developing benefits of AI to meet the interests of society.

4 Current state of AI policy in Romania and a comparison with the EU

The national public authority responsible for digital policies in Romania is Autoritatea pentru Digitalizarea României (ADR)/Authority for Romanian Digitisation. On 1 February 2023, the authority launched a national public debate on the need and options for AI regulation (ADR, 2023).

Romania launched on 26th of September this year a draft of a specific strategy for the adoption of digital technologies in the economy and society while respecting human rights and promoting excellence and trust in AI. It is in public consultation (economedia.ro, 2023). Several analyses of the strategic framework for AI have been elaborated, some of them through EU-funded projects, others as a result of collaboration between national public authorities (Authority for Digitisation of Romania) and universities, mainly aiming at the effectiveness the public administration's activities (OPAC, 2023). A specific line of funding an EU project was designed from the Operational Programme Administrative Capacity for the National Strategic Framework in the AI Field – Activity A6.1, from the last Multiannual Financial Framework, 2014–2020. The results of these projects were released in early 2023.

On the national level of Romania, the Strategic Framework for AI takes into consideration: the Government Programme 2021–2024, which mentions AI for the strategic and digital transformation of public administration and the economy, and as well as for national policies, which includes measures for digitalisation and intelligent specialisations, such as: the Romanian National Plan for Recovery and Resilience (PNRR, 2021), Romanian Industrial Policy (OPCA 2018), E-government Romanian Public Policy 2021–2030 (OPCA 2020), and Romanian Educational Policy – Educated Romania (Presedintia României, 2021). Other strategies, as mentioned in the OPCA report (2023), complete the general framework for Romania's approaches to AI, such as: the National Strategy of Research, Innovation, and Smart Specialisation 2022–2027 (Romanian Government, 2022); Employment Strategy (Romanian Government, 2021a) and Romanian Cyber Security Strategy for 2022–2027 (Romanian Government, HG 1321/2021, 2021b).

The national strategic framework for AI in Romania is organised around 6 general objectives to which 13 specific objectives are assigned, which support the design of the proposed measures. These objectives are detailed in Table 2 below:

Table 2: **Objectives of Romania's AI strategy**

General objectives	Specific objectives
OG1. Supporting education for RDI and the training of specific AI skills	OS1.1. To increase the training capacity and training level of an AI specialist
	OS1.2. To increase the level of basic understanding of the population regarding the benefits, use and regulation of AI technologies
OG2. The development and use of efficient infrastructure and data sets	OS2.1. To develop AI-specific hardware infrastructure and ensure transparent and fair access to it, to facilitate the processes of RDI and production in this field
	OS2.2. To expand the use of data sets, with application in various sectors of activity
OG3. The development of the National Research – Development – Innovation System in the field of AI	OS3.1. To develop fundamental and applied scientific research specific to the field of AI, as well as on an interdisciplinary level
	OS3.2. To reduce the fragmentation of R&D resources and efforts in IA by conjugating and synchronising them within some centres and national specialised innovation groups connected to the centres and international AI resources.
	OS3.3. To support and promote AI innovation.
OG4. Transfer insurance technologically through partnerships	OS4.1. To improve the exploitation of research results by developing technological transfer capacities
	OS4.2. To establish and organise a national network of spaces for testing and experimentation (TEF) with solutions developed in the field of AI
OG5. Facilitating the adoption of AI across the whole of society	OS5.1. The adoption of AI technology in the public sector
	OS5.2. The adoption and exploitation of AI technologies in economic priority sectors
OG6. Developing a system for the governance and regulation of AI	OS6.1. To ensure the governance framework for the development of AI
	OS6.2. To facilitate the development of AI through regulation

Source: Universitatea Tehnică din Cluj-Napoca, 2021.

Each of the general and specific objectives in Table 2 are detailed through measures containing responsibilities and a deadline for their realisation. The general architecture of the strategic framework for the elaboration and implementation of the AI strategy in Romania will provide an integrated approach, systematised and divided by action fields and interested parties/ responsible categories, so as to build the foundations for a coherent action plan for the strategy implementation period that will follow. In this way, the national objectives for the adoption of AI by the whole of society will be accomplished and a harmonised transformation will be assured.

It could be stated that the approach taken by the AI policy in Romania is new, yet it has been on the agenda of public authorities and private actors. Romania must fill the gap at the European level by elaborating the national strategy for AI and appointing a national authority in this regard. Through its expertise, the newly appointed Scientific and Ethical Council and the responsible Ministry will put AI regulation and the creation of a suitable legislative framework among their first tasks.

A study case for the use of AI in private sector, in Romania, is the private healthcare system. The Medical Imaging Center Regina Maria is pioneering the use of AI for improving the identification of patological processes. In this way, the volume of work for medical personeel is reduced and the accuracy of results and the detectation of serious condition, such as cancer, is improved (Regina Maria, 2023). It is expected that the use of AI in medical imaging will revolutionize the entire sector, by helping doctors to diagnose more accurate, but also to predict better the evolution of different pathologies and to recommend more efficient and more personalized treatments. The newest project of Regina Maria medical center is called „Lunit INSIGHT MMG", which is a computer-aided detection/diagnosis (CADex) system based on AI algorithm designed to help detect, locate, identify and characterize suspicious areas of breast cancer on mammograms (Regina Maria, 2023). The project is using the DeepcOS AIM, the newest medical platform implemented by the network Regina Maria, which allows to the imaging department the access to AI, in this case to the mammography AI analysis. As the medical doctor responsible for this task, ms. Dr Aurelia Cristina Bilbie reports in the interview, the access is established through a software component (deepcOS Gate), installed on the website of imaging department and a second component, the AI platform in deepc Cloud. The mammographic images (DICOM data) are automatically sent to deepcOS Gate through PACS (Picture Archiving and Communication System), a complex archiving system and access to medical imaging that the Center has within the network. They have also DICOM (Digital Imaging Communication in Medicine), international data exchange standard for biomedical imaging. Then, the data are pseudonymized and automatically sent to AI platform. The data is processed by the system and the mammographic images that will be analyzed by the AI system, will help the radiologist in drafting the final result (Regina Maria, 2023).

It needs to be underlined that the software device is an auxiliary support for the detection and a valuable diagnosis aid, not an interpretative one. That means that the system cannot make an autonoumous diagnosis, but it comes with a pertinent, fast analysis of the huge amounts of data obtained from imaging scans, that will facilitate the final resolution of radiologist and will increase the degree of precision of imaging interpretations and radiological diagnosis (Regina Maria, 2023).

Regarding the use of AI in finance sector in Romania, the experts from the field indicated the following benefits: (a) optimising the operational efficiency, (b) a faster decision-making process and (c) improved financial modelling (Făniță, 2023). Regarding the first benefit, the author mentioned as example an AI-based solution which could model a company's credit risk by predicting its performance under different market conditions. In this way, the company receives a support by avoiding costly mistakes and make better decisions. For Romania, statistics showed that 30% of financial companies have already adopted AI based technology, reporting improved operational efficiency (Făniță, 2023). For the second benefit, the AI-technology based system could support banks making better decisions by predicting which loans to grant, faster that humans. These systems allow them to renegotiate or liquidate loans without affecting the overall bank sheet. In Romania, over 35% of banks use AI to speed up the decision-making process, according with reports (Făniță, 2023). For the third benefit, AI can help investors make informed decisions about a company's performance and long-term prospects. Also, it improved the accuracy of analysis and provides options for investors, by making more accessible to all types of investors. The statistics show that approximately 25% of Romanian investors use AI solutions to improve the financial modelling process (Făniță, 2023). When it comes to risks of using AI in financial sector, the author indicates the following elements: potential data breaches, algorithms errors and lack of treasability (Făniță, 2023). The concerns regarding the use of AI is revealed by an IBM study released on September 2023, which emphasized that 75% of Romanian companies believe that the organizations with the most advanced generative AI will have a competitive advantage, but more than half of the investigated Romanian companies (57%) are concern with data safety and 48% are afraid of the bias or accuracy of data (financialintelligence.ro, 2023).

Recently, the Romanian Minister of Digitalization declared that the AI instruments will help increasing the VAT collection up to 1%. He mentioned that the ministry has identified as short term objectives that the use of ERP, robots automatically processing data and correlate for obtain relevant financial data, could increase the taxes collection for the Government. The estimation are between 0,9% and 1% increase of VAT collection by the AI instruments (economedia.ro, 2023). Another use of AI instrument in central administration is the national electronic invoicing system, e-invoice, which is under working phase between the representatives of Ministry of Digitalization and of Finance, according to the Minister (economedia.ro, 2023).

5 Challenges and opportunities for AI policy in Romania and a comparison with the EU

The biggest challenges for AI policy in Romania are similar for other MSs; namely, databases, financing, the qualifications of human resources, and certifications (Universitatea Tehnică din Cluj-Napoca, 2021). Each of these elements is discussed below.

Databases represent a challenge due to the need for data collections that guarantee integrity and access to data in compliance with the legislation on the protection of personal data. The access to public data must be open, but in accordance with legislation, assuring trust and control with regard to the information provided (Universitatea Tehnică din Cluj-Napoca, 2021).

Financing for AI development and its policies should come from internal and external sources. Research and innovation must be encouraged through financial incentives to facilitate the development of AI solutions in Romania, such as startups or early adopters. In this way, some of the risk for the new companies/products will be carried by the state through this financial incentive, encouraging an openness towards the development of AI. Financing professional trainings and certifications could also be a good opportunity to stimulate AI. Practically, it is about investments in different areas, such as research institutes, education (STEM education), working spaces such as incubators for accelerating businesses and grants for companies which are already investing in AI (Universitatea Tehnică din Cluj-Napoca, 2021).

Noting its big potential in this area, the greatest challenges and opportunities for AI in Romania refer to skills/competencies training. The human resources who benefit from these trainings should come from outside the AI field, such as employees or consultants on the technical and managerial levels. As opportunities arising from this challenge, one could mention access to talents, personnel orientation toward professional training from an ethical perspective, and the development of specific competencies (Universitatea Tehnică din Cluj-Napoca, 2021). The Romanian private sector is supporting the public authorities, especially for the education field. For example, Microsoft was involved in the AI education of students from Academy of Economic Sciences from Bucharest. Two years ago, in 2020, the tech giant supported the lauch of a laboratory equipped with artificial intelligence which offers to students, professors and researchers access to unlimited computing resources and computing power (Neagu, 2021).

For the education field, the use of AI will change, even revolutionate the system (Jurma, 2023). The new generation of AI based on LLM (Large Language Models) allows the automation of didactic functions. These system are open accessible, and it could be considered the beginning of AI revolution (Jurma, 2023). The launching of ChatGPT version 3.5 in November 2022, a generative intelligence based on GPT's LLM marks the beginning of AI-assisted education, which has to be approached differently than computer or Internet-assisted education, still present in Romanian education. It could be considered that up till now, Romanian educational system/school is one of the most conservative systems, struggling to maintain its course in line with the digitalization of society. In Romania, situation is more complex, due to the inflexibility of the system and its underfunding. The use of AI in Romania educational system is likely to dimminish the mentoring activities, which allowed to young people to compensate the difficulties from school and will amplify the shortcomings of Romanian education system (Jurma, 2023). The changes will be happening, so it will be important how the transformation will be approached. The investigations of scenarios made by the author generated the conclusion that education and reforms in social structures will help to bear the impact of changes.

Practically, we are witnessing the debut of a new market, which gives a chance to start-ups, which might be important players in the next 3-5 years. Romania has the advantage of a good IT infrastructure and has aready innovative project in AI field. This advantage could be use to position better in the race for AI dominance, if it is use sooner (Jurma, 2023). The author also mentioned that a big disadvantage for Romania in AI field is education and governance, so AI and education will need transformations. For example, an educational AI has to be trained and prepared to function in Romanian educational environment, reflecting Romanian goods and values (Jurma, 2023).

Another important challenge for education will be the teacher's role. AI won't replace the teachers, contrary, it will need them more. But its role will be dramatically changed and will be more a mediator between pupils and machine. Over time, educational AI will learn from teachers and pupils and will understand and address their needs in a personalized way. Also, the Romanian private lessons market will be strongly affected, because pupils will have free access in any fields to tutors more competent and patient than their teachers. Then exams will become unnecessary, because evaluation will be continuos and in real time. Education itself won't be confined to a period spent in an institution, but will become continuos, throughout life and diffused throughout society (Jurma, 20023).

New jobs will emerge from the AI's use, so educational system will have to prepare the necessary skills of tomorrow's employees, for example prompt engineer (specialized in dialogue with AI, which carefuly has to be trained before being released into the real world). Communication competences will be very important for these new jobs, equal to expertise in the field in which AI is being trained. The author ends his article with a warning, that AI requires training and supervision, and will be more useful to treat it "like a person than a computer" (Jurma, 2023).

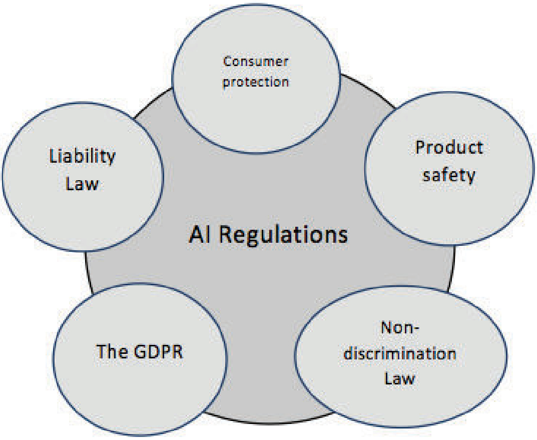
Regulation in the field of AI in Romania represents another challenge. The adaptation of a legislative framework which will support the adoption and development of an AI ecosystem, as well as public policies that support innovation, also taking into consideration ethical codes, and conduct good-practice dissemination, could lead to concrete solutions to be implemented in Romania (Universitatea Tehnică din Cluj-Napoca, 2021).

Other opportunities for Romania's development of AI policies are: (a) the development of the research-development and innovation sector – human resources, expertise, international and national recognition; (b) the consolidation of training and education capacities for AI specialists in the education system; (c) the generalisation of AI basic knowledge and skills among the population and enterprises; (d) developing specific AI structures (investments, regulations, data sets); (e) developing the institutional ecosystem with AI expertise (research centres, firms, testing spaces and experimenting with solutions); (f) adopting AI solutions in the public environment for services and in the private environment for economic competitiveness; and (g) the consolidation of AI governance and regulation (Universitatea Tehnică din Cluj-Napoca, 2023). It could be summarised that the main elements of the AI policy are human resources, knowledge, infrastructures, and institutions.

As relates to AI policy development in Romania (opportunities), the following areas could be considered: (1) the research and innovation area – which has indicated the need for an interdisciplinary approach and the establishment of a different centre for research, development and innovation of this technology type; (2) partnership areas – between public and private entities, facilitating integration into the European network of research; (3) architecture and infrastructures – easy access to devices and infrastructure supported by the cloud networks offered by international companies, which means national solutions developed in this sense could provide an answer to this challenge (Universitatea Tehnică din Cluj-Napoca, 2021).

When it comes to the regulation of AI technologies, it is crucial to connect it with the existing regulations such as consumer protection (OUG 58/2022), the GDPR (2016), liability law (which has not been transposed into Romanian legislation), product safety, non-discrimination law (OUG no. 45/2020). The figure below illustrates the relationships between the regulations in the AI field and other fields:

Figure 15: **Interplay of AI regulations and other related regulations**



Source: Universitatea Tehnică din Cluj-Napoca, 2021.

6 Policy recommendations for Romania

The strategic framework for Romania regarding AI policies is mature and ready for elaboration and implementation. Specialists in the fields from academic communities, especially technical universities and public authorities, mentioned in the reports elaborated several recommendations concerning the adopting of AI on a larger scale in Romania, including:

- To monitor the timing of the progress with AI in Romania compared with international and European developments, taking national specificities into consideration.
- To ensure the cooperation of all stakeholders (academic communities, public administration, business environment) on the implementation and monitoring of the measures and regulations.
- To implement a national AI strategy with a specific degree of flexibility to allow operational adaptations, according to the dynamics of the field, unpredictable developments, technological leaps, developments in the regulatory framework, the actual pace of of understanding and adopting technologies in society, effective progress in creating/launching/implementing projects and programmes regarding AI.
- The control-evaluation function while implementing the AI strategy and policy should be actively supported and consolidated as the key element in the monitoring of progress and rapidly informing national decision-makers, thereby assuring successful implementation.
- Strong national leadership is needed to harmonise the objectives and approaches concerning implementation and maintenance of the course of measures proposed. Romanian expertise could be important for the EU and its objectives regarding positioning the country at the global forefront of AI innovation. In Romania, there is a significant number of specialists, representing an asset in Romania's preparations for the adoption of AI by the whole of society. The national strategic framework is an expression of the public's awareness of the power of technology and its impact on the daily functioning of society and the proactive approach taken by central public authorities in Romania to manage the field in the direction of achieving the objectives and aligning them with international trends. Romania has a chance to build new pillars for technology adoption by continuing the initiatives and performances of the entrepreneurial environment and of research-development-innovation in the AI field, which have advanced in the last few years despite the lack of a regulation framework and a supportive architecture for the innovative in AI. By elaborating the strategic national

framework for AI and its implementation, Romania will take an important and qualitative step forward, positioning the country on the international map of AI and modernising its society and economy. In this way, Romania may considered to be an expertise and innovation regional centre in technology, consolidating the national potential of its talents (Universitatea Tehnică din Cluj-Napoca, 2023).

- To establish a national authority to regulate AI, as suggested in the title of the report prepared by the Technical University of Cluj-Napoca Authority for the Regulation of Artificial Intelligence (ARIA) (Universitatea Tehnică din Cluj-Napoca, 2021), aiming to accelerate innovation in AI in Romania, assuring public trust in systems based on AI, following the line AI made in Europe, AI software application developed in Romania must become export products and ensure protection of the EU's values and harmonisation with EU regulation.

The establishment of ARIA, as the national authority for AI regulation, is a useful idea. Some directions for activities could be (Universitatea Tehnică din Cluj-Napoca, 2021):

- Establishing the procedures for the evaluation, designation, and notification of conforming entities.
- Certification as the centre of the framework for AI regulations.
- Conformity evaluation entities that will certify the systems based on AI, including those with a high risk level.
- Regulating the functioning of testing places for AI regulation. These places are zones where regulations are limited and favourable for testing AI-based applications. Such places are particularly important for introducing new technologies and innovative products to the market.
- Developing more transparent, predictable and verifier guidelines for AI applications.
- Providing grants for startups/experts for the standardisation process.
- Organising training programmes and accrediting authorised auditors for the certification of AI-based systems.
- Providing lists containing experts in different fields of AI.
- Monitoring and regulating new rights of citizens in the framework of interactions with AI-based systems, as an explanation right or the right to know.
- Promoting and running the Romanian Data Space portal in line with the European Data Space.
- Organising public debates and consultations before adopting regulations, offering all interested stakeholders the possibility of formulating opinions

and submitting observations of the proposed measures.

Another interesting proposal for the activity for ARIA could be the creation of a national data space with different approaches to data such as, for example, public data, which should be available and useful for the economy and population, such as statistics, data on the environment, mobility etc., while defining the legal framework for data partition between parties should be compulsory. The policies to be designed and the agreements for the use of data must specify who has a right to access the data and what its aim is, to ensure the maximum benefits of using the data for society. This legal framework must assure the conditions for data providers and users regarding the intersectoral use of data. Regulations for intermediaries are also necessary. The legal framework must establish trust between the users and providers of data.

Sandboxes for the development and implementation of AI policy are important and may amount to examples of good practice for other MSs. Below are a few examples:

Table 3: **Examples of a regulatory sandbox**

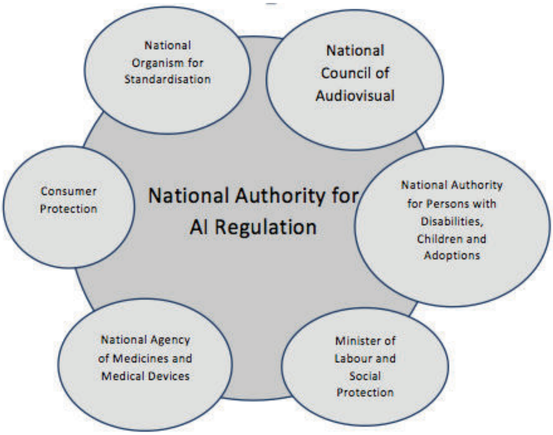
Regulatory sandbox	Examples
AI regulatory testing spaces create a controlled environment for testing innovative technologies for a limited period, based on a test plan agreed by the authorities	Testing space for AI regulation: • Area for testing autonomous systems for product delivery • Area for testing autonomous vehicles (e.g., drones) • Data spaces
	The right to an explanation: • Refusal by the public authorities to grant social assistance to a person based on the calculations or recommendation of an AI-based system • Refusal to issue a visa based on the recommendation of an AI-based system • Refusal by the local administration based on the recommendation of an AI-based system to approve the organisation of an event • Refusal by the legal system to approve parole based on an evaluation based on an AI application

Source: Universitatea Tehnică din Cluj-Napoca, 2021.

Another interesting development of AI policy is the Catalogue for AI incidents that is practically a database of incidents with AI systems, where an association with press reports could be made. In this way, the vulnerabilities of AI system use will be identified, supporting and improving activities that are using an AI system from industry, the economy, public administration, education, medicine etc.

An important policy recommendation regarding AI policy is to ensure coordination and cooperation with other national actors, among which it is worth mentioning: standardisation agencies (e.g., the National Organisation for Standardisation) – for the elaboration of AI standards in various fields; the National Council for Audiovisual, for news elaborated by AI systems, with filters for child content; Consumer Protection for cases of consumer manipulation by algorithms which generate false reviews; the National Agency for Medicines and Medical Devices from Romania, for monitoring and certification of the AI system installed on medical devices; several ministries, such as the Labour and Social Protection Ministry, the Education Ministry, The Ministry of the Economy for the establishing of national strategies for the development of competencies (similar to other MSs); the National Authority for the Rights of Persons with Disabilities, Children and Adoptions for AI toys, AI voice assistants, video games and emotional AI; and audit centres for monitoring and accreditation (Universitatea Tehnică din Cluj-Napoca, 2021). Further, on the international level the AI policy and authority could cooperate with data provider agencies such as the European Data Space or the national agency for regulating AI of other MSs (Universitatea Tehnică din Cluj-Napoca, 2021). The figure below illustrates the cooperation among agencies in the field of AI regulation.

Figure 16: **Cooperation of agencies regulating AI**


National Organism for Standardisation
National Council of Audiovisual
Consumer Protection
National Authority for AI Regulation
National Authority for Persons with Disabilities, Children and Adoptions
National Agency of Medicines and Medical Devices
Minister of Labour and Social Protection

Source: Universitatea Tehnică din Cluj-Napoca, 2021.

An important actor for cooperation in the elaboration of AI policy is the academic community, and continuous collaboration will improve the adaptation of policy to the latest trends in the fields. It could develop AI good-practice models and standards, providing the competencies for the better understanding and use of AI, increasing the awareness of the public administration and economy regarding the new risks and technologies, signalling cases where AI developers do not comply with the ethical norms for AI, promoting partnership with the local administration for the development of competencies regarding AI use, support for the introduction of AI laboratories on the high-school level (pre-university level in Romania).

7 Conclusions

Regulation of the AI field in Romania is needed to increase public trust in AI applications and create the framework for data sharing. Without data, there cannot be the development of automatic learning or innovations. It is also necessary to establish a national authority to manage the regulations in the field of AI capable of reducing the barriers to the development and use of AI: open access to databases and open linked data.

The regulations for public administration will support the provision of open data and application norms, including the once-only principle (which forbids the public administration from twice requesting data). The principles regulating the AI field on the global and European levels must also be transposed in Romania.

An important role in the elaboration of AI policy in Romania will be played by the Scientific and Ethical Council, which will act as a valuable liaison between public authorities (Ministry of Research, Innovation and Digitalisation), academic communities and the business environment.

The exponential development of AI, the potential Romania holds in terms of expertise, human resources, the motivation of the business environment to support the development of providers, are important indicators of the future that Romania could have on the European level. Yet, some of these objectives will not be achieved by Romania unless the strategic approach becomes operational in the sense of the elaboration of a national strategy for AI and other legislative components, such as by-laws and measures.

Even though AI is associated with the progress and development of societies and economies, one should consider the negative impacts it might bring in terms of the security of users, ethical rights, which explains why control and monitoring measures as well as restrictions should be clearly stipulated.

In conclusion, the changes to our lives are approaching sooner that we could have foreseen, the AI technology will impact our day-to-day life, and thus from the strategic and legislative perspectives it is necessary to be prepared to better face the near future. Romania today has a chance to build its future in terms of technology by continuing the initiatives of entrepreneurial environment and of research, development and innovation in the AI field, which has occurred outside the legislative framework. Efforts should be made by Romania to achieve a spectacular qualitative leap so that acquires its own place on the international map of AI.

REFERENCES

- Afina, Y. (2023). AI governance must balance creativity with sensitivity. Retrieved 4 July 2023 from: https://www.chathamhouse.org/2023/06/ai-governance-must-balance-creativity-sensitivity.
- Autoritatea pentru Digitalizarea României (ADR). (2023). Autoritatea pentru Digitalizarea României a lansat dezbaterea privind necesitatea şi opţiunile de reglementare în domeniul inteligenţei artificiale. Retrieved 21 July 2023 from: https://www.adr.gov.ro/autoritatea-pentru-digitalizarea-romaniei-a-lansat-dezbaterea-privind-necesitatea-si-optiunile-de-reglementare-in-domeniul-inteligentei-artificiale/.
- Chen, C., Shi, Y., Zhang, P., & Ding, C. (2021). A Cross-Country Comparison of Fiscal Policy Responses to the COVID-19 Global Pandemic. Journal of Comparative Policy Analysis: Research and Practice, 23(2), 262-273.
- Clark, J., Murdick, D., Perset, K., Grobelnik, M. (2022). The OECD Framework for Classifying AI Systems to assess policy challenges and ensure international standards in AI. Retrieved 4 July 2023 from: https://oecd.ai/en/wonk/classification.
- Economedia.ro (2023). Ministrul Digitalizării estimează că instrumentele de inteligență artificială vor duce la o creștere a colectării de TVA cu până la 1%. 6 November 2023. Retrieved 20 November 2023 from: https://economedia.ro/ministrul-digitalizarii-estimeaza-ca-instrumentele-de-inteligenta-artificiala-vor-duce-la-o-crestere-a-colectarii-de-tva-cu-pana-la-1.html.
- Elliffe, C. (2021). Taxing the Digital Economy: Theory, Policy and Practice. Cambridge University Press.
- EURES (2020). Patru feluri în care COVID-19 ne-a schimbat modul de lucru. Retrieved 23 July 2023 from: https://eures.ec.europa.eu/four-ways-covid-19-has-changed-way-we-work-2020-06-18_ro.
- European Commission. (2018a). Coordinated Plan on Artificial Intelligence COM/2018/795. Retrieved 21 July 2023: https://eur-lex.europa.eu/legal-content/RO/TXT/HTML/?uri=CELEX:52018DC0795.
- European Commission. (2018b). Artificial Intelligence for Europe COM (2018) 237 final. Retrieved 21 July 2023 from: https://www.eumonitor.eu/9353000/1/j4nvke1fm2yd1u0_j9vvik7m1c3gyxp/vknuq8ls10yp/v=s7z/f=/com(2018)237_en.pdf.
- European Commission. (2018c). Artificial Intelligence for Europe. SWD(2018) 137 final. Retrieved 21 July 2023 from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018SC0137.

- European Commission. (2020a). WHITE PAPER On Artificial Intelligence - A European approach to excellence and trust. Retrieved 21 July 2023 from: https://commission.europa.eu/system/files/2020-02/commission-white-paper-artificial-intelligence-feb2020_en.pdf.
- European Commission. (2020b). A European Strategy for Data. COM/2020/66 final. Retrieved 21 July from: https://eur-lex.europa.eu/legal-content/RO/TXT/PDF/?uri=CELEX:52020DC0066.
- European Commission (2020c). Digital Education Action Plan (2021-2027). COM/2020/624 final. Retrieved 21 July 2023 from: https://eur-lex.europa.eu/legal-content/RO/TXT/PDF/?uri=CELEX:52020DC0624.
- European Commission (2021a). AI Act. COM(2021) 206 final. Retrieved 21 July 2023 from : https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0023.02/DOC_1&format=PDF.
- European Commission. (2021b). The European Fiscal Board assesses the appropriate fiscal stance for the euro area in 2022. Retrieved 5 October 2021 from: https://ec.europa.eu/info/publications/european-fiscal-board-assesses-appropriate-fiscal-stance-euro-area-2022_en.
- European Parliament. (2023). MEPs ready to negotiate first-ever rules for safe and transparent AI. Press release on 14.06.2023. Retrieved 21 July 2023 from: https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rules-for-safe-and-transparent-ai.
- Făniță, A. (2023). Inteligența artificială în industria financiară. Retrieved 19 November 2023 from: https://pkffinconta.ro/inteligenta-artificiala-in-industria-financiara/.
- Financialintelligence.ro (2023). Studiu: 75% dintre companii cred că organizaţia cu cea mai avansată inteligenţă artificială generativă câştigă. 20 September 2023. Retrieved 20 Novermber 2023 from: https://financialintelligence.ro/studiu-75-dintre-companii-cred-ca-organizatia-cu-cea-mai-avansata-inteligenta-artificiala-generativa-castiga/.
- GDPR – General Data Protection Regulations (2016). Retrieved 23 July 2023 from: https://www.dataprotection.ro/?page=noua%20_pagina_regulamentul_GDPR.
- Jurma, O. (2023). "Este pregătită România pentru educația asistată de Inteligența Artificială?", republica.ro. 18 May 2023. Retrieved 19 November 2023 from https://republica.ro/este-pregatita-romania-pentru-educatia-asistata-de-inteligenta-artificiala.
- Lopez Cobo, M., De Prato, G. (Eds.) (2022). AI Watch Index 2021. Retrieved 21 July 2023 from: file:///C:/Users/Zsolt/Downloads/jrc128744_ai_watch_index_2021_1_1%20(1).pdf.

- Makin, A. J., & Layton, A. (2021). The global fiscal response to COVID-19: Risks and repercussions. Economic Analysis and Policy, 69, 340-349.
- Malatras, A., Dede, G. (2020). AI CYBERSECURITY CHALLENGES. Threat Landscape for Artificial Intelligence. ENISA. Retrieved 4 July 2023 from: file:///C:/Users/Zsolt/Downloads/ENISA%20Report%20-%20Artificial%20Intelligence%20Cybersecurity%20Challenges.pdf.
- Ministerul Cercetării, Inovării și Digitalizării/Ministry of Research, Innovation and Digitization of Romania. 2023. România are un Consiliu științific și de etică în inteligența artificială. Retrieved 23 July 2023 from: https://www.mcid.gov.ro/romania-are-un-consiliu-stiintific-si-de-etica-in-inteligenta-artificiala-11413/.
- Neagu, L. (2021). Doar 6% dintre companiile din România folosesc aplicații cu inteligență artificială. Economica.net. Retrieved 20 Novermer 2023 from: https://www.economica.net/doar-6-dintre-companiile-din-romania-folosesc-aplicatii-cu-inteligenta-artificiala_505419.html
- OECD. (2022). "OECD Framework for the Classification of AI systems", OECD Digital Economy Papers, OECD Publishing, Paris. Retrieved 4 July 2023 from: https://doi.org/10.1787/cb6d9eca-en.
- Operational Program Capacity Administrative (OPCA) (2023). Cadru strategic pentru adoptarea și utilizarea de tehnologii inovative în administrația publică 2021 – 2027 – soluții pentru eficientizarea activității. Retrieved 21 July 2023: https://www.adr.gov.ro/wp-content/uploads/2023/05/Propunere-Cadru-strategic-national-IA.pdf.
- Operational Program Capacity Administrative (OPCA) (2020). E-România O politică publică în domeniul e-guvernării. Retrieved 21 July from: https://www.adr.gov.ro/wp-content/uploads/2020/08/Livrabil-A12_Propunere-de-politica-publica-in-domeniul-e-guvernarii.pdf.
- Operational Program Capacity Administrative (OPCA) (2018). Document de Politică Industrială a României. Retrieved 21 July 2023 from: https://oldeconomie.gov.ro/images/politici-industriale/SIPOCA7/Document%20de%20Politica%20Industriala%20a%20Romaniei.pdf.
- Ordonanță de Urgență nr 58/2022 pentru modificarea și completarea unor acte normative din domeniul protecției consumatorilor. Retrieved 23 July 2023 from: https://lege5.ro/gratuit/geytcnzvhayds/art-ii-ordonanta-de-urgenta-58-2022?dp=gq3diobtgmztioi.
- Ordonanța de Urgență nr. 45/2020 pentru completarea Ordonanței Guvernului nr. 137/2000 privind prevenirea și sancționarea tuturor formelor de discriminare. Retrieved 23 July from: https://www.cncd.ro/wp-content/uploads/2021/02/OUG-45_2020.pdf.

- Pascu, C., Lourenco, M.B. (2023). Artificial Intelligence and Cybersecurity Research. ENISA. Retrieved 4 July 2023 from: file:///C:/Users/Zsolt/Downloads/Artificial%20Intelligence%20and%20Cybersecurity%20Research%20(1).pdf.
- Planul National de Redresare si Rezilienta. (2021). Retrieved 21 July 2023 from: https://gov.ro/ro/stiri/unda-verde-de-la-comisia-europeana-pentru-pnrr&page=1.
- Presedintia României. (2021). România Educată. Retrieved 21 July 2023 from: http://www.romaniaeducata.eu/wp-content/uploads/2021/07/Raport-Romania-Educata-14-iulie-2021.pdf.
- Regina Maria. (2023). Centrele de imagistică Regina Maria – pionier în domeniul aplicațiilor de inteligență artificială în imagistică medicală. Retrieved 17 November 2023 from https://www.reginamaria.ro/articole-medicale/centrele-de-imagistica-regina-maria-pionier-do-meniul-aplicatiilor-de-inteligenta-artificiala.
- Romanian Government. (2022). Strategia Națională de Cercetare, Inovare și Specializare Inteligentă 2022-2027. Retrieved 21 July from : https://www.mcid.gov.ro/wp-content/uploads/2022/12/strategia-na-ional-de-cercetare-inovare-i-specializare-inteligent-2022-2027.pdf.
- Romanian Government. (2021a). Strategia Natională pentru ocuparea forței de muncă 2022-2027. Retrieved 21 July 2023 from: https://mmuncii.ro/j33/images/Documente/MMPS/SNOFM_2021-2027.pdf.
- Romanian Government. (2021b). Strategia de securitate cibernetică a României, pentru perioada 2022-2027. HG nr. 1321/2021. Retrieved 21 July 2023 from: https://www.cybercommand.ro/webroot/fileslib/upload/files/conducere/hotararea-nr-1321-2021-privind-aprobarea-strategiei-de-securitate-cibernetica-a-romaniei-pentru-perioada-2022-2027-precum-si-a-planului-de-actiun-e-pentru-implementarea-strategiei-de-securitate-ciberne%20(1).pdf.
- Universitatea Tehnică din Cluj-Napoca (2021). Elaborarea cadrului strategic național în domeniul inteligenței artificiale. Raport de consultare generative[. Retrieved 21 July 2023: https://strategie-ia.utcluj.ro/docs/POCA-CSN-IA_ConsultareGenerativa_Raport.pdf.

A liberal future in a united Europe

/europeanliberalforum
@eurliberalforum
#ELFevent

liberalforum.eu

This publication was co-financed by the European Parliament. The European Parliament is not responsible for the content of this publication, or for any use that may be made of it.