United Nations
Educational, Scientific and
Cultural Organization

UNESCO Institute
for Information Technologies
in Education

Digital Transformation of Education

Artificial Intelligence

Media and Information Literacy, Human Rights and Freedom of Expression

THE NEXTMINDS

COLLECTION OF PAPERS

Authors
Igor Shnurenko
Tatiana Murovana
Ibrahim Kushchu

Artificial Intelligence

Media and Information Literacy, Human Rights and Freedom of Expression

Editors
Ibrahim Kushchu
Tuba Demirel

UNESCO Institute for Information Technologies in Education in cooperation with the TheNextMinds

Authors: Igor Shnurenko, Tatiana Murovana, Ibrahim Kushchu

Editors: Ibrahim Kushchu, Tuba Demirel

The choice and the presentation of facts in this publication and the opinions expressed therein are not necessarily those of UNESCO or TheNextMinds and do not commit the Organizations. The designations employed and the presentation of material throughout this publication do not imply the expression of any opinion whatsoever on the part of UNESCO or TheNextMinds concerning the legal status of any country, territory, city or area of its authorities, or the delimitation of its frontiers or boundaries. Whilst the information in this publication is believed to be true and accurate at the time of publication, UNESCO or TheNextMinds cannot accept any legal responsibility or liability to any person or entity concerning to any loss or damage arising from the information contained in this publication.

Published by TheNextMinds for the UNESCO Institute for Information Technologies in Education

UNESCO IITE, 8, Bldg. 3, Kedrova Street, Moscow, 117292, Russian Federation

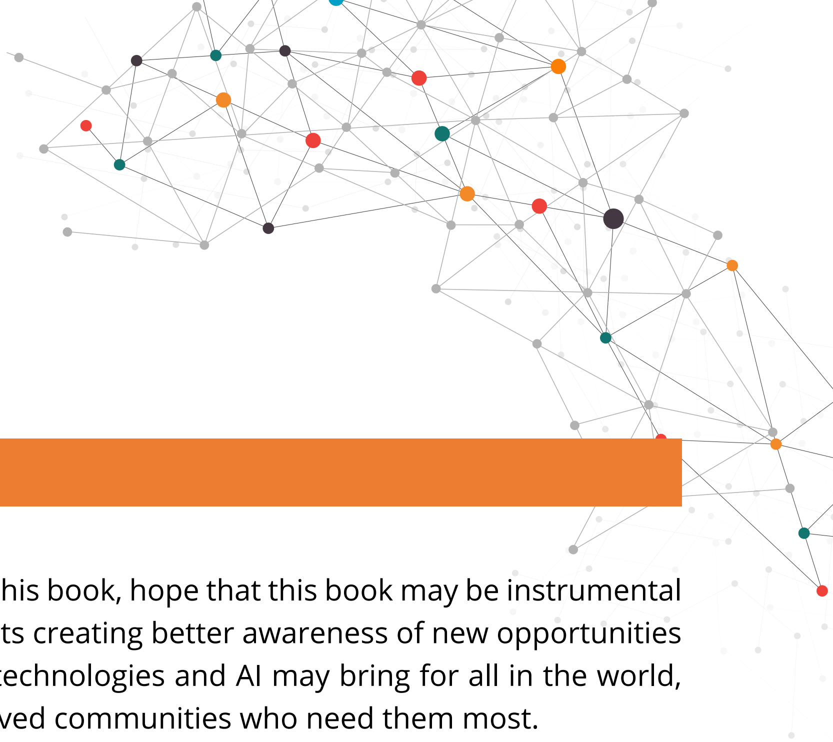
Tel: +7 499 1292990, Fax: +7 499 1291225, E-mail: Liste.info.iite@unesco.org, www.iite.unesco.org

TheNextMinds, Suite 10, 42 Brunswick Terrace, Hove, BN3 1HA, United Kingdom

Tel: +447740143135, Email: info@thenextminds.com, www.thenextminds.com

ISBN: 978-1-912037-64-3

Cite this content as: UNESCO IITE and TheNextMinds. (2020). Artificial Intelligence: Media and Information Literacy, Human Rights and Freedom of Expression. Collection of Papers. Authors: Igor Shnurenko, Tatiana Murovana, Ibrahim Kushchu. Editors: Ibrahim Kushchu, Tuba Demirel.



DEDICATION

We, editors and authors of this book, hope that this book may be instrumental in supporting all other efforts creating better awareness of new opportunities and risks that information technologies and AI may bring for all in the world, especially for the under-served communities who need them most.

CONTENTS

ACKNOWLEDGEMENTS

A number of people and organizations have been very helpful in contributing to the production of this book. We are thankful to UNESCO IITE for leading the work and funding the publication process.

The core content of this book is based on TheNextMinds' training, which was held during the 22nd, 23rd and 24th of January, 2020, under the auspices of the Ministry of Culture and Media of Serbia, on 'AI: Freedom of Speech and Media and Information Literacy' for various stakeholders, including public sector institutions, civil society organizations, media industry and academia. We are grateful to the Ministry and the leaders, Ms. Maja Zaric and Ms. Slavica Trifunovic, making this training possible and all attendees being present and contributing to the training.

We are also grateful to Ms. Ruth Clar Spence Gorman, Ms. Lale Omay, Dr. Ahmet Koçak and Mr. Ghislain Charles Sireilles who devoted a considerable time in helping with proof reading the English of certain sections of the book. Also supportive was Ms. Ayse Aydinoglu in preparing a background research for certain sections of the AI-related content, for which we thank her very much. Other helpful hands included, among others in the larger team, proof editors and graphic designers.

EXECUTIVE SUMMARY

The relationship between media and information literacy (MIL) and human rights (HR), especially, the right to access to information, education and freedom of expression (FoE), is undeniably strong. MIL, supported by the right to education, is an essential requirement for citizens to be able to access, understand, analyse, create and express media content as well as to be able to improve on the realisation of their relevant human rights. So, MIL would be dry and naked without the envelope of human rights perspective.

Digital tools and platforms have taken MIL further from being unidimensional to being interactive and dynamic. Various (digital) media, citizens, content producers, regulators (i.e. governments) and other stakeholders now operate in a dynamic MIL ecosystem, which is continuously changing and evolving. Without any doubt, in recent years, artificial intelligence (AI) with its supporting technologies, such as the cloud, big data, Internet of things (IoT) and (mobile) connectivity, are all having a disruptive influence on this ecosystem. The successes of MIL efforts primarily rely on the successes in understanding AI and its adoption.

The book in your hand, to serve this purpose, presents an exploration of dynamic relationships among AI and relevant emerging technologies, MIL, human rights and freedom of expression. This dynamic relationship is explored from perspectives of three major areas of concern in MIL:

- **Passive MIL:** accessing, using and adopting media and information,
- **Active MIL:** creating, disseminating, analysing, evaluating, interacting with and influencing media and information, and
- **Influential MIL:** realising and practicing media and information rights.

The first part of the book creates technical basis for AI and relevant emerging technologies as they relate to MIL and human rights. First, the strength of AI stemming from new technologies such as big data, the IoT and all forms of connectivity is presented. In this way, the power and the limits of such power is established as they relate to the above three areas of MIL. While most AI techniques have not been advancing drastically since their birth, the significant developments in emerging relevant technologies are revolutionizing the AI systems. However this revolution is limited with narrow AI where bounded domain-specific problems are solved competently. Still, AI plays a disruptive role in extending and augmenting the capabilities of citizens and other stakeholders of the MIL ecosystem. Regardless of who the stakeholder or the actor is, passive, active and influential MIL are enhanced by the use of AI in various digital and media platforms and environments.

The second part of the book is a discussion exploring how AI may contribute or hinder the development of MIL competencies. This part starts with presenting a brief on the views, efforts and position of UNESCO on MIL. This brief contains a dateline of activities in making MIL a widely recognised essential line of activity as well as in developing MIL by closely following the changing digital environments. Then the role and significance of MIL in social and economic context is presented from a modernized point of view. According to this point of view, there are increasing complexity and uncertainty in developing MIL competencies, which may require a special set of tools, approaches and practices. Also important are the mind-blowing changes in digital media environments, which brings about impositions to MIL to discover new ways of learning, using and adopting of competencies as they relate to information creation, access and utilisation. Next, the issues around what is and what is not 'true' in media are discussed in terms of their vagueness and departure from reality. These discussions are essential for developing approaches to how digital technologies, especially AI, may create challenges for and may support MIL efforts in empowering citizens. Digital technologies with AI have a significant impact on MIL and its future development in terms of negatives and positives. They change the media environment and consequently the experiences of people in these environments. While media and content become conveniently accessible and usable by professionals (i.e. journalists) and ordinary citizens, nature and most importantly, the quality of experiences change drastically. While AI and digital tools may help with faster, precise and more directed media use and content delivery, it is not very clear how the users are benefiting from these advances. A greater complexity arises when the powers and intentions of data holders to serve profit-making machines are recognised. This part closes with detailed discussions on how these issues come about as challenges and what may be the opportunities to empower people via MIL using new technologies.

The third part of the book explores in great detail how AI influences human rights and freedom of expression. It has three chapters:

- The first one deals with an extensive discussion of AI in the context of human rights and freedom of speech; what kind of AI functions and techniques are being used in generating, distributing and using information on line and how this relates to human rights; and what the future trends look like in this area. More specifically, after providing a context for AI in human rights, the chapter goes through various tools and techniques of AI as they relate to human rights and freedom of expression. For example, positive and negative impacts of personalisation, efforts in machine understanding of human emotions, content moderation, especially, in social media, surveillance, propaganda and disinformation, fake news and deep fakes are all presented here. Also presented are some of the ethical concerns related to privacy and use of personal data, bias, information equality and discrimination caused by data and AI systems and creating responsible AI with appropriate justifications. Finally, the chapter ends with the discussion of some of the positive and negative approaches that AI can help with improving human rights. On the positive side, AI can help with reducing discrimination, bias, improving focussed delivery and equality in access to information and finding ways to improve social life. But AI may also imply that human rights and free will may be in significant danger, and new approaches to designing human rights and education for MIL may be needed.

- The second chapter in this part deals with the need for regulating AI at local, national and also global levels. In this respect, some of the regulatory efforts in the EU, USA and China are evaluated. It is argued that some of the critical topics for regulations are directly related to the prevention of potential risks that may arise from the fast and widening spread of AI systems. As explained in this chapter, such risks range from being part of AI developmental processes to existential or unintended consequences.

- The third chapter of this part presents possible responses to human challenges posed by AI. These responses may be in terms of legislations and policies. One major concern is that intentions and interests of governments and large tech companies may not be in agreement and this may require strong regulatory and legislative actions. Other concerns include issues related to creating human entered AI systems with human oversight while keeping the progress of AI beneficial and safe for humanity.

The core content of part three in this book is related to active and influential MIL, where an extensive analysis of how people interact with media and information content in advanced technological environments while preserving, if not, improving situations with human rights.

The book ends with a summary and conclusions focussing on AI and the most recent AI trends such as brain uploading (i.e. The Human Connectome Project) and brain-machine interfaces (i.e. Neuralink) and their influence on MIL and human rights while the man and machine converge to create a trans-intelligence.

1. Artificial Intelligence & Emerging Technologies

Ibrahim Kushchu

Artificial intelligence (AI) has already come into our daily lives mainly through mobile devices and the Internet. Similarly, governments and businesses are increasingly making use of AI tools and techniques to solve business problems and improve many business processes, especially online ones. Such developments bring about new realities to the social life that may not have been experienced before. These realities lead most of us to spend considerable time on various media environments and social media platforms for professional or social reasons. Most of these online interactions are designed and driven by new technologies and especially by AI.

The sections that follow provide a non-technical introduction to AI and how it may relate to media and information literacy (MIL) and human rights (HR). The purpose is providing a context for understanding AI's core concepts and capabilities and then relating the powers and realities of AI to its influence on MIL and HR. So, the first section is all about AI and description of relevant technologies. Readers familiar with AI and its supporting technologies such as data science, big data, the cloud and industry 4.0 may read the following section selectively and perhaps skip the section on emerging new technologies supporting AI. However, the section on AI and its relation to MIL and HR provides a basis for further discussions in the book's following chapters and, therefore, is important.

1.1 Artificial Intelligence

The interactivity on the Internet has evolved in two broad dimensions. One is mobility, and the other one is the intelligence. With the developments in connection capacity and speed, the networked digital solutions moved to the cloud and intercommunicating devices. Mobility has already proven its benefits, and with intelligence added, it will provide much better benefits to individuals, businesses and societies.

What is the nature of intelligence in today's artificial intelligence (AI)? Conventional wisdom holds that AI is aimed at imitating human intelligence. But we still do not know much about how human intelligence or the mind works. In addition, any design is limited to the designer's abilities. That goes for the design of intelligence, too. Recently, however, AI is more concerned with simulating *any* intelligent task or behaviour that can be observed in the worlds of humans, animals or plants. By doing so, AI started to show promising solutions for industry and businesses as well as our daily lives. *AI, then, is concerned with any intelligent activity that the machines can show.* These activities can range from a robot's behaviour to grasp an object and move into another position to chatbots conversing with consumers via natural language and helping them to make a purchasing decision.

Although at the moment AI focuses on many different tasks in narrow domains, the idea of simulating the whole of human intelligence is still there. AI systems could be generally divided into three main categories [8]:

- **Artificial Narrow Intelligence (ANI) or Domain-Specific AI (Weak AI):** This category is almost all AI systems that are in use at the moment. This type of AI is characterised by specific domains for which a model can be built based on rules or boundaries governing the domain. For example, it is no longer difficult to have an AI which is excellent in chess or other games as we already know the rules of the games describing chess or a similar game. AI systems can master quickly such specific domains and provide intelligent solutions for that domain. This is often called Narrow Intelligence or *The Weak AI* as the intelligence attained here is far from general human-level intelligence and is only domain-specific.
- **Artificial General Intelligence (AGI) or Strong AI:** In this type of AI systems, the intelligence learned in one domain can be generalised to similar domains or an unrelated domain as human beings do. For example, humans learn to walk mostly indoors and on smooth surfaces. But this behaviour, as we grow, is generalised to walk on uneven terrains

or roads with even some degree of slopes. Generalisation in AI is very important both for domain-specific tasks and also for reaching human-level intelligence. However, for so many years in the history of AI work, the issue of generalisation is proven to be hard especially at a human level. Therefore, such problems are often named as *hard AI* or *strong AI* problems to differentiate it from domain-specific AI (weak AI) with narrow intelligence.

- **Artificial Super Intelligence (ASI):** In this type of AI systems, machines are considered more intelligent than humans in almost every aspect. This is an imaginary scenario that may be possible, but given the current state of AI such machines are not around yet. This scenario also touches issues such as robots (i.e. artificially intelligent machines) controlling the humans and the world of self-replicating super intelligence dominating future civilisations. At this moment in time, there is not enough evidence to support this hypothetical situation.

It is important to differentiate these AI terms and approaches in order to understand the real capabilities of AI and the boundaries of hype around AI. Some of the above concepts may be misleading when they are used out of the context of Narrow AI, where most of the successes of current AI systems exist.

1.1.1 Revisiting The Narrow Artificial Intelligence

Currently, most of the artificial intelligence problems are solved for a specific task or a domain. Although there are degrees of generalisation within a specific domain or for a given task these are still considered as part of the Weak AI with narrow intelligence. The intelligence exerted in these AI solutions are highly credible but are often task-specific. Sometimes, they are optimal solutions for a model world for which all the rules defining it are identifiable as they are finite and exact such as in chess. AI systems could learn these rules and reach a best winning strategy.

If we keep the discussion in narrow artificial intelligence or the weak AI, it is possible to see generalising (not completely general) AI solutions which can be at a human level or better than human level. It should always be kept in mind that these are all for a given task or a domain. So, as long as AI performs better than humans for specific tasks, we can infer that AI may show *super intelligence* only for that specific task at hand. The same is true for general intelligence in particular domain-specific tasks. This is often referred to as 'generalisation' as opposed to 'general intelligence'. For example, given past data (millions of pictures of cats and dogs), an AI system may learn to differentiate between cats and dogs. When the same system is presented with a new previously unseen picture(s) and when the system correctly identifies these unseen pictures as cats and dogs, this AI system is said to have generalised what is learned from the past data used as input (seen by the system) to an unseen pictures or cases. In this respect, it is possible to see better generalisation than humans can do, but this does not mean that we have reached a general AI intelligence. It only means that there may be a general intelligence for that particular task. This would never mean super intelligence either.

We know that AI systems are much better in predicting brain tumours [3] than conventional radiological methods. We also know that AI now can beat the world champions at the games of Chess and Go. None of these, yet, means that we have super intelligence or general intelligence. What it all means is that AI has been better performing than humans on particular tasks. As we witness artificial intelligence being better in discovering brain tumours, it is interesting to ask these questions - how is the system expected to improve in the future and will it be replacing human doctors?

1.2 Augmented Intelligence: Extending Human and Machine Intelligence

With the developments in the digitally supported businesses and life events, one type of intelligence will probably be very significant: augmented intelligence. It refers to a synergistic intelligence achieved through human and machine collaborations. This creates room to improve both machine intelligence through human support and human intelligence through machine support.

Conventionally, one of the most significant developments in AI has always been intelligent decision support systems. In decision support systems, a 'machine learns' about a particular domain and makes inferences to provide solutions. For example, doctors have used such systems in identifying and diagnosing some illnesses. The same idea is still being used to support not only doctors but also other experts such as lawyers. In this scenario, the machine helps a human expert in making better decisions.

Similarly, there are cases where machines learn from human experts. There are ongoing works where robots observe a human expert doing a task and the robot learns from hand movements, etc. In this case through interaction with a human expert a robot can learn to perform a task by imitation.

There are also numerous examples of using intelligence to replace and control missing hands, arms and legs and so improving abilities of disabled portions of the population.

With the developments in AI together with infrastructure and services offered via the cloud and IoT, there will be more and more examples of machine and human collaborations for improved Intelligence. These will be in the form of machine to machine, machine to human and human to machine interactions. The degree of interaction may even get as intimate as having artificial body parts to augment human behaviour [11].

1.3 How Does AI Work?

AI systems are basically software systems (or controllers for robots) using such techniques as machine learning and deep learning to solve problems in particular domains without hard coding all possibilities (i.e. algorithmic steps) in a software.

The figure below shows the components of a typical AI system. AI systems work on data gathered from various sources such as email, online data, internally held past data, etc. and that may have various forms, for example, audio, video or text.

The data scientist's job is to collect, store and understand data (making simple analysis via visualisation and descriptive statistics) to prepare the data for AI models. Data scientists' job quality is essential for AI systems to properly work. In fact, there is a saying that 'your AI is only as good as your data' and the data will have a direct influence on actions or decisions produced by AI systems.

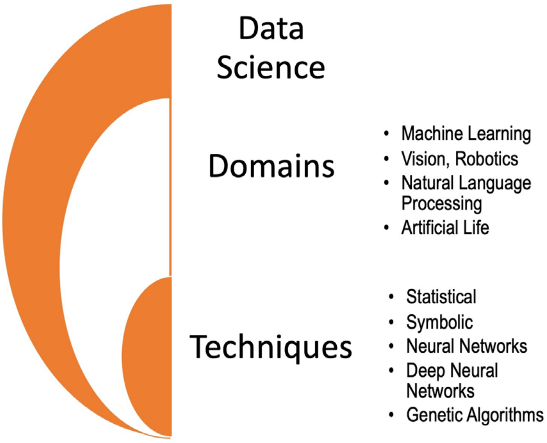


Figure 1. Components of AI solutions [5]

Once the data is prepared and there is some understanding about the data and an AI technique (i.e. a model) is chosen in relation to the problem at hand. AI techniques range from some well-known statistical techniques to those that are specifically developed to tackle AI problems such as Artificial Neural Networks or Genetic Algorithms. These techniques essentially perform some prediction, classification or clustering. The AI system's job is basically to construct patterns in order for the system to perform these tasks. Which technique to choose depends on the data at hand and the question(s) to be answered about the data or the nature of the problem. Statistical techniques are good for data preparation or even for simple clustering and classification problems, whereas AI-specific techniques are good for complex data and sophisticated problems. In such cases, capturing the data patterns is not possible or extremely difficult for the conventional statistical techniques.

The AI domains closely follow human cognitive, motor and vision abilities. For cognitive abilities machine learning, language understanding and speech processing can be typical techniques to use. Vision and robotics relate to software that controls movement, seeing and recognition. The nature of data, choice of technique and application domain goes hand in hand but in general any appropriate AI technique could be used for a given problem in any AI domain.

Machine learning is a special case of AI domain, where an AI *learning algorithm* (such as back-propagation) is used with a technique or model (such as artificial neural network) to learn from already available data and create an abstraction of the data. The model is first 'trained' on some portion of the available data and then is 'tested' on the data which was left out during the training and was unseen by the model. During training the model creates a representation of the data by discovering the rules or patterns that define the data set in order to make predictions or classifications related to the domain. Then this model is tested to make similar predictions or classifications over the unseen data. If the model works well with the unseen data, this AI system is said to have 'learned' the problem and it can be used in actual implementations to provide solutions to those new instances of the problem during actual use.

1.4 Digital Technologies, AI and The Societal Implications

Advances in computing and digital technologies have always had a direct influence in the ways we live our lives and how businesses and the social dynamics operate. This influence is also seen in the way we perform our daily routines such as using mobile devices for communication and actively participating on social media.

Perhaps all start with advances of digital technologies in the business world. Early applications of digital technologies helped many businesses to automate data and information flow, especially in the banking industry. This supportive role with the advances in Internet technologies have influenced businesses to discover new ways of operations from selling online to management of logistics with the help of advances in network and telecom technologies. Mobile technologies coupled with cloud computing have enhanced business operations in such a way that most businesses are now available anytime anywhere and most of the required computing infrastructures could easily be outsourced via cloud service providers. This picture reflects advanced methods of large data collection, processing and analysis and techniques to use processed data to improve business processes. This active influence of digital technologies transforms businesses drastically and put severe pressures on them to be fast, flexible and timely.

Inevitably, the changes in businesses due to digital technologies have direct influence on how governments and the society operate. Governments try to keep up with the changes in the private sector in order to improve its operations. And this creates demands and expectations in society to receive services and conveniences supported by the digital technologies.

All these changes around individuals and the society have been continuously changing the way we live our life and perform our daily tasks. For many, no day passes without checking emails, using mobile devices and applications for communication and making decisions and getting online for social media interactions.

AI systems are the newest and perhaps most influential of digital technologies. With the AI systems, the businesses are not only able to handle large data sets and provide speedy essential input to the operations, they are also able to predict, adapt to constant changes and be more flexible. Due to the complexity of the processes and the amount of large data in various forms, some of the critical processes cannot be automated using conventional programming methods and statistical techniques. Such complexities also arise through availability of large numbers of sensory devices used in, for example, production and logistics – the Internet of things (IoT). By introducing AI systems into these devices, new business processes often eliminate some of the jobs that human workers do. A new paradigm, industry 4.0, emerges as a result of such intelligent automation, which now dictates not only *how* the businesses operate but also *who* does the job. Many manufacturing sites can now operate fully automated with robots and without any human workers. AI now brings in unheard and unexpected innovations to the business world that many organisations will need to integrate, at least, to remain competitive and if not move further to lead the competitors.

All these technological developments influence and shape the society. Therefore, AI systems, in the first instance, shape our lives and social interactions through these technological progresses. In addition to these, there are many AI applications which are specifically developed for providing better services to individuals such as those used on mobile phones and in social media.

The advances towards industry 4.0 lead to creation of society 4.0. At the moment and increasingly in the future, there will be no individual members of the society who will not be connected to intelligent AI applications and services through mobile devices or intelligent IoT. We are already extending our body and mind through intelligent applications such as personal assistants, intelligent wearable devices and other systems and delegating our activities to AI systems helping us at home such as cooking or cleaning robots and systems which operate household apparatus.

The influence of AI on businesses and socio-economic dynamics is inevitable. This may have an unexpected form and consequence in the future. Therefore, there are important responsibilities of governments, non-profits and other international organisations in preparing for AI developments in terms of legalities, social responsibilities and ethics [5].

1.5 Emerging Automation Technologies

In this part, a number of the most important core components of technology influencing economic and social conditions of living in a networked and dynamical environment are presented. Nowadays' digital interactions are often faced with having to deal with large amounts of structured and unstructured data in many forms including audio and video. This has led to the developments of tools and techniques to deal with the 'big data'. Digital industry developed various ways to support utilization of big data via 'the cloud computing' or the cloud for short. The offerings of outsourced hardware and software at the cloud gave rise to 'IoT' and automation of operations and processes. This whole picture of this digital revolution is now converging to what is called 'industry 4.0'.

In the sections that follow, each of these core components are explained briefly with references to how they all shape the society, governments, businesses and the wider environment.

1.6 Introducing Automation Concepts

Industry 4.0 with its revolutionary approaches to automation is a new development influencing business dynamics significantly and in such a way that is not seen before. This influence is largely owed to developments in building and using 'the big data', 'the cloud' computing infrastructure and services and a connected planet with huge numbers of connected devices as the 'internet of things (IoT)'. Added to this connected world are intelligent automation and robotics as the core strengths of a new concept 'industry 4.0' [2].

The following sections visit all relevant technological components influencing AI, one by one and they are:

- big data,
- the cloud,
- IoT, and
- industry 4.0, automation and robotics.

Each of the sections below builds up to an understanding of new dynamic capabilities of AI being supported by any or all of those technological advances above.

1.6.1 Big Data

The recent developments in digital technologies and the social and business uses of these technologies have resulted in almost unimaginable accumulation of data from various sources. This accumulated data is now known as 'big data' to differentiate it from conventional and limited ways of collecting, storing and analysing data.

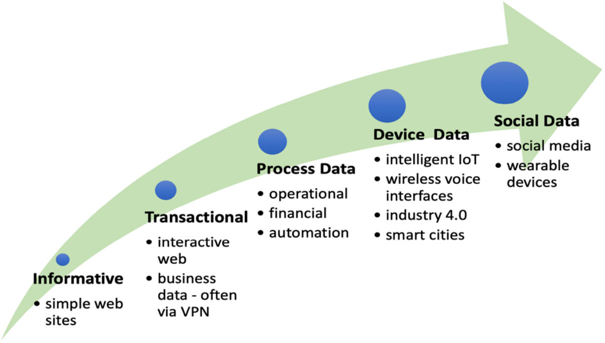


Figure 2. Evolution of networked data [5]

Figure 2 provides a picture of evolution of data types and sources. The big data can be structured, unstructured and semi-structured. They are gathered from multiple sources. Conventionally, some of these sources are intranets, spreadsheets and databases. However, recently these sources include emails, PDF documents, photos, videos, audio, social media posts, data resulting from various business operations and transactions on billions of pages of the internet sites on the World Wide Web.

The sources and the data are therefore often distributed but connected, related and often have inter dependencies. These properties provide an opportunity to create a platform for *shared relevant data* and this in turn makes big data so essential, significant and so valuable as to solve numerous problems and create new opportunities arising from a good data analysis and understanding of insights. There are in fact two major reasons why we are interested in big data:

1. discovering new and original insights in the data to support social and business lives in changing dynamical environments, and
2. solving an existing social or business problem better by gathering more useful information through big data.

In order to successfully benefit from these, it is important to carry out good data analytics using techniques learned from data science.

Big data has various properties which are defined as Vs of data and here we present 5Vs of data [6]:

- Volume: this refers to the size of the data, which is typically very large. Working with connected distributed big data is a huge challenge for data engineers.
- Velocity: this refers to speed of generating and processing data. The ideal speed in some cases requires businesses to have the ability to analyse and use the data in real time as it is generated, say on a website or during a production process.
- Variety: this refers to types and structure of data. While the technology allows processing of different types of data such as text, audio and video, the challenges related to analysis of unstructured and semi-structured data still remain.
- Veracity: this refers to quality and reliability of data. This is perhaps one of the most significant property that needs to be seriously taken into consideration in using big data. This relates to reassurances that the data

at hand is relevant to the problem in consideration and that the data can reliably help solve the business problem.

- Value: this refers to the worth of data. There are costs involved in collecting, analysing and using data and such costs should be justifiable when evaluating the value (i.e. return) of big data for the business.

The business of taking advantage of big data is becoming increasingly promising through the support of artificial intelligence and machine learning techniques, which significantly help to find unimaginable insights in the data and to create intelligent autonomous solutions based on data such as recommender systems or intelligent chatbots. Also important are opportunities created by data sharing apps and solutions that can support social and business processes. When such processes are connected and autonomous, many daily and business operations become fast and more accurate.

1.6.2 The Cloud

Another important development in the connected world includes the opportunities provided by 'the cloud' technology. The cloud technology is an essential Internet-based infrastructure and digital services to store data (though optional) somewhere on the planet (i.e. data centres) and to use relevant software services for gathering, sharing, analysing and using data. This is made possible with the developments in data transfer bandwidth and processing through the Internet. The data is stored on physical servers maintained and controlled by cloud providers who may also offer various software platforms and other relevant services.

There are various types of activities at the cloud [1]:

- Infrastructure as a Service (IaaS): in this case, the businesses may choose to outsource their requirements for computing hardware and computing capacity.

- Platform as a Service (PaaS): businesses may also decide to have a shared platform for running various software and application development environments at the cloud rather than purchasing, say, license-based solutions.

- Software as a Service (SaaS): similar to utilization of platforms, businesses may also choose to take benefit of using any software available on the cloud. This is sometimes called 'software on demand' service.

- Recovery as a Service (RaaS): while typical service level standards are offered by the cloud service providers, additional provisions for uninterrupted services in the face of any unfortunate events or disasters may also be offered.

Some of the cloud computing providers include well-known world leading companies such as Apple (iCloud), Google (Gmail and GoogleDocs), Microsoft (OneDrive) Facebook, Amazon, etc.

While the cloud offers encrypted services for data protection and for security of cloud operations, it is still important to make considerations for the protection of privacy and security. Some businesses may choose to keep data within their computing facilities but use some of the cloud services. So, businesses may need to decide on a certain degree of computing operations to outsource. In any case, legal authorities may still demand access to a business' cloud content and services.

Cloud computing provides better resources and computing power compared to average in house computing facilities, laptops, smart-phones, etc. especially for heavy automation applications. In addition, cloud computing may serve as a platform with other types of provision of digital services where businesses can automate some management operations through this platform. Such process of automation reduces drastically any possibility of human errors. The capabilities of cloud operations may also be enhanced by using AI techniques for automation. Companies such as Google, Amazon, Microsoft and IBM offer artificial intelligence capabilities in the following forms.

- machine learning platform,
- computer vision,
- speech recognition,
- text analysis, and
- dynamic translation.

Businesses enhance their competitive advantages by signing up for cloud-based AI services.

1.6.3 The Internet of Things

Another important digital development underlying the developments of industry 4.0 and processes of automation is what is called *the internet of things (IoT).* IoT is simply a system of interrelated computing devices with sensors. Devices may be objects such as mechanical and digital machines, or may be used on plants, animals or people with unique identifiers (UIDs). These devices have the ability to transfer data over a network and often without human interaction. This is also called machine to machine communication or M2M for short. It is also possible that these devices are interfaces interacting with humans via voice or text, in which case it is a human-machine communication (H2M or M2H).

A good example for this is the MIT's Oxygen project which explores human-centric pervasive computing where information is seen as freely available like the oxygen [7].

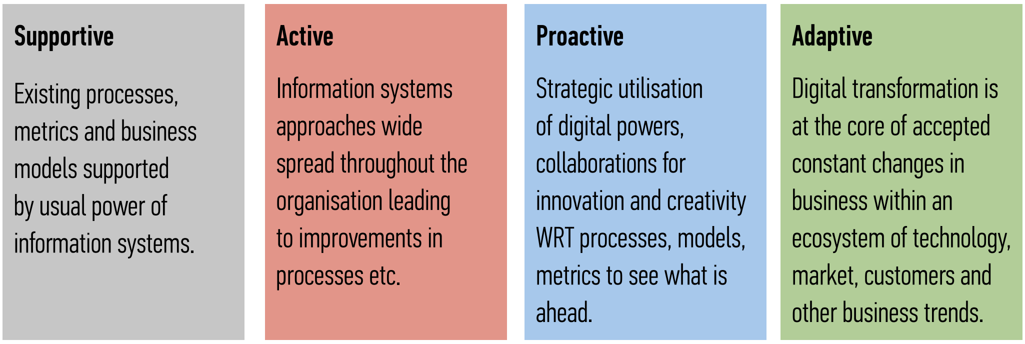
Coupled with artificial intelligence, which can analyse data that smart sensors and devices produce, IoT has proven useful in many operations. Some of these include making measure of temperature, pressure, humidity, air quality, vibration and sound. Yet some others enable, for example, making operational predictions many times earlier and with greater accuracy than threshold-based monitoring systems. It this way, IoT help to avoid unplanned downtime and increase operating efficiency. Speech recognition and computer vision used in IoT can help eliminate human review. While enhancing risk management, IoT can also open new approaches to development of new products and services owing to availability of data.

The IoT creates an essential environment for automation, for example, in production, marketing and logistics domains and also for the connection of businesses to the environment such as in smart cities and smart government. These processes can be improved via horizontal and vertical digital systems integration. The enablers for such automation include big data analytics and use, cloud computing and artificial intelligence as an enhancer. In this case the overall systems may sometimes be called *The Industrial and Intelligent Internet of Things*. In addition, the enhancements in IoT and automation may extend to intelligent simulations, additive manufacturing or 3D printing and augmented reality – a composite view of computer-generated images on top of real world views.

1.6.4 Industry 4.0, Automation and Robotics

Industry 4.0 is considered as the fourth industrial revolution after the steam engine, the age of science and mass production and digital technologies and the Internet. Although it is often characterized as the automation and data exchange in manufacturing with factories without human and lights, it is more than that.

In order to understand what is meant by this, it is important to briefly explain how digital revolution has transformed businesses. This is depicted in the figure below. Over the years digital technologies transformed businesses by initially taking passive and supportive roles but more recently this role has changed to a proactive and an adaptive one.



Supportive

Existing processes, metrics and business models supported by usual power of information systems.

Active

Information systems approaches wide spread throughout the organisation leading to improvements in processes etc.

Proactive

Strategic utilisation of digital powers, collaborations for innovation and creativity WRT processes, models, metrics to see what is ahead.

Adaptive

Digital transformation is at the core of accepted constant changes in business within an ecosystem of technology, market, customers and other business trends.

Figure 3. Evolution of digital business transformation [5]

Industry 4.0 and automation, therefore, goes beyond the realm of any business organization by accepting and aiming to respond competitively to constant changes in the connected environment which is made up of the technology, the sector dynamics that the business is in, customers, other businesses and legal requirements. This state of business operations is made possible by utilization of big data, the cloud and the IoT as presented earlier.

In its narrow sense industry 4.0 can be seen as a revolutionary approach to enhancing capabilities of an organization through improving business processes as well as employing new methods of techniques to enhance operations of various departments.

In this sense, major developments are in the areas of Robotic Process Automation (RPA) and Intelligent Automation. RPAs in simplest sense are robots that work hard to perform certain and especially repetitive tasks without getting tired and with enhanced speed and precision. With an added intelligence, robots or chatbots (i.e. intelligent software which understands and responds to natural language interactions) perform tasks better than a human worker, in cooperation with a human or just as supporter in making business decisions. The automation then may help to significantly increase process efficiencies and improve customer experiences. A good example for the latter is online recommender systems or sales-bots which observe the patterns of behavior of customers and guide them through a purchasing processes. For example, customers could make an image-based search of a similar product that they like on a website. Also possible is to use virtual reality to allow customers to visualize how a certain furniture would look like in their room.

Industry 4.0 helps to build production systems that are more adaptive to changes in provision supplies or in any other critical interferences. It allows more efficient, more flexible and more reliable production processes. Some of these systems are available off the shelf. For example**,** Siemens' MindSphere is a cloud-based, an open IoT operating system to link products, plants, systems and machines. Using such a system can help businesses to design and control sophisticated and innovative production plants.

The developments in the field of robotics are primary causes which have driven innovations in industry 4.0 and automation. Robotics is a discipline that deals with design, construction, operation and use of robots through control, sensory feedback and information processing. There mainly two types of robots:

- **Industrial robots**: these are conventional robots used in factories and may come in different types such as Cartesian, SCARA, cylindrical, delta, polar and vertically articulated robots.
- **Learning robots:** These robots use AI techniques to acquire novel skills or adapt to the environment through some automated learning mechanisms. These may be used in various scenarios:
    - Self-learning robots take certain tasks and learn to perform these tasks very well after countless repetitions and training.
    - Robots learning from each other and cooperating with each other are examples of collective intelligence that exist in swarm of robots to perform certain tasks.
    - Robots learning from humans are those which observe a human expert performing certain tasks and then learn from the expert such as welding.

In its narrow sense industry 4.0 influences all businesses and the processes in business operations. In its broader sense when businesses live within a connected dynamical world, industry 4.0 influences not only businesses but also every aspect of our life. One of the biggest influences will be in the job market. As shown in Figure 3, most of the repetitive jobs will be handled by robots and only unique, one time tasks will be carried out by humans. Some of the examples of the jobs that robots will be helping humans do (quadrant II) include the job of lawyers and nurses. Some of the jobs that only humans will do seem to include art work and similar creative jobs [5].

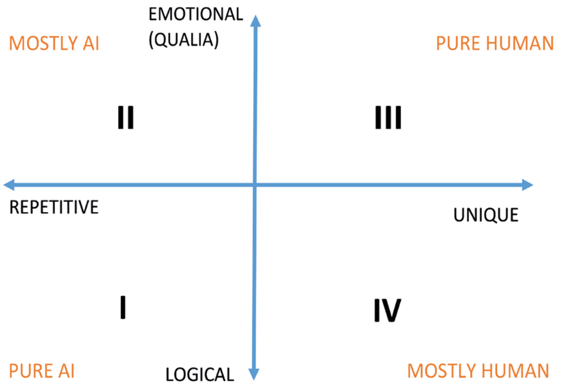


Figure 4. The future of jobs after automation [5]

While the technology evolves to allow faster and high capacity data transfer over network, there will be more and more independently operating small intelligent devices with their own wireless and voice-activated operating systems. Therefore, industry 4.0 and automation will not only improve on business operations in organisations, it will also have impact on the dynamical connected world of billions of intelligent devices all around us.

1.6.5 Section Summary

In this section, an introduction to industry 4.0 and automation is presented through visiting core developments in the digital world leading up to revolutionary ways of doing business. The roots of this revolution, called industry 4.0 is based on other influential recent developments namely big data, the cloud, IoT, Robotics and automation. Without any of these developments, industry 4.0 would not be possible. The big data provides an invaluable input to the ways businesses can be improved and be more innovative. The cloud provides all necessary infrastructure and services to work on data and acquire valuable insights for businesses. Based on these, automation of processes made possible via IoT and robotics, not only influence any businesses but also their environment. Industry 4.0 is a name given to a revolution that aims not only to improving organisations but also to the contribution to the creation of dynamic environments for businesses and customers.

All of the infrastructure and auxiliary new technologies play an instrumental role in the successes of AI not only in the business world but also in shaping our competencies and life styles leading to new forms of social interactions in media environments. In the next section that follow, what is so special with AI which can make this technology so influential in our life is presented by explaining its inherent strengths and mechanisms.

1.7 What Is So Special with AI?

For the first time in the human history we have a powerful tool that can aim at finding solutions to complex problems that no one or any particular group of human beings could imagine or could have thought. Basically, given a complex number of parameters of problem, AI may surpass human solutions in many domains.

Yet, we are just using AI only in a limited way. Current AI successes are highly dependent on 'the big data' coupled with the availability of high-speed processors and high-capacity internet connections. We are yet to witness the wonders of 'strong AI' going from applications to the areas such as vision and perception; natural language processing; many forms of machine learning such as analogy, collaboration and imitation based; strong reasoning and planning leading to common sense and, perhaps consciousness. Given all these, perhaps only half-explored areas and limited systems that are in action at present, AI will always have such capabilities and abilities that can transform our lives in many unimaginable ways now and in the future.

We owe these unprecedented successes of AI systems to a few distinct features, two of which are:

- AI has certain highly capable tools and mechanisms, such as machine learning, and
- the performances of AI mechanisms are amplified by the availability of large data sets, high speed of processors and capacity of connected networks, including the Internet.

Most widely used inherent mechanisms of AI are related to machine (deep) learning and these mechanisms make it possible to:

- classify (measure relevance or relationships),
- predict (making assertion about what is next or what will happen in future), and
- prioritise or optimize, especially through evolutionary AI methods such as genetic algorithms.

These are results of inherent functionalities of, at present, widely used AI systems (such as deep learning) to classify (measure relevance or relationships) and/or to predict (making assertion about the next or future). One other critical, though less pronounced, ability of AI systems is optimization especially through evolutionary AI methods such as genetic algorithms. Such strengths mostly come from the techniques of machine learning, whether it be reinforcement, supervised or unsupervised learning, using large data sets—verbal, textual, image or video streams. Perhaps most importantly, some of AI systems may be working in real time.

Any system which can provide information about the relationships around us; can tell us what the future will be like and what is the best action or sequences of actions to take will naturally be an influential system.

The influences of AI systems, with their inherent capabilities, on our present lives can be attributed to certain critical tasks that are complex and often are beyond capabilities of human-based resources, statistical techniques or simple algorithmic automation, digital or otherwise.

1.7.1 The Two Super Powers of AI

The above introduction is hardly the half of the story why AI is so special. The real success of current AI systems is mostly attributable to the fact that these systems are working like ‘a data hungry monster’ on the one hand and then like ‘a gentle tailor’ on the other. These refer to two different powers of AI systems:

- The Power of Pull—the hungry monster, and
- The Power of Push—the gentle tailor.

These two powers, exercised mostly on big data, at present make up the sweet looking specialties of AI making our lives sometimes better but at other times worse when we think of AI systems used in digital environments.

- **The Power of Pull:** AI systems can be used to perform a number of intelligent discovery tasks, including but not limited to:

  - collecting data from various sources such as life events, organizations, happenings, movements, actions and behaviours of people, the uses of machines and objects, social media, web sites or open sources,
  - collecting data in various forms such as text (regardless of language), voice, image or video streams,
  - determining unknown insights, patterns or relationship in, often, complex and multi-dimensional, data,
  - refining or simplifying data into useful forms, and
  - determining the optimum or useful sequences of actions.

In this respect, AI can be used as an effective search mechanism to reach out to needed and desired data and information sources and create useful sources as big data.

- **The Power of Push:** Similar to pull power, with the help of inherent functionalities, AI systems are able to produce and deliver intelligently tailored solutions or actions:
  - with appropriate type and amount of data or content,
  - at a specified, often, sensitive time,
  - for carefully determined target individual(s) and/or groups,
  - through suitable channels,
  - at specified locations, and
  - triggered by pre-set reasons.

In this situation AI is used to disseminate carefully and intelligently designed content, perhaps mostly from sources created via ‘pull power’, according to the needs and desires of users at appropriate times or intervals.

Over the past decade or so the world has witnessed all positive promises and actual impact of AI systems as a result of exercise of these two distinct powers in many application domains from industry to the health sector, from commercial solutions to individual solutions. The existing evidences have also built a noticeable confidence in AI. But the implications of these powers could be positive or negative. For example:

  - in the not so far future of education, AI powers could help content producers to group and classify topics and information in a precise and modular way in order to deliver content to learners according to their needs and requirements at a time suitable for them. Similarly, students will be able to access content as much as they need at any time. Plus, AI could guide them through what is actually needed for them, and

- similarly, the same powers can be used in the same precision for surveillance and control of masses and individuals. A government or an influential organization could gain more power over people through effective censoring using the 'hungry monster', while an activist could share exactly the right info for the right person(s) tailored according to a pre-designed protest or campaign.

More of such examples can be listed where the 'pull and push' powers of AI is exercised to help members of the society with their interactions in the digital world but also to make use of people's data for purposes which may negatively influence the quality of lives.

1.8 Co-Evolution of AI with Society

It is now well understood that the evolution and progress of the technology is moving quickly into a stage where neither our planet nor the members of the society can easily adapt. AI is one the most critical part of such changes in technology.

The speed of developments and the expectations from AI systems have an uncertain and blurred trajectory to the future. As we are not quite sure how, when and by whom those powers of push and pull can be exercised, there is a growing vagueness around potential benefits and harms of AI systems. There are issues about governance AI systems and their use globally. There are increasing numbers of implementations and applications (i.e. products) and many sectors and industries, stakeholders, standards, plans, actions and proposals in the complex world of AI systems in various locations of the world. We are not very sure how this fast changing complexity will evolve into the relatively slow changing future societies. The regulatory organizations are slow and also lack the required resources to effectively regulate AI developments globally.

Recognizing the urgency and significance, many organizations - large and small, for profit and not for profit - NGOs and governments are trying to gain control of almost free flowing AI developments around us. Most of these efforts may seem to be driven by the potential benefits of AI but they also carry important concerns about the potential harms of AI systems. It is now mostly clear that we are dealing with a potentially very unsafe technology unlike any other seen in the history of humankind. The risks of AI systems are minimally attributable to the way in which the key AI mechanisms such as neural networks or their variants and conventional machine (deep) learning functions. The degree of advances in these tools and mechanisms have really been minimal since the beginning of AI or their inventions. What is worrying is the fact that AI tools and mechanism can now produce unexpectedly successful and, sometimes, beyond human solutions owing to the availability of huge amounts of data, capacity and speed of the Internet connection and the speed of processors. This has been leading to great advances in the ways that data is gathered, designed and fed to AI systems for various purposes. These accumulating concerns and worries are more about the way the 'push and pull' powers of AI are exercised in practice rather than AI's inherent capabilities of models and mechanisms.

Is this new technology so strong that it can influence and determine human destiny? Are there any possibilities that we will face a form of intelligence misaligned with societies, communities and, in general, with our life on earth in terms of purpose and consequences? The answers to these questions are directly related to how the AI will advance into the future. It is very difficult to make predictions about the possibility of artificial general intelligence, let alone super intelligence or singularity. What is, however, clearly evident is that we will be living with AI systems, processes and applications which will be widely spread and available as a result of push and pull powers of AI.

AI will have even more significance when machines and humans converge. For example, robots converge to be more like humans and humans will be more like robots. The first signs of these are emerging through Neuralink [9] project where brain interfaces are being built with the aim of implanting tiny electrodes and chips into our brain and reading our brain signals. Also important are nano-bots swimming in our blood [4] and collecting information about chemicals and hormones within our body [10].

It is very clear that technology is not progressing in parallel to the progress in the geographically and technologically disperse global society. Therefore, in order to avoid possible sufferings and harm for human beings and to rather promote positive contributions of AI, all expected risks of AI should be managed with appropriate safeguards.

1.8.1 The Risks of AI

Potential risks of AI can be systematically grouped into four categories:

- **Regulatory Risks** arising from non-compliance or lack of regulations.
- **Technical Risks** that are part of the process of development from data gathering to design, development and implementation.
- **Consequential Risks** are those faced, perhaps, long after using the AI systems, most of which could be unintended.
- **Unknown Risks** will always be there as we are dealing with a system, which seems to have very unpredictable future. AI systems coupled with mind-blowing progress of relevant technologies such as the Internet, IoT and the cloud may evolve to something too complex to monitor and control.

Descriptions of each of these risks are beyond the scope of discussion here but it is worth mentioning some generic issues about AI that influences media and information literacy and human rights.

1.8.2 Barriers to Beneficial AI

The risks of AI can be summarized into the three major issues that prevent AI to be globally beneficial tool:

- **Capacity of AI—**due to a number of successful solutions to some of the key problems, such as playing chess or the game of Go, after many decades of so-called ‘winter’, AI has become over-hyped. This has created an overconfidence in AI but in fact the essential AI techniques has not changed drastically since its birth. AI is still strong only in the ‘narrow domains’. What has changed is drastic improvements in supporting technologies of AI such as speed and capacity of processors, connectivity and big data handling. Especially, the power of collecting, storing and processing various forms of data (i.e. text, voice, image and video) for AI systems making AI systems appear as if they are being successful outside the limits of ‘narrow AI’ and moving from domain-specific solutions to general AI. What, in fact, is happening is that the capability of having large data sets (i.e. instances of problems in millions or more) is strengthening AI to solve much more and a greater variety of domain-specific problems. Although there are a number of efforts in creating ‘general AI’ systems, the majority of successful AI solutions are still in the realm of ‘narrow AI’.

So, the intelligence in AI is still limited to the solution of domain specific problems. With this limited capacity and only within this limited capacity, AI has been functioning extremely well and has been very instrumental in all aspects of our lives.

- **Ownership of AI—**it is critical to understand that whoever dominates and controls AI systems will largely determine the kind and volume of generation of results and degree and boundaries of benefits shared. Currently, it is very clear that ownership of AI systems are in the hands of technologically advanced countries and those incorporations who hold the 'big data'. This is inevitably leading to various biases and also demographical and geographical inequalities, which seem to be getting worse over time. This raises an important question. Who are we developing the AI systems for? Are we developing AI for equal access to all in the planet regardless of the geographical locations and demographics of the people? So far, the implementations and the uses we witness do not appear to be so.
- **Memory of AI—**the memory of AI systems is hidden in the accumulated data that has been used in the past couple decades in generating solutions across variety of domains. Almost all of AI systems that exist now largely depend on the data prepared and presented to it. Whatever data you have will determine the nature and the success of solutions of AI systems. Therefore, how the big data is gathered, accumulated and used significantly determine how AI systems are serving the businesses, the society and individuals. It is now a well-known issue that the accumulated data sets that are employed by AI systems are largely determined by the present owners of AI systems and controllers of relevant IT infrastructure. This is especially evident in how the influential search engines and social media companies treat data.

Possibilities for beneficial AI: In order to have AI systems that are beneficial for all, it is absolutely necessary to create a context in which the ownership and memory of AI are not biased in any form. In this context, both are regulated with strong policy backing internationally and nationally to serve the common good. Although the capacity of the AI is limited with domain-specific problems, being able to collect and process large amounts of data for a problem is expanding the capabilities of AI in solving more and more difficult problems. Thus, by removing, or at least minimizing, the limitations or constraints of capacity, ownership and memory from AI systems, it is possible to solve complex problems with almost neutral and unbiased AI systems to serve much wider communities according to their actual needs and desires.

1.8.3 Beneficial AI and MIL

This book closely follows the efforts of UNESCO in developing MIL so that it empowers global communities and individuals to realize and exercise their human rights in accessing and utilizing reliable and correct information. The chronological development of MIL may be seen as having the following sequence:

- **Passive MIL—**at the beginning all MIL efforts were aimed at promoting possibilities to accessing, using and adopting media and information. This is just a one-way and therefore passive, flow of information from producers of content to the users. MIL would be seen as successful if individuals and communities had the competencies in using IT tools, including the Internet and accessing desired information.
- **Active MIL—**the next stage of efforts in developing MIL involved creating, disseminating, analysing, evaluating, interacting with and influencing media and information. Especially with digital transformations in the media ecosystem and education, MIL is becoming increasingly instrumental in enabling communities and individuals to participate in the creation and dissemination of information and useful content as well as governance of information and relevant MIL competency development. In this way, MIL efforts create more active and dynamical media and information environments.
- **Influential MIL—**the ultimate aim of the MIL efforts may be achieved only if competencies of communities and individuals have been developed in such a way that they can realise and practice media and information rights while effectively interacting in the media environments. In this context, it is almost possible for everyone to access the information more or less as needed and desired and information is mostly reliable and clear from manipulations or fakes. The whole ecosystem of media and information works as properly as possible. This state of MIL may only be possible if accessing reliable and useful information is secured as part of the human rights. Therefore, all present and future MIL efforts should be channelized fully to make objectives of MIL accepted as part of human rights by influencing relevant stakeholders including the private sector and large corporations. In this way, influential MIL will provide a piece of mind to individuals and communities when using information to improve their lives.

Given these various efforts in MIL, how can AI support active and influential MIL? The first and foremost important step in utilizing AI to support MIL is aiming to reduce, if not remove, issues related to ownership and memory challenges associated with AI. This is an extremely challenging task given the strength of those large data-holding technological giants, some of which are stronger than governments. Supposing that the general AI challenges related to ownership and memory of AI are resolved to a reasonable extent at least within a limited scope in certain areas of MIL, there are a number of ways AI could help with influential MIL and human rights. Or as AI itself is neutral technology, it may be possible that, in good hands, AI can help drastically.

It is worth remembering that AI systems are best known for their speed and precision in providing solutions in connected environments. AI may sometimes be:

- a data-hungry monster handling huge amounts of data from various sources in media environments and for individuals, professionals (i.e. journalists) and communities, and
- also a gentle tailor for moderating and disseminating content through fast and precise personalisation and customisation according to needs and desires of actors in MIL including ordinary citizens.

Although it will be elaborated extensively throughout this book, AI can play various significant roles in contributing to influential MIL efforts and human rights, for example, by:

- gathering and disseminating precisely customised reliable information for all actors in MIL according to their needs and desires,
- designing and delivering MIL learning environments for empowering individuals and improving MIL competencies,
- providing new and effective means for creating, accessing and using information,
- providing labels or descriptions with respect to sources and measures of degree of reliability of those sources,
- monitoring and filtering all forms of fakes, disinformation, propaganda and manipulations to ensure quality experiences for all actors in media environments,
- providing tools to facilitate evaluation of potential benefits and harms of these experience, and

- creating systems that can help detect and reduce potential surveillance, all forms of biases, discrimination and improve participation, especially in, democracy.

Most of these supports that can be received from AI systems for influential MIL are related to improving human rights and supporting individuals and communities to realize and exercise their rights, including freedom of speech.

Using AI for influential MIL can also support regulators. Currently regulators have at least three main issues when it comes to regulating AI in general and also in particular regulating AI for the purposes of improving MIL experiences. These challenges are:

- lack of sufficient AI expertise and resources,
- lack of speed to keep up with changes in AI, and
- dominating powers of those data-holding tech corporations.

Using AI for MIL and improving human rights may help to create necessary grounds for the regulators to understand and respond to 'AI for MIL' objectives and implementations through various supporting policies.

1.9 Conclusions and Principles of AI for MIL

It is important to realize that AI techniques have improved only marginally since its birth in the 1950s and 1960s. What has been truly revolutionary is all about advances in power of processors, speed and capacity of data gathering, storing and transferring in networked environments. Plus, there has been significant improvements in how various AI relevant technologies such as mobile networks and devices, IoT, the cloud and big data handling have advanced and taken place in our business, social and individual lives. Given such advances AI is now performing much better than what is actually expected from AI. One of the major reason for this is that provision of big data in various forms including text, voice, image and video, is extending capabilities of AI in solving complex problems for which large amounts of data can be collected. Until recently it was extremely difficult for AI to process image data but now intelligent cameras can do face recognition and even sentiment analysis from facial expressions. While the AI techniques are still humble, the availability of data makes AI systems capable of solving complex problems.

Understanding AI with its close dependency on availability of data reveals a realistic view of capacities and capabilities of AI for now and for the future. As for now, it is very reasonable to say that AI can solve complex problems that comes with making MIL successful and widely available to all. But it also has implications for intended and unintended misuse that may create challenges for MIL as well as human rights.

As discussed, AI sometimes function as ‘data-hungry monster’ which can make collection and processing large and complex data not only possible but also extremely useful with ‘intelligence’ injected in the processes. AI also work as a ‘gentle tailor’ where a collected and processed data can be disseminated in an ‘intelligently’ customized and targeted way.

These ‘push’ and ‘pull’ powers of AI can be extremely useful for MIL in education for competency building as well as accessing relevant and reliable information in media environments. Equally, they may create challenges for MIL, if in the wrong hands, in terms of, for example, bias, censorship, disinformation and fakes. These will have severe influence on how human rights may be protected and sustained. This in turn, will open a new chapter in making MIL to move from passive to active and then to an ‘influential’ one. In the chapters that follow, detailed discussion of these challenges will be provided.

Minimizing the negative influences of AI on MIL and HR (i.e. to achieve a positively ‘influential MIL’), is a significant challenge. This challenge is primarily related to directly opposing some of the activities of large technology corporations which hold the majority of data in the world and control the progress of AI technologies. It is therefore important for MIL professionals and relevant stakeholders to make collective and coordinated strategies and action plans in resolving issues relevant to, especially, ‘ownership’ and ‘memory’ of AI. The ownership of AI technologies has to be as equally and equitably distributed as possible and the memory (i.e. globally accumulated data) that is available to AI should represent all and be available to all, globally. In the chapters that follow, possible ways of policy, regulatory and legal responses to these are presented.

References

1. Atos (n.d.) *IaaS, PaaS, SaaS (Explained and Compared)*. Retrieved from https://apprenda.com/library/paas/iaas-paas-saas-explained-compared

2. Crnjac, M., Veža, I., & Banduka, N. (2017). From Concept to the Introduction of Industry 4.0. *International Journal of Industrial Engineering and Management,* 8(1), 21-30.

3. German Cancer Research Center (Deutsches Krebsforschungszentrum, DKFZ) (2019). *Artificial Intelligence Helps to Better Assess Treatment Response*

of Brain Tumors. *Sciencedaily*. Retrieved from www.sciencedaily.com/releases/2019/04/190402215624.htm

4. Kurzweil, R., & Miles, K. (2015). Nanobots In Our Brains Will Make Us Godlike. *New Perspectives Quarterly*, *32*(4), 24-29.

5. Kushchu, I. (2020). *AI Training Notes,* TheNextMinds, (can be requested via https:// thenextminds.com/get-in-touch/).

6. Marr, B. (2015, March 19) Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance. *IBM Bigdata & Analytics Hub.* Retrieved from https://www.ibmbigdatahub.com/blog/why-only-one-5-vs-big-data-really-matters

7. MIT (n.d.) *MIT Oxygen Project.* Retrieved from http://oxygen.csail.mit.edu

8. Mueller, J. P., & Massaron, L. (2021). *Machine learning for dummies*. John Wiley & Sons.

9. Neuralink (n.d.) *The Neuralink project.* Retrieved from https://neuralink.com/

10. Thiruchelvi, R., Sikdar, E., Das, A., & Rajakumari, K. (2020). Nanobots in Today's World. *Research Journal of Pharmacy and Technology*, *13*(4), 2031-2037. doi: 10.5958/0974-360X.2020.00366.2.

11. Torres, L. (2019). *Improved Prosthetic Hand Has a Lighter Touch and Easy Grip*. Retrieved from https://www.npr.org/sections/health-shots/2019/07/24/744601440/improved-prosthetic-hand-gains-a-lighter-touch-and-easy-grip?t=1564244241417

Key Readings

- ATCC Finance. (2015). *Industry 4.0 Challenges and Solutions for the Digital Transformation and Use of Exponential Technologies.* [Audit Tax Consulting Corporate Finance Report]: Zurich, Switzerland.

- Daugherty, P., & Carrel-Billiard, M. (2019). *The Post-digital Era is Upon Us-Are You Ready for What's Next.* (Technical Report). Retrieved from Accenture: https://www.accenture.com/_acnmedia/PDF-94/Accenture- TechVision-2019-Tech-Trends-Report.pdf

- Frankish, K., & Ramsey, W. M. (Eds.). (2014). *The Cambridge Handbook of Artificial Intelligence*. Cambridge University Press. Retrieved from http://assets.cambridge.org/97805218/71426/frontmatter/9780521871426_frontmatter.pdf

- Rojko, A. (2017). Industry 4.0 Concept: Background and Overview. *International Journal Of Interactive Mobile Technologies (IJIM), 11*(5), pp. 77-90. doi:http://dx.doi.org/10.3991/ijim.v11i5.7072
- Yáñez, F. (2017). *The Goal Is Industry 4.0. Technologies and Trends of the Fourth Industrial Revolution.* Independently Published. ISBN 9781973413172.

Key Videos

- AJ+ (2018, January 24). *Robots And AI: The Future Is Automated And Every Job Is At Risk* [Automation, Pt.1] [Video file]. Retrieved from https://www.youtube.com/watch?v=rnBAdnNIIXk
- Board of Engineers (2018, March 7). *The 9 Pillars of Industry 4.0* [Video file]. Retrieved from https://www.youtube.com/watch?v=iW9M7YywOLA
- Growth Tribe (2018, March 28). *AI Impact on Jobs & the Skills of the Future* [Video file]. Retrieved from https://www.youtube.com/watch?v=B5l_vNEcFWg
- Modelec (2019, July 18). *Trends and Challenges in Smart Factory* [Video file]. Retrieved from https://www.youtube.com/watch?v=rnBAdnNIIXk
- New Scientist (2017, July 24). *Tiny Robot Could Swim Through Your Blood* [Video file]. Retrieved from https://www.youtube.com/watch?v=cOEFNVdhZDA
- Ramesh R. (2017, August 13). *Artificial Intelligence In 5 Minutes* [Video file]. Retrieved from https://www.youtube.com/watch?v=2ePf9rue1Ao
- Ramesh R. (2017, July 29). *How to Apply AI in Business* [Video file]. Retrieved from https://www.youtube.com/watch?v=N_eHmaRf9T4
- Ramesh R. (2018, September 20). *So you want AI for your business. Where do you start?* [Video file]. Retrieved from https://www.youtube.com/watch?v=dfvycjYIMPo
- Simplilearn (2019, April 30). *Artificial Intelligence In 5 Minutes What Is Artificial Intelligence? AI Explained* [Video file]. Retrieved from https://www.youtube.com/watch?v=ad79nYk2keg
- TEDx Talks (2018, August 30). Jennifer Brooks. *Making a Social Impact in Fourth Industrial Revolution* [Video file]. Retrieved from https://www.youtube.com/watch?v=cp-ricDKPPA
- The Artificial Intelligence Channel (2017, December 17). Kai-Fu Lee. *Artificial Intelligence and Business* [Video file]. Retrieved from https://www.youtube.com/watch?v=OojaXyvksFc

2. Media and Information Literacy & Artificial Intelligence

Tatiana Murovana

Media and information literacy aims to empower professionals and individuals to improve lives through access to reliable information and its correct use, while digital technologies, in particular AI, can enable the change of the MIL role from passive one to active and even influential role. Thus, for media literate people, media environment might become more accessible and content creation and dissemination more feasible and efficient. Most of these opportunities are related to developments in the ability to collect, store, mine and analyse the big data, and to discover insights from this data that can be used in many ways. AI comes as a significantly appropriate tool for the discovery of data insights due to its ability to transform large volumes of data into meaningful patterns and rules. AI-based technologies have changed media environment substantially. Those factors, along with growing monopolisation of digital platforms, has given rise to new challenges for MIL.

For many years UNESCO invested efforts into developing MIL to have an active role in improving the lives of many in various geographical locations. The sections that follow starts with the description of perspectives and activities of UNESCO on MIL. In order to form a basis for further discussion on the digital technologies and AI interaction with MIL, an evaluation of MIL in the modern socio-economic context and related changes on the perspectives of MIL is presented. Then, the pros and cons of digital technologies and AI in empowering citizens are detailed in the final part.

2.1 UNESCO Perspectives on MIL

Approaches to understanding of literacy has been gradually changing during the last few decades. Literacy is no longer seen as simple abilities to read, write, count and comprehend language. With this change, media and information literacy takes up a critical position to literacy, too. It does value in-class as well as out-of-class learning via various platforms relevant to information, media and technology, which enables people to develop and practice a critical thinking about what is being learned.

UNESCO, for the last 40 years or so, has always been a leading organization for contributing and improving critical competencies of people in this respect. For example:

- Between 1930 and 1955 'better broadcasting' principle led to the term media literacy.
- In 1974, the term 'information literacy' was coined by Paul Zurkoskwi.
- In 1982, Grunwald Declaration stated that "political and educational systems need to recognize their obligations to promote in their citizens a critical understanding of the phenomena of communication".
- In 1997, Paul Gilster introduced 'digital literacy' to refer to people's understanding and use of information through various digital sources.
- In 2008, UNESCO proposed the umbrella term 'media and information literacy' to embrace all the developments above and to refer to information, media and technological competencies.

UNESCO's strategy in this process has unified the study competencies under the concept of MIL which now involves a combined literacy of information and media including digital or technological ones. UNESCO promotes significance of people's competencies for work or social life in terms of knowledge, skills and attitude towards media and information. UNESCO's MIL aims to improve people's understanding of:

- creation, dissemination and use of information,
- sources and targets of information such as libraries, digital media platforms,
- purpose, intention and capabilities in dissemination, adoption and use of information, and

- appropriate management of information for improving work or personal lives.

These UNESCO MIL notions are clearly stated in the 'Belgrade Recommendations on Draft Global Standards for Media and Information Literacy Curricula Guidelines':

"The use of the term 'media ' … refers to two dimensions. Firstly, there is the news media as an institution, the 'fourth estate ', having specific professional functions that its constituents pledge to fulfil in democratic societies and which are necessary for good governance and development. This includes radio, television and newspapers, whether online or offline, as well as includes journalistic content on the Internet. Secondly, there is media as the plural of the term 'medium ' and which here refers to multiple communication modes such as broadcast and cable television, radio, newspapers, motion pictures, video games, books, magazines, certain uses of the Internet etc. MIL encompasses engagement with all these modes. For its part, UNESCO is particularly concerned with information and news, but recognizes that there is other content such as entertainment, interpersonal communications and advertising."

UNESCO's MIL activities are supported by a strong strategy to enable societies all around the world to reach as much high level of media and information literacy level as possible. Various countries have various levels of MIL. MIL policies and strategies globally may be categorized as follows:

- advanced in many countries in Europe, North America and Australia,
- fair with no systematic implementation, and
- none or very close to none.

Most countries are at the basic level where they provide IT/digital/computer literacy, which is still an important foundation for much broader concept of MIL. UNESCO is working consistently with various initiatives to improve MIL globally. Some of these initiatives include:

- preparation of model 'Media and Information Literacy Curriculum for Teachers',
- the facilitation of international cooperation,
- development of 'Guidelines for Preparing National MIL Policies and Strategies',
- articulation of a 'Global Framework on MIL Indicators',

- setting up a 'MIL University Network',
- articulation and establishment of an 'International Clearing-house on MIL' in cooperation with the United Nations Alliance of Civilizations, and
- provision of 'Guidelines for Broadcasters on Promoting User-Generated Content and MIL'.

UNESCO also runs various training and development programs for improving competencies:

- for teachers two major programs are 'Media and Information Literacy and Inter-cultural Dialogue' and 'Media and Information Literacy in Journalism' to:
  - create a high-level awareness of the significance of MIL in the educational processes,
  - empower teachers in integrating MIL into teaching via provisions of pedagogical methods, curricula and other resources, and
  - in particular, develop massive open online courses (MOOCs) on fake news, intended and unintended information anomalies disinformation, misinformation and mal-information.
- for capacity-building programs and resources:
  - curricula development,
  - policy guidelines and articulation and assessment frameworks,
  - free and open online courses for self-paced MIL.
- networking and research facilitation through:
  - Global Alliance for Partnerships on MIL (GAPMIL),
  - MIL University Network,
  - MIL CLICKS as a social media initiative.

All of these initiatives are long-term commitments stated in the UNESCO strategy in order to promote media and information literacy worldwide.

Being an integral part of the UNESCO Education Sector, UNESCO Institute for Information Technologies in Education (UNESCO IITE) focuses its activities in the field of MIL on:

- raising awareness and policy advocacy on the significance, role and scale of MIL for education,
- contributing to the development of MIL-related policies and professional strategies at international, regional and national levels,
- providing MIL-related training for educators, university and school librarians and developing relevant education materials and tools, and
- facilitating cross-sectoral and interdisciplinary collaboration among stakeholders.

UNESCO's efforts in fostering MIL serve the intended purposes for all may not be successful without a proper account of the influence of new digital technologies and AI on how media environments and their use change access to information.

2.2 Media and Information Literacy and Socio-Economic Context

As presented in the previous section, according to UNESCO, MIL is largely concerned with:

- Processes related to information: how information is produced, disseminated, by whom and for what purposes.
- Ability and desire to use information and media: how and why people use or don't use information, how they engage with libraries, media and technology, or if not, why?
- Critical capacity to evaluate information and media: what knowledge, skills and attitude do people need to critically evaluate information?
- Successful interaction with information and media: how can people manage their interaction with information, media and technology for desired outcomes in their social, political, economic and cultural lives?

Thus, MIL includes all technical, cognitive, social, civic and creative capacities that allow people to access, critically understand and evaluate the media and consciously and effectively interact with it. All of these actions unfold in specific socio-cultural and economic contexts that cannot be ignored. As the socio-economic features of our time, these contexts, in turn, are largely determined by several multifaceted and ambiguous processes that can be summarized under the term 'digital transformation' in the world. The term in itself is somewhat populist and not rigorous enough, but its use allows us to highlight at the operational level some important MIL trends and characteristics from the viewpoint of current reality. We suggest to note the following aspects of the digital transformation:

- complexity and uncertainty,
- changes in the media environment,
- post-truth era, and
- ‘Attention Economy’ and ‘Surveillance Capitalism’.

It is very clear, that in the last decade, the digital transformation of the world is mostly driven by big data, the cloud, IoT and AI. This is also valid for every aspect of private and professional life, including media environment and people’s interactions with media and by media. For this reason, the following discussions on the above aspects of digital transformation should be taken in the context of new technologies and AI.

2.2.1 Complexity and Uncertainty

Formal education system for many years had been oriented towards shaping intelligent and educated people, which meant the ones who had right answers. This would have remained true, had we not been already living in the fast-changing unpredictable world where the volume of new knowledge and information, as well as the number of innovations is growing exponentially. Today, there are no single right answer and viewpoint, one comprehensive explanation of what is happening that would convince everyone. The room for straightforward answers is steadily and constantly narrowing. It is not easy to find what we need in the huge information flow, in a wide range of opinions, views and approaches. No less difficult is to define how correct and effective the chosen position and preferred option will be.

We live in a fast changing world that it is difficult for even partial comprehension and almost impossible to fully understand and grasp (of course, we are not talking about people who would make their misunderstanding of the world less acute coming forward with some nice all-explaining conspiracy theory or religious dogma, as well as about those in a state of mental distress). The unexpected, the unplanned, the unpredictable is becoming the norm. Nassim Taleb’s ‘black swans‘ [6] have become an integral part of our culture. Almost every day we learn that something ‘unprecedented‘ has happened. The unprecedented has become commonplace to the degree that the word itself has already lost its old-time emotional load and would not transmit an energy impulse any more.

There is a growing gap between the complexity of the phenomena and processes we face, the complexity of various social or economic systems and our limited ability to cope with them, both at the level of the individual and the society level. Our reality has become hybrid; it is practically impossible to distinguish 'natural' and 'virtual' realities. These spaces of our existence are so intertwined and interlinked that, in essence, they are no longer separate from each other. But we have not yet developed a coherent toolkit, methods, strategies and practices for managing that hybrid reality. In the legislative, psychological, social or some other context, we still use old, traditional approaches and paradigms, slightly adapted to the 'digital' by a way of some decoration.

All of these issues related to 'complexity' and 'uncertainty' are mainly driven by the advances in digital technologies but most effectively by the AI and related new technologies. For example, while speed and variety of information in the media environment are largely the result of capabilities of digital technologies, exposure to the 'unexpected' may be the result of intelligent analysis and processes of data by AI and resulting intelligent delivery. The core lesson to take from all of these is that AI is one of the major contributors to uncertainties and complexities around MIL and to manage these more proactive approaches and strategies are required.

2.2.2 Changes in The Media Environment

The same situation also applies to the media sphere. Digital technologies have completely changed the way we create, deliver and consume or use content. Quite a fair amount of work have been written on this topic, but since our goal is not a rigorous study of the media, but rather a general understanding of the modern context for MIL, we will note only some fundamental changes:

- Data explosion: the amount of created and circulated information is growing exponentially. Every day an unimaginable amount of content and data is produced.

- Source and verification ambiguity: publicly accessible content is no longer created and delivered by a limited number of professional journalists and authors, publishers and television and radio companies. Content may now be created by individuals but are collected and regenerated mostly by large data-holding companies. Procedures of thorough selection and verification of information, editing and proofreading are now becoming a history.

- Diminishing power of regulatory authorities: although many countries declared themselves free of censorship, government authorities used to have a wide arsenal of means to restrict the dissemination of unwanted materials. Now the national authorities of most countries are losing control over information flow and the role of actual content regulators is being transferred to the owners of the largest digital platforms.

- Wide geographical spread: printed texts were previously distributed in a limited number of copies and usually within one country or region, one culture and language. Now, information can be spread worldwide with no limitations.

- Ambiguity in traceability: previously, names of the media product creators and distributors, as well as clear lines of their responsibility were known to everyone.

- Distributed collections of data: in the analogue era, libraries, archives and museums maintained and preserved the world's memory. Still nobody knows how to deal with a tremendous and endless amount of digital information.

- Media convergence: established industry services and work practices transform and entirely new forms of content emerge. Long-established media industry and content repository eroded. Increasing uncoupling of content from particular devices presents challenges for public policy and regulation.

- Proliferation of devices and platforms: with the rise of smart-phones, video, social media and live broadcasts merge transforming and creating novel and variety of ways to consume and create media.

- Directed and customized content: AI systems make digital content and services more customized and personalized. This leads to the end of the 'one size fits all' approach of conventional TV, radio, newspapers or magazines and of the mass dissemination of the same information and the same advertising to large audiences.

All of these drastic changes are emphasized and augmented by the use of AI and its supporting technologies. Any of the above concerns either using AI as a supportive tool or AI is the core technology giving rise to the emergence of these fundamental changes around MIL.

2.2.3 Post-Truth Era

The situation where objective facts are less important for shaping public opinion than appealing to emotions and personal beliefs has been called 'post-truth'. In 2016, this concept was recognized by the Oxford Dictionary as the 'word of the year' due to its active use in the media when describing the process and results of Brexit in the UK and the US presidential elections. Then it became obvious that the picture of reality, constructed by the media, has a decisive influence on how people make important decisions regardless of whether 'media reality' reflects actual reality at all and to which degree it is substantiated and confirmed by serious data. In the world of post-truth, it has become challenging to rely on objectivity and reliability, which seem less attractive upon being lost in the information noise. Assessments, opinions, comments, rumours, tales, myths and legends come to the fore, which create a feeling of uncertainty and incompleteness.

Thanks to the Internet, we have gained access to almost limitless amounts of information and tools and space for creating and distributing one's own content has become commonplace. Instead of information shortages, the humankind is facing information overload and noise which is not easy to manage. People have no time to contemplate over every event, at most, they get to run through the headlines. On the Internet, you can find confirmation or refutation of almost any idea or position and most of the information space in which we live is an arena for manufacturing opinions, not the search for truth.

Although the term 'post-truth' itself is relatively young, in a certain sense it can be said that humanity has existed in a post-truth situation throughout all its recorded history. There have always been people who, for the sake of their goals, created myths, misleading others, interpreting events in their way and influencing decision-making. If you carefully analyse any significant historical event in which a large number of people participated (wars, revolutions, elections, referendums), it becomes clear that the masses of people are guided to one degree or another by emotions. Soldiers are easier to control if you provide them with an image of the enemy, rather than let them know whose interests they should risk their lives for, allowing them to wonder why the people they fight with are considered enemies. The voter will rather react to a beautiful slogan, rather than analysing the content of the actions proposed by the candidate and the consequences of their implementation. Nevertheless, over time, mass literacy, the spread of the scientific method of cognition, public libraries and professional media made it possible to gain access to reliable facts making informed conclusions from them. But these achievements of civilization have historically coincided with the spread of an attitude towards truth as something that cannot be explained unambiguously; truth has become multifaceted and relative. Each person has an opportunity to independently decide what values to accept, what to believe or not.

Of course, propaganda and disinformation are nothing new and the history of humankind is the full of myths and misconceptions. However, in the context of ubiquitous communication and AI-powered personalization, threats and risks of post-truth acquire new scale and character. Digital technologies naturally are very suitable to automatically disseminate the 'post-truth' cases and content to wide variety of receivers. AI, on the other hand, enhances and amplifies such cases using intelligent methods to change the character of truth significantly. A great example is fake videos, created using AI techniques, where it is almost impossible to verify the originality of such videos after they are being copied once. In these videos, content, voice and image could be played around rather flexibly in the way the creator prefers and are being increasingly close to the real videos. Perhaps these are the most recent and effective examples of how media realities are changing.

2.2.4 'Attention Economy' and 'Surveillance Capitalism'

At the dawn of humanity, food was the most important economic resource. Then the land has become such a resource, along with the people who cultivated it. Later, economic power passed into the hands of those who owned the means of industrial production—equipment, machinery and structures. Now the market is being captured by new industries that could not be predicted several decades ago: in 2019, the online retailer Amazon overtook Microsoft in terms of market capitalization. Together with them, Apple, Alphabet (Google's parent company), Facebook and the two Chinese digital giants, Alibaba and Tencent [1], are among the world's top ten largest companies. Only one holding company from this list (Berkshire Hathaway Inc.) is not directly associated with the information industry. What is the basic resource of this new economy?

Companies that offer expensive high-tech services such as search engines, social networks, detailed maps and navigators, translators, instant messengers, office applications, etc. are becoming financially strong leaders. Yet, we use Google maps or WhatsApp messenger for free. So, how do these companies make money? They sell users' attention, extracting profit from the time that users spend while browsing and using the services, from the sale of advertisements and other information that users get to view and from the use and sale of data that users generate in the process of constant interaction with services, applications and gadgets. In this way, we pay for the services and opportunities that we get with our attention that is the primary resource and currency of the attention economy. One would have hoped that a certain balance could be achieved between the relevant high-quality content received and the amount of attention given. But something went wrong.

The resulting economic model of the Internet, where information and communication services are provided for free but a profit is derived from advertisements, user data, forming opinions, attitudes and beliefs, habits and behavior, has largely determined the information space in which we currently live. Since human attention is limited, there is a struggle for any portion of the consumer's attention. It is not the creation of high-quality information that becomes economically profitable. But rather, manufacturing content and the inception of such forms of activity and interaction that are designed to attract and retain audience and making them come back again results in more profit.

Viral videos and scandalous fake news achieve this goal more effectively than in-depth articles with detailed and professional analysis. Games and social networks with their calculated likes, comments, friends and other elements of the 'vanity fair' give incomparably more ad views. Further, they become a lifestyle much faster than mastering educational materials, studying the masterpieces of world culture, or reading the quality press. Emotionally charged content not only inevitably attracts attention, but also provokes people to distribute it themselves in such a way that they become carriers of, by and large, meaningless and useless information.

New technologies such as big data, machine learning and AI are changing the rules of the game in the MIL world. Now significantly widespread, every online activity produces data instantly and this data from various sources are collected and accumulated carefully. The ad-based business models are now changing to data-based information provisions. Data has become an important asset and AI with machine learning has become an invaluable tool for discovering useful information from the business data that is collected from various sources. Often on social media, the data that is given to tech businesses in return for certain conveniences is aggregated, analysed and then sold as is or after being refined based on some parameters.

The raw data is now creating essential foundation for understanding much about people. AI is taking this further and helping businesses to create predictive models of people's behaviours. So, the use of our data is not limited to the data collectors with monopolistic IT power but also is taken advantage by a variety of businesses including insurance, retail, hospitality, healthcare, finance, entertainment, education, logistics and transportation etc. This, now, is a new ecosystem of market actors—from producers to consumers.

With the modality of using new technologies, data owners and users are clearly differentiated. In the background, almost invisible, are powerful IT companies holding the data and the more visible are those producing data. Data sources and forms are continuously monitored by the 'invisible'. This set up has now become an extreme pressure, which is slowly leading to undesired surveillance and social control. This in turn may influence the quality of lives and freedom as well as the power relations and degree of unfavourable domination in the society – possibly harming democracy. This issue is extensively discussed in 'The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power', a book by Shoshana Zuboff (see the sections 'The Corporate Power and the Behavioral Surplus Economy' and 'Prediction Factories' of this book).

2.3 MIL-Related Challenges Posed by The AI

In the previous sections, some changes in the social, economic and communication environments that are relevant to MIL were reviewed. This section highlights the manifestation of these changes in different MIL-related contexts.

In 2018, the Pew Research Center published a paper on the impact of advances in AI on the essential elements of being a human [5]. The paper reports that experts interviewed during the research, expressed their concerns about the followings:

- technology is not neutral and it replicates and reinforces biases,
- people are losing control over their lives as decision-making in digital systems is automatically ceded to 'black box' tools,
- most AI is controlled by companies or governments whose focus is on profits and power, not on human-centred values and ethics, and
- people's deepening dependence on machine-driven networks will erode their abilities to think for themselves, take action independent of automated systems and interact effectively with others.

These observations are also relevant to understanding how people interact with and through AI-driven systems. Let's consider them with a focus on information and media.

2.3.1 Filter Bubbles and Echo Chambers

Humans often desire to prove the correctness of her or his own opinion. This desire becomes a basis for the emergence of confirmation bias—the tendency to seek, use and memorize the information that confirms our beliefs and not the one that refutes them. Under the influence of confirmation bias, new information is interpreted in such a way that it can reinforce a person's existing ideas about reality or a common truth regarding a particular issue. The information that obviously conflicts with these perceptions is ignored or discounted.

Confirmation bias evolved as a human evolutionary advantage facilitating rapid decision-making, effective communication, cooperation and socialization within tribes. We need this tendency to confirm our pre-existing attitudes, to verify self-image, to avoid cognitive dissonance and to boost social identity. However, being pushed and amplified by digital technology, confirmation bias enables the expansion of two social phenomena— 'filter bubbles' and 'echo chambers'.

Notable examples of the power of these phenomena include the rise in populist politics and mass protests against immigration in Europe, the polarizing election campaign of Donald Trump, the anti-vaccination movement or the vote for Brexit.

The heart of the description of a filter bubble comes down to providing content for consumption that is closely associated (most likely created by AI-based personalization systems) with personal preferences so that a bubble is formed to present a restricted view of the bigger picture.

Search engine results based on the gathered data on users' preferences, behavior, habits, etc. include only the pertinent information. Search results, news feeds, recommendations, services, products that appeared on our screens are personalized and do not reflect a wider picture. 'Personalized' means that they are tailored to be liked by users (a response to confirmation bias) and to get them to consume content, products or services (to spend their attention, time and money and to provide an additional portion of private data). The more data is collected from viewers' habits and the more tailored is the content they get, the stronger is filter bubble they have got in. This often is result of AI systems built for these purposes.

Personalization, with the help of AI, gives people an opportunity to consume products that seem to be designed just for them. This fits well to inner belief system. We do not like to waste time scrolling. These AI systems produce such tailor-made results that one does not see alternative concepts, viewpoints, or even services and products and exposure to information that challenges individual attitudes is minimized.

Confirmation bias and filter bubble create comfort zones while, at the same time preventing us from getting the whole story. They reduce the diversity of ideas exposed to a person and provide the bases for misinformation at an escalated scale. Deep fake technology strengthens misinformation and its persuasiveness.

People who tend to share a common filter bubble with like-minded friends, over time form communities in which content and communication confirm certain ideologies. Unanimity and consensus are echoed from all sides making the community particularly prone to becoming of radical groups and having polarization.

Strong overlapping of filter bubbles of interacting people is called the echo chamber. Increasingly radicalised online groups may at some point turn to real-life violence and terrorism to achieve their goals.

In the circles of communication that develop in this way, ideas and beliefs are reinforced by multiple repetitions. Like-minded people support the common wisdom of their respective circles without trying to question them. The echo chamber gets closed, so no alternative information may penetrate inside it. Everyone located outside the circle or the echo chamber, is perceived as outsider and is being treated accordingly. The unanimity reigning within such closed network communities, with regard to certain phenomena, convinces members that their views are correct. Alternative opinions or facts that contradict the accepted concepts are not simply muted by the community members, but are actively rejected. Having radical views within such closed communities is aggravated by the fact that, due to, mostly AI-based, personalization of social networks news feed, search results and recommender systems, information that can make participants doubt their viewpoints simply will not get into view.

2.3.2 AI and Decision-Making in Communication

Another area of the interaction of AI and MIL is the decision-making process. One of the important advantages of being media and information literate person is to have the capacities to make informed decisions. However, due to their predictive abilities, intelligent techniques are offering increasingly more ready-made decisions to people.

In the area of media and communication, we have equipped AI with great decision-making power, despite the understanding that AI systems function in ways that may not be predicted or explained even by their designers and makers. This is the AI who decides what we see in the search results and in the news or social media feeds. This is the AI who decides what kind of content to read or to watch and receive via the recommendations. AI decides what kind of products or services should be proposed to choose. AI produces content, creates art and disseminates it at a much larger scale than we could do on our own.

From the MIL perspective, it is important to understand how the medium of communication works. However, AI-powered algorithms work as a black box and sometimes the processes of making decisions in AI is hidden even from its creators.

Media and information literate approach to the analysis of communication implies deciphering of purposes of the other party. In cases where automated decision-making is used, we can be fantastically wrong about it.

Not only media businesses have adapted to a 'data-driven' and 'AI-driven' approach for operational and even strategical decision-making. Human resources, healthcare, insurance, finance, tourism, retail, entertainment, education, transportation, criminal justice and other industries and sectors bring AI to automate decision-making processes. The predictive ability of AI is highly in demand, despite its non-transparency and unaccountability. The situation is worrying, as in the case of human error, one can complain about a person who make a mistake. But who to blame in the case of software's mistake or bias?

Famous American writer Robert Shekley wrote in his short story *Watchbird*: "Admit that we were wrong trying to cure human problems by mechanical means. Start again. Use machines, yes, but not as judges and teachers and fathers". We are definitely not following that testament. AI is moving into our lives and decision-making processes slowly, silently and surely.

2.3.3 Programmed to Be Addictive

Andrew Sullivan underlined the power of AI in his article published in the New York Magazine that "No information technology ever had this depth of knowledge of its consumers – or greater capacity to tweak their synapses to keep them engaged" [6]. This consideration moves us to one more intersection of AI and MIL – the addictive potential of AI-enhanced products and services.

As we have learned from the attention economy piece of our conversation, the profit of digital media companies depends on the amount of time we spent online. The surest way to that is to make the product a part of our life. That is why persuasive design and other tricks to keep us hooked on platforms are so popular. Technology evokes the mob-like behavior in people by getting us into intrusive online habits.

Here are some statistics that illustrate our relationship with gadgets [3]:

> 44% of teens check their phone as soon as they wake up.
>
> 67% of teens check their phones every 15 minutes.
>
> 50% get anxious when they cannot.
>
> 68% of people suffer from phantom vibration syndrome.

World Health Organization included gaming disorder as an addictive behavior disorder into the ‘International Statistical Classification of Diseases and Related Health Problems’ since such dependence also causes quick and sickly addiction, subdues the will, snatching people away from a fulfilling life. Is this a behavioural addiction?

Our brain reacts to smart-phones the same way as it does to a slot machine – each notification, refreshing the email or scrolling social media feed activate a dopamine reward system giving us a small portion of anticipation pleasure. We scroll the news feed for fear of missing something important – for brain it is a trigger similar to the one that makes a gambler pull the handle over and over again, out of fear of missing the jackpot.

This is a natural reaction because evolution has formed our brain in a natural environment when for survival it was important to learn how to instantly respond to changes, give priority to new information, take note of the fellow tribesmen experience getting their approval and tap for the sources of energy by mobilizing emotions. Biologically, we are not very much adaptable to exist in an artificial environment such as the information society in its present form. Our brain’s evolutionary advantages become easily exploitable. For example, reduction of uncertainty around us is a must for survival – in an evolutionary sense to know who run away from and whom to hunt. The connected world with smart-phones and media platforms provides us with information to reduce uncertainty anywhere at any time. Manufacturing digital products involves not only content producers, who make products look interesting and relevant, but also designers who make them beautiful and attractive. They are supported by marketers who know what’s in demand and the best practices of selling, by psychologists with their knowledge of the properties and vulnerabilities of the psyche, by neuro-physiologists who understand how to get the brain to respond in a desired or programmed manner by releasing the neurotransmitter dopamine.

As MIL includes different aspects of our interaction with media, we should understand what underpins our actions. Nir Eyal, author of the world best-seller 'Hooked. How to Build Habit-Forming Products' explains in detail how digital platforms influence and manipulate human habits and behavior. In 2019, he published a new book written from the other side of the fence that is called 'Indestructible: How to Control Your Attention and Choose Your Life'.

2.4 How MIL Can Practically Empower Citizens in The Digital Era

A number of negative or alarming features were highlighted above. What could be done to improve the situation? There are some options – from utopian to practical ones.

For example, the current economic and political systems could be reinvented to better help humans in expanding their capacities and capabilities. Or, people could join forces to innovate widely accepted approaches in order to make them aimed at open, decentralized and intelligent networks. Mass activism could produce tangible results towards ethical, trustworthy and responsible AI [5].

Such considerations of AI are 'would be very good' developments, but what are the most practical and realistic steps that are within our grasp? It is definitely necessary to raise awareness and educate people in order to help them understand how modern information and communication systems work and how to deal with them, specifically focusing on AI.

Let's be reminded of the UNESCO definition of MIL:

> "MIL is a composite set of knowledge, skills, attitudes and practices that allow effectively access, analyse, critically evaluate, interpret, use, create and disseminate information and media products with the use of existing means and tools on a creative, legal and ethical basis".

Although these skills are very important, penetration of AI and big data technologies into communication and information domains has given rise to new challenges to be addressed by MIL. AI changes both the media environment and our experience within this environment.

There is less and less information in this changing media environment that could be "accessed, analysed, critically evaluated, interpreted, used, created and disseminated" within the conventional set of MIL-related competences. Here are the six questions that are used for media analysis and which should lead to an understanding of media message:

- Who created, or paid for, the message? When?
- What is the message about and why is it being sent?
- Which techniques are used to attract my attention?
- How might different people interpret this message?
- What are the direct and indirect messages?
- What is omitted from the message?

These questions are good to analyse an article in a newspaper – they allow making some assumptions about the explicit and implicit motivation of the author of the media resource, de-constructing some biases (by omission, by placement, by spin), etc. Do these questions work with Facebook feed? No, they are not enough and even might be misleading in case we ignore the software dimension in the social media structure, i.e. how algorithms of Facebook feeds work. Spotting facts from opinions, fact-checking, text analysis (denotation, connotation, etc.), critical inquiry of headings and captions, illustrations and info-graphics, all these important and powerful MIL tools are still in use and still work. Are they enough to analyse Google search results? The answer is no. Classic MIL skills work with separate posts and separate social media accounts, separate sites, which might be included into search results. But to be truly media and information literate, to be real actors in the digital media environment, one should be aware of how the digital platforms function, how they are integrated into the data industry, advertising industry, etc.; how various kinds of AI-based technology operate, what kind of social effects they provoke, what kind of ethical issues they raise.

It is important to note that not only smart-phones, tablets and computers are communication devices. With growing 'datafication' and the development of the Internet of things, any consumer device, any household or city appliance that sends and receives data is a communication device. Even a small thermostat with a microphone and camera can record information and data and transmit it to a third party and as such it is a communication device, a medium. Then, it requires not only technical skills to set the temperature, but also MIL skills to manage this data mining and sharing activity and even to decide whether one is ready to use such kind of devices or not.

Therefore, digital content is packed within digital services, which leads users to do things that people would not planned to do – to part with their time, attention and money, to share their personal data and digital footprints. This side of technologies is hidden, not obvious and hardly can be identified or reflected on through direct experience. People born in the digital era often consider that kind of communication architecture as natural and the only possible. That all makes the need to adapt MIL skills and attitudes to new technological reality even more essential.

There is a very famous phrase coined by Marshall McLuhan, a brilliant Canadian philosopher and communication thinker: “The medium is the message”. His idea was that a medium itself, not only the content it carried, should be the focus of study. He said that the features of media were important to understanding its impact on the society. That idea has given a push for the development of media studies and contributed significantly to MIL.

Further to the idea, new media theorist Lev Manovich argues, “The software is the message” as “a universal engine which the world runs on” [4]. Thus, the communication process, that is the main subject of research and teaching for MIL specialists, should be studied and taught at three interconnected dimensions, or layers:

- media texts,
- media tools and channels, and
- software (technology).

The MIL-related set of competencies and attitudes should be updated accordingly. In addition to the conventional MIL topics, MIL education programme should cover AI, big data, Internet of things and other industry 4.0 technologies, as well as cultural, economic and even neuro-biological dimensions of modern media environment. This is the approach that is used in this publication, which opens with explanatory materials about what AI is and how it works, continues with reflections on the ethical, legal and social consequences of its application and its impact on fundamental human rights.

Given the deep penetration of digital technologies into our lives and the current economic model of their development described by the notions of attention economy and surveillance capitalism, new MIL, or digital MIL, should include a set of knowledge, skills, attitudes and practices that enables people

- to retain not only critical, but cognitive and behavioural autonomy
- to manage attention and behavior in media environment
- to 'hack', or at least identify and minimize the damage of biases and personalization effects, such as filter bubbles and echo chambers
- to understand how technology influences the social fabric
- to update skills and competences according to the life-long learning approach

The successes of the MIL depends on the appropriate mindset that allows to tolerate and accommodate the increased complexity and uncertainty of the world. This mindset ensures the ability to live in a world where models cease to be deterministic, such as in AI.

MIL-mindset includes understanding that there is no longer the one right way or just one right answer. The robustness should be prioritized before efficiency, meaning that multiple solutions should be prioritized before one right solution. There needs to be a change in the mindset towards adaptation, variation and invention. In this context, the path towards a goal is more important than expected results. We should be more oriented to 'the way' or 'the journey' than 'the result'.

2.5 Critical Cognitive and Behavioural Autonomy

Our digital communication and interaction are framed by intelligent techniques that are used by social media platforms, digital services, etc. So, to be media and information literate, people should be provided with practical ability to raise their sovereignty in the information and communication contexts:

- be able to focus on performing their own task without distraction and ignoring non-thematic media messages or procedures,
- interact and to be a part of the media communication on their own rules according to their own agenda (their own Whats, Whys, Whens and Hows).

The important part of the Digital MIL is, therefore, critical thinking. What should critical thinking include? First of all, good intention and fact-checking skills are not enough. To be instrumental in new media environment with sophisticated techniques and persuasive design, critical thinking should include four 'selfs':

- self-directedness,

- self-disciplining
- self-monitoring, and
- self-corrective thinking.

Accordingly, critical thinking is an ability to actively and skilfully conceptualizing, applying, analysing, synthesizing and/or evaluating information gathered from or generated by observation, experience, reflection, reasoning, or communication [2]. As biases and filters limit capacity of meaningful consideration, to de-bias and de-filter cognitive ability, it is crucial not to be rigid, continuously question own convictions, beliefs and opinions. Critical thinker should have flexible mental frames, cultivate cognitive complexity, imagination and experimenting. Instead of looking for the only correct answer, critical thinking helps noticing the fuzziness where others are only able to see the obvious and the banal. By doing so, one can better comprehend the complexity of the world, its rapid fluidity and uncertainty.

To avoid limited reasoning, one should be attentive to opponents views using 'consider the opposite' strategy. One should be open to receiving additional information beyond typical realm of understanding. From the point of view of critical thinking, 'what we do not know' is much more important than 'what we know'. This somewhat paradoxical idea reminds us that the experience and knowledge of each of us is limited and the point of view is just a point that allows seeing only a part of reality from a certain position. Nobody's point of view can be exclusively correct. This understanding creates space for a more tolerant world-view that allows for different visions, positions and opinions. It is the interest in what is unknown that becomes an engine for development and a more complete understanding of how the world works.

In order to adequate modern challenges, critical thinking should also be supported with theoretical and practical knowledge about the principles of operation of the sources of information and communication, the ways information interaction is mediated by digital platforms and devices and the modes of functioning of the AI systems of search engines, social networks and recommender systems. This pattern of thinking would enhance understanding of the process of collecting, distributing, comparing and using data. It also helps to understand how AI may be used in creating and personalizing the distribution of images, texts, video and audio recordings, including the so-called fakes. Combined with the traditional skills, attitudes and practices of MIL, these new competencies should make it possible for citizens to become more self-reliant, skilful and empowered.

2.6 Concluding Remarks

For the last couple of decades, the meaning of literacy has gone beyond the abilities to read, write, count, and so forth. Now that the nature of information is dependent on various media, literacy has taken up a critical position in every part of life for people of all ages. UNESCO has taken a pioneering role in different countries for people to improve media and information literacy and develop critical competencies. In order to sustain such activities, training programs, strategies and resources, it is a must to rely on new digital technologies and AI for the management of rapidly evolving media environments.

Socio-economic state prefigures the components of MIL competence; technical, cognitive, social, civic and creative capacities and hence how people understand, evaluate and productively interact with media. The term 'digital transformation' implies important perspectives for MIL which depends largely on big data, the cloud and the AI. The unpredictable technological advances flood into every layer and system of life so rapidly that, especially through AI and related technology, the emerging reality brings 'complexity and uncertainty' around MIL domains and, hence, life. Similarly, AI provides new ways of data collection, dissemination and regulation which suggests drastic 'changes in the media environment' creating an urge to change MIL approaches as well. AI fosters increasing amount of data to flow ubiquitously which evokes subjective perception of truth, hence objective facts are not determinant in the 'Post-truth Era'. Finally, the nature of economy is being influenced by a new 'attention economy' with the wide application of big data, machine learning and AI technologies. The biggest companies cultivate 'attention' through high-tech expensive services they give for free but profit from advertisements, collecting user data, forming opinions, attitudes, beliefs and habits. Among these dynamics 'surveillance capitalism' is triggered by continuous monitoring of data owners which results in undesired surveillance and social control.

How people interact with and through AI-driven systems cause MIL-related challenges; 'filter bubbles' stands for the limited or biased access to information due to personalization and recommender systems. Similar filter bubbles create communities by time, and overlapping interaction create 'echo chambers' where like-minded people support one another simply through repetitions. The nature of these closed communities may lead radicalisation of opinions and facts. AI-powered algorithms have changed 'decision making' for MIL, where data driven decisions are hidden to the user. So, the question remains about the responsibility of any probable bias or mistake. Also notable is that through AI systems, actions on media are shifted from communication to a strange 'addiction'.

A new approach to MIL can practically empower citizens in the digital era with a mindset which involves adaptation, variation and invention but also multiple solutions rather than one answer. In order to reach 'cognitive and behavioural autonomy' it is a must to be critical about theoretical and practical knowledge of information and communication work, how digital platforms and devices mediate our information interaction, and how the AI systems of search engines, social networks and recommender systems operate.

References

1. BCG Henderson Institute (2019, April 12). *Winners are Changing Rapidly*. Retrieved from https://bcghendersoninstitute.com/winners-are-changing-rapidly-chart-of-the-week-15-2019-33e9ce28d331
2. The Foundation for Critical Thinking (n.d.) *Defining Critical Thinking*. Retrieved from https://www.criticalthinking.org/pages/defining-critical-thinking/766
3. Jiang, J. (2018, August 22). *How Teens and Parents Navigate Screen Time and Device Distractions.* Retrieved from https://www.pewresearch.org/internet/2018/08/22/how-teens-and-parents-navigate-screen-time-and-device-distractions/
4. Manovich, L. (2014, March 10). *Software is the Message.* Retrieved from https://journals.sagepub.com/doi/abs/10.1177/1470412913509459/
5. Pew Research Center (2018, December 10). *Artificial Intelligence and the Future of Humans*. Retrieved from https://www.pewresearch.org/internet/2018/12/10/artificial-intelligence-and-the-future-of-humans/
6. Sullivan, A. (2016, September 19). *I Used to Be a Human Being*. Retrieved from https://nymag.com/intelligencer/2016/09/andrew-sullivan-my-distraction-sickness-and-yours.html
7. Taleb, N. N. (2007). *The Black Swan: The Impact of the Highly Improbable* (Vol. 2). Random house.

Key Readings

- Carlsson, U. (2019). *Understanding Media and Information Literacy (MIL) in the Digital Age: A Question of Democracy*. Gothenburg: University of Gothenburg.
- The Grünwald Declaration on Media Education, (1983). Educational Media International, 20:3, 26, DOI: 10.1080/09523988308549128

- Ptaszek, G. (2019). From Algorithmic Surveillance to Algorithmic Awareness: Media Education in the Context of New Media Economics and Invisible Technologies. *Media Education as a Challenge*, pp. 59-73. Retrieved from http://www.interreg-danube.eu/uploads/media/approved_project_public/0001/42/666106e88e1438c2b0cc4866cb31d7d8909933bd.pdf
- UNESCO (2011). *Fez Declaration on Media and Information Literacy.* Retrieved from http://www.unesco.org
- UNESCO (2012). *The Moscow Declaration on Media and Information Literacy.* Retrieved from http://www.unesco.org
- UNESCO (2013). *Global Media and Information Literacy Assessment Framework: Country Readiness and Competencies.* Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000224655
- UNESCO (2014). *Paris Declaration on Media and Information Literacy in the Digital Era.* Retrieved from http://www.unesco.org
- UNESCO (2016). *Riga Recommendations on Media and Information Literacy in a Shifting Media and Information Landscape*. Retrieved from http://www.unesco.org
- UNESCO (2016). *Khanty-Mansiysk Declaration on Media and Information Literacy for Building a Culture of Open Government*. Retrieved from http://ifapcom.ru/files/2016/The_Khanty-Mansiysk_Declaration_web.pdf
- UNESCO (2017). Courier #2 *The Media: Operation Decontamination*. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000252318/PDF/252318eng.pdf.multi
- UNESCO (2018). *Journalism, 'Fake news' and Disinformation: A Handbook for Journalism Education and Training*. Retrieved from https://en.unesco.org/fightfakenews
- UNESCO (2019). *Belgrade Recommendations on Draft Global Standards for Media and Information Literacy Curricula Guidelines*. Retrieved from https://en.unesco.org/sites/default/files/belgrade_recommendations_on_draft_global_standards_for_mil_curricula_guidelines_12_november.pdf
- UNESCO (2019). Media and Information Literacy at UNESCO Portal. Retrieved from https://en.unesco.org/themes/media-and-information-literacy
- Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* Profile Books.

Key Videos

- Internetstiftelsen (2019, November 25). *The Social Vulnerabilities of AI*. [Video file]. Retrieved from https://www.youtube.com/watch?v=spoDsPrvQNY
- Netflix (2020). *The Social Dilemma (Documentary)*. [Video file]. Retrieved from https://www.netflix.com/ru-en/title/81254224
- TEDGlobal (2017). Zeynep Tufekci. *We're Building a Dystopia Just to Make People Click on Ads.* [Video file]. Retrieved from https://www.ted.com/talks/zeynep_tufekci_we_re_building_a_dystopia_just_to_make_people_click_on_ads
- TNW (2014). Nir Eyal *Building Habit-Forming Products*. [Video file]. Retrieved from https://www.nirandfar.com/new-video-hooked-the-psychology-of-how-products-engage-us/

3. AI Impact on Human Rights and Freedom of Expression

Igor Shnurenko

In this chapter, a huge societal transformation caused by the arrival of AI is placed into the perspective of human rights (HR) and basic universal freedoms. AI could be leveraged to create new global opportunities and we investigate some AI use cases when civic groups, businesses and international organizations apply AI solutions to fight discrimination, inequality, poverty and bias. But in an increasingly turbulent world, we share as a global community, ubiquitous and sometimes indiscriminate AI adoption creates new uncertainties and risks. We look into the infringements on freedom of expression (FoE) and other human rights which are explained not only by unintended consequences of AI application but also by some of its inherent features.

The big tech corporate overreach and behavioural surplus economy may impact freedom of expression in yet unexpected ways. There exist also existential human rights dangers originating in some fundamental aspects of AI. The dual aspect of AI systems is examined, one that offers users more choice, another that reinforces biases and diminishes the individual agency to seek and share diverse opinions and ideas. AI could be exploited by unscrupulous state and corporate actors to embed and perpetuate bias and discrimination.

In the sections that follow, we start with creating a context for AI in relation to human rights. Next we present inherent features of AI that have direct relationship with HR and we conclude with discussion on AI trends influencing

HR. The discussions include major AI-related risks and uncertainties as well as how human rights due diligence helps work out responsible approaches to answering all of these challenges.

3.1 The Context for Human Rights and AI

3.1.1 Human Rights in Information Age and The Arrival of AI

Human rights are a set of universally accepted norms and moral principles that stitch together the fabric of modern society. They describe those inherent aspects of an individual human being that define certain standards of human interaction in and with society and state. Although the debate is still ongoing about their exact content, nature and justifications, human rights constitute the foundation, commonly understood as inalienable, upon which many other institutions were built, including the modern state itself.

It's universally recognized that human rights are applicable to all persons regardless of their sex, age, ethnicity, language, religion, or any other status every time and everywhere. Human rights are defined in the Universal Declaration of Human Rights. In 1966, the United Nations adopted the International Covenant on Civil and Political Rights (ICCPR) [8]. In many countries, they are inscribed into national constitutions and in others, they are part of the common law. By all means, human rights are protected by law enforced by international and national bodies.

Human rights encompass a variety of rights, including freedom of expression, free speech, the right to liberty and security of person, the right to liberty of movement and freedom to choose a residence, the right to a fair trial and to remedy if any of the rights were violated. Privacy as an ability of persons to express themselves selectively is also considered one of the fundamental rights. According to the ICCPR, there are not only individual but also collective rights, such as the right of self-determination, the right of peaceful assembly, prohibition of slavery, etc. Over the last few decades, social rights have acquired ever greater importance, especially in developing countries, for example, the right to safe housing or clean water.

As is often the case with the arrival of new technology, the advent of artificial intelligence has triggered a huge transformation in the information environment surrounding humans. As a result, human behavior is also going through a process of modification, changing not only the balance between society, science and technology but also the whole fabric of society at large.

Accordingly, the scope of human rights and the practice of their application are undergoing significant transformation.

While there are many observable cases where AI-driven systems serve for the benefit of individuals and communities, AI is often linked to transferring the decision-making process from humans to machines resulting in control over citizens.

This 'constellation' of technologies and processes enables computers to replace or complement specific tasks otherwise performed by humans. The increasing independence, speed and scale associated with the automated, computational decision-making by AI-driven systems may cause possible violations of certain human rights and liberties.

Freedom of speech is a principle according to which an individual or a community may articulate their opinions and ideas without fear of censorship, legal sanction, or retaliation. The principle includes the right to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers. Freedom of speech is essential to democracy since the informed electorate is a necessary condition for self-government by the people. That's why in democratic countries, the freedom to hold opinions is considered an absolute right that permits no exception or restriction whether by law or other power.

However, the ways in which information is stored, transmitted and secured in the digital age affect the exercise of this right. Opinions are being expressed now mostly in the digital form mediated by privately-owned platforms and applications where AI systems play a growing role in their selecting, ranking, distribution and even erasing and sanctioning the users for their perceived faults. The fabric of opinions now strongly depends on AI-controlled browsing activities, AI-mediated algorithms for search queries and the news feed in social networks, on keeping and having access to the user's digital records in the cloud and on using email and messenger communications. Governments and big private corporations who control or own the means of communication may interfere with these mechanics and processes of forming and holding opinions by individuals and within communities.

Freedom of expression is particularly important for the mass media that in a democracy plays a special role as the bearer of the people's right to know and freedom of expression for all. It's important to note that this right also requires freedom from undue coercion in holding or development of beliefs and ideologies. In his work 'On Liberty' (1859), the philosopher John Stuart Mill argued that "...there ought to exist the fullest liberty of professing and discussing, as a matter of ethical conviction, any doctrine, however immoral it may be considered." [5]

The right to expression is not an economic good that may be exchanged for money, security or the opportunity for employment. This is true in periods of peace and prosperity but even more so in the times when lives and security of people are at stake. But there are numerous examples of how algorithms and AI-mediated networks deem some information inappropriate and take down the content that would not be considered as such on the public square. AI systems so far lack contextual understanding, but nevertheless they are tasked with filtering and effectively censoring perfectly legitimate content from platforms.

As a general principle, freedom of expression should not limit the right to privacy as well as the honour and reputation of other persons. Greater latitude is normally given when someone criticizes public figures. Public debate should not be completely suppressed even in times of emergency such as for example Covid-19. In all cases, the value of freedom of speech, freedom of expression and, more and more, freedom of the press, depends on the way people conduct their social interactions.

Today, AI-driven systems are having increasingly more and more control on the information environments. The products and processes of artificial neural networks (ANNs), algorithms and AI applications are now found in every corner of the Internet, in search engines, on social media platforms, digital devices and various technical systems, messaging applications and public information mechanisms. Algorithms and AI applications determine now how widely, when and with which audiences and individuals content is shared and in many cases, how it is created. In the situation when AI is so ubiquitous, it's much easier for big corporations that essentially own the online space, to exert their control not only over the technical, but also the substantial side of communications. The flow of information becomes more and more dependent on how ANNs select content for users and by them and on commercial impulses to promote selected content. In this respect, the phrase "The medium is the message" coined by the Canadian philosopher Marshall McLuhan becomes the rule.

One example of AI-driven restrictions and controls is the Golden Shield Project, an initiative by the Chinese Ministry of Public Security that automatically filters even potentially unfavourable data flow coming from foreign countries.

Human agency is integral to AI, but the distinctive characteristics of AI deserve human rights scrutiny with respect to at least three of its aspects: automation, data analysis and adaptability.

Automation removes human intervention from parts of a decision-making process, completing specific tasks with computational tools. This can have positive implications from a human rights perspective if a design limits human bias. For example, an automated border entry system may flag individuals for scrutiny based on objective features such as criminal history or visa status, limiting reliance on subjective (and bias-prone) assessments of physical presentation, ethnicity, age, or religion. But the experience shows that sometimes AI-governed systems may reinforce existing biases instead, creating discriminatory effects.

3.1.2 Uncertainty of AI and Human Rights Due Diligence

AI is an intrinsically ambiguous technology with a lot of uncertainties in its methods of knowledge acquisition, model construction, data processing and decision-making. In a highly turbulent world we live in, the human rights framework for inquiring into the development and use of new technologies becomes ever more important. The information environment is a complex ecosystem of technologies, platforms, private and public actors. It has become digital so that both access to information and its dissemination are now performed mostly through AI-driven means of communication. The potential of AI is great and so is the consequences of its unintended uses should something go wrong, especially with an uncertain future we all face now.

Growing AI adoption may lead to the unforeseen and unintended consequences. The lack of transparency and 'explainability' of AI, the lack of correction mechanisms as well as some other factors may infringe freedom of expression and free speech.

Even the experts' knowledge and intuition and the experience of practitioners may not be sufficient to evaluate the effects of applications not previously anticipated. That's why a society-wide dialogue is needed. It should involve activists and business people, public figures and scientists, AI experts, tech companies representatives and officials.

To help experts and civil society, it is important that innovative methods of human rights due diligence be developed and tested. Human rights-based standards and methodologies should form an essential part of business policy and practice. They provide a robust framework for the responsible development and use of AI.

As AI is becoming ubiquitous, the attention to issues of human rights should not be limited only to the companies of so-called 'big tech' and technological start-ups. Companies working in other sectors such as logistics, fin-tech, retail, transport and services should also be proactive in using the human rights framework for controlling their AI-driven systems. Governmental and non-governmental organizations should lead the developing responsible approaches to human rights challenges.

Civil society as a whole should be continuously evaluating the risks that new AI-based systems may pose for human rights and should be working along with governments, non-profits, social entrepreneurs and individuals to use the robust HR due diligence framework to uncover blind spots and prevent dangerous developments.

3.1.3 AI and The Right to Privacy

To train, calibrate and refine AI systems, big data sets are mined. Collection of data may interfere with rights to privacy and data protection and their following analysis may reveal private information about individuals.

Information should be treated as sensitive even if derived from big data sets fed from publicly available information. For example, researchers have developed machine learning techniques that can accurately guess individuals' age, gender, occupation and marital status just from their cell phone location data. Even more could be derived from Facebook's emoticons. It is possible to predict a person's future location from past history and the location data of his or her contacts. A lot can be foretold using quite simple data. British researchers even demonstrated the ability to predict 'satisfaction with life' from Facebook messages.

To protect the right to privacy and other human rights, the information derived from raw data and processed must be treated the same way as any other personal data.

3.2 Features and Uses of AI in The Context of Human Rights

3.2.1 AI's Lack of Predictability: AGI vs. Narrow AI

Some authors suggest that society is headed towards the so-called 'technological singularity' that is the arrival of artificial general intelligence (AGI) and perhaps the Artificial Super Intelligence (ASI). There are many definitions of what it means, but the most known one is that it is the ability of a computer system to approximate (AGI) or surpass human intelligence (ASI) across multiple domains.

While popular culture depicts images of AGI-governed dystopia, most scientists and experts agree that this capability seems to be still-distant for computers designed the way they are today. So, for foreseeable future, this threat seems to be exaggerated, while there are many recent advancements with respect to narrow AI, where computer systems act is limited to specific domains. They perform certain tasks, mostly using a complex machine learning systems and human-developed algorithms. For example, narrow AI supports voice assistance on mobile devices and customer service chatbots, online translation tools and self-driving cars, search engine results, mapping services and so on.

With the recent Covid-19 pandemics engulfing many countries, apocalyptic scenarios regarding even narrow AI seem to become more pronounced. While we shouldn't fall for superficial repetitions of popular 'memes' meant to frighten and amuse, let's not discount some clearly alarming tendencies.

Lack of predictability is an aspect of AI that may hold some promise for global technological transformation, especially in machine learning-driven ANNs. But we already have enough evidence to illuminate its risks. As humans are progressively excluded from defining the outputs and even objectives of AI systems, it becomes much more difficult to ensure their transparency, accountability and access of users to effective remedy if their rights are violated.

3.2.2 The Dual Aspect of AI-Driven Personalization

AI-driven personalization of products and services has a dual aspect in it.

1. It offers users more choice and caters to their needs, helping them get access to variety of opinions. People also have more opportunities for expressing their own views and, by using global platforms, making them known to the many.

2. But personalization clearly has another, potentially dangerous aspect to it. It may interfere with the individual agency to seek and share opinions and ideas across political, ideological and societal divisions by minimizing exposure to diverse views. Large private social networks may personalize content by reinforcing biases, creating filter bubbles. In order to sustain users' online engagement, they also incentivise the recommendation and promotion of inflammatory content and outright disinformation. The media could also have conscious and unconscious biases that are reinforced by AI-driven recommender systems.

Corporations design AI-governed systems for personalization with the aim of micro-targeting. But, for example, the deployment of micro-targeting through social media platforms creates a curated world-view that hinders pluralistic political discourse. Although this is not a direct violation of freedom of expression and freedom of the press, it created an environment inhospitable to free speech.

The corporate monopoly of the online search market makes it almost impossible for users to opt-out of the algorithmic ranking of search results. The search monopolists intend to make people believe that their search results are truly the most relevant or objective information available on a particular subject. As a rule, this notion is misleading. It is closely related to another misconception that AI objectively presents factual information. Efforts to automate content moderation through AI also come at a cost to human rights.

3.2.3 Responsibility Delusion in Automated Decision-Making

In an AI-driven system, the dissemination of information and ideas is governed by opaque forces that invoke the black box analogy. Their priorities may be at odds with an enabling environment for media diversity and independent voices. Combined with undue media dominance or concentration by privately controlled groups, it may be harmful to a diversity of sources and views thus infringing on the freedom of expression and free speech. The UN Human Rights Committee and other international rights bodies have found the situation dangerous and argue that states should take appropriate action in this respect, preventing monopolistic tendencies and controlling the application of new technologies.

Growing reliance on and confidence in automated AI-governed decisions creates a situation of transfer of responsibility from those who are authorized to make decisions to machines. It creates a delusion that the machine has, in fact, the capacities of moral agents. In that case, machines could have more rights than humans that serve only as subjects to the actions of AI-induced agents. Humans are considered unreliable, biased and erroneous, while algorithms supposedly can be fully trusted and do no wrong. The responsibility delusion leads to alienation of people from their own experience as what counts more is the machines' act. It also disables individuals from accessing remedies to adverse AI-driven decisions that may infringe on their rights. It also undermines scrutiny of AI outcomes which is very important when new systems are being developed.

3.2.4 Dangerous Use of Consumer Data

AI-driven decision-making systems depend on the collection and exploitation of data, ranging from non-personal information to the data that could be used to identify people. The vast majority of data used to feed AI systems falls in the middle—as, for example, anonymised personal data or behavioural products manufactured from personal data. Companies use data extracted from algorithmically inferred digital fingerprinting and online profiling. To feed AI systems, they also buy or trade datasets from third parties including aggregators.

International data protection standards rest on notions of consent, limited use and clear purpose, transparency and accountability. AI challenges all that. AI-governed autonomous systems and other consumer products like a home or medical devices are often equipped with sensors. Around the clock, they collect vast amounts of real-time data on all individuals within their reach. As it has been shown in academic publications, global social media platforms could use AI-driven methods to infer sensitive information about persons without their awareness. By doing that, they create or refine existing profiles adding inferred information about health conditions, family relationships, religious views, sexual orientation, or political affiliation.

The resulting behavioural products could use the inferred data that allows the platforms or their customers to manipulate people, nudge and herd them or subject them to social conditioning. “No one shall be subjected to arbitrary or unlawful interference with his privacy, family, home or correspondence, nor to unlawful attacks on his honour and reputation”, says article 17 of ICCPR and this practice clearly infringes on this right. It also violates ‘the right to hold opinions without interference’ (Article 19) and ‘the right to freedom of thought, conscience and religion’ (Article 18).

Collected datasets used to feed AI-driven systems serve only as raw material for manufacturing new products. That’s why individuals can hardly exercise any degree of control over their data. So in the context of AI, the mentioned notions are basically deprived of any practical meaning. Once data are re-purposed in an AI system, they lose their original context, depriving individuals of the ability to repair or delete them. Thus, the risk that data about persons becomes inaccurate or out of date increases. Using the data, AI-based systems make important decisions, in many cases profoundly affecting people’s lives. Yet, individuals have few avenues to exercise control over decisions based on products that have been derived from their personal data. Corporations speak of anonymisation techniques, but as some scientists have shown, they are inadequate to the task and can only be used as a smokescreen.

In her book, 'Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy', the publicist Cathy O'Neil exposes how AI-centred systems are being used for sentencing people, for choosing who gets the job, who gets the interview, who gets the loan and who gets the house [6]. Similar approaches are applied in China's system of social rating that as the Chinese authorities say has proven its effectiveness during the corona-virus crisis.

In any case, human rights due diligence is required to make sure AI-based systems are not a distraction from making responsible decisions and do not hinder the rights of individuals to remedy. "The era of blind faith in big data must end", says O'Neil.

3.2.5 AI-Based Surveillance and Tracking

When journalists working for the British newspaper Daily Telegraph arrived at their newsroom on January 11, 2016, they discovered small, black rectangular boxes labelled 'OccupEye' attached to the underside of every desk. They were baffled by those unusual pieces of office equipment and suspected a foul game. Having googled the brand name, they discovered that the boxes were wireless detectors that collect information about individuals. Their sensors, 'ultra-sensitive, yet ultra-reliable',—as the company's website boasted—were triggered by motion and heat. As a result, the management could track their workers real-time [10].

When the news went public, an official explanation was sent to staff saying that the devices were installed only temporarily to make the building energy-efficient as a part of The Telegraph's commitment to green energy. This explanation didn't hold water, however. The OccupEye's official website focused on savings that their clients could make by cutting their personnel while using the device. Journalists also suspected that the management could use the data gathered about the employees to exert control without their awareness. The devices were, in fact, part of a system of 'automated workspace utilization analysis' designed to provide detailed metrics on worker attendance.

The journalists of The Telegraph have finally had their surveillance devices removed, but the trend persists. Tracking activities of employees come a long way from the physical surveillance of the private detectives of the 1850s, to the closed-circuit cameras and email monitoring of the 1990s and to AI-based apps that go beyond merely monitoring productivity in the workplace.

Technological advancements in several fields—AI, mobile devices, DNA testing and biometrics—have dramatically expanded capacities for worker surveillance both on and off the job. Now employers get access to the health data, individual behaviours and personal characteristics of their employees. Even when giving their consent to data collection, workers rarely understand what exactly those productivity apps and wellness programs do. With intrusive surveillance by AI-driven systems much more achievable and economical than before, 24/7 tracking is becoming a reality. As the new technological advancements often go unnoticed by NGOs, governments and even organized labour, employees rarely can do anything about it. From the position of human rights due diligence, there should be sweeping legal changes to address these concerns.

Personnel tracking starts with data-mining for head-hunting that may include not only background checks, credit checks, criminal records, but also predisposition to health risks, genetic testing, psychometric testing, drug testing and biometrics. There are various AI-based applications that mine, compile and process all that data resulting in important personal decisions. Specific applications may track productivity, behaviours, personal characteristics, use of company resources spying on communications and geo-location of employees, their interaction not only with customers and other employees but also with members of their families, people on the street and so on. In most cases, the black-box technologies presented in such applications won't allow users to exercise moral or legal agency so all this may constitute 'arbitrary or unlawful interference' with privacy, family, home, correspondence, or 'unlawful attack' on honour and reputation. This may also constitute an infringement on the 'right to the protection of the law against such interference or attacks'.

3.2.6 AI-Enabled Affect Recognition

The next stage of facial recognition technologies is 'affect recognition', that claims to 'read' our inner emotions and motives by measuring and interpreting the micro-expressions on faces, body movements and gestures. By doing this, not only emotions, but also personal values, motives and character traits are deduced from the physical behavior of individuals. Critics compare the logic of its method to the discredited physio-gnomic views and race pseudo-science proclaimed official in Nazi Germany and used to justify racial and national inequality. As affect recognition encodes biases, some experts say it lacks robustness to make sure its results are accurate or even valid. However, AI-enabled affect recognition technology is spreading at scale. It is used in classrooms, job interviews, personnel tracking to make judgments who is ‘attentive’, ‘productive’, ‘honest’, ‘good worker’, etc. Often it happens without people’s awareness. Again, there’s no way for a person to seek a remedy of a decision harmful for this person which also constitutes a human rights violation.

3.2.7 AI in Content Moderation

Social networks and other major platforms use automation to moderate content and that comes at a cost to human rights. The moderation in social networks has a three-level structure, with most algorithmic and AI-based measures taken on two lower levels. First, special filters assess whether the uploaded content possesses certain characteristics of ‘unwanted’ material. If the answer is ‘yes’, the content is blocked from being published and depending on the algorithm, its author or purveyor may be punished. If the material passes through the filter, it goes online, but AI still decides which particular user will see which piece of it for how long.

This ranking of the published material essentially constitutes the second level of its filtering. Algorithmic decision-making is one of the most guarded secrets of the system, with the history of interactions, characteristics and data about the user and other criteria are taken into consideration. Basically, a black box decides whether you see the post or not.

Human-in-the-loop appears on the third level of filtering. Here, users could report ‘inappropriate’ content that triggers a reviewing procedure by human moderator. Although the content’s review is judged based not on any national or international constitution or legislation, but rather on the platform’s internal rules known as Notice-and-Take-Down procedures (NTD), the remedy is still possible in principle.

By all accounts, even the third level of filtering has insufficient transparency. As for the first two, the machine learning techniques lack even basic senses and capacities of moral agents able to make judgments. Thus, the whole procedure should be carefully looked upon in terms of its decision-making process and human rights due diligence approach applied.

Automatic content filtering is not limited to social networks. For example, YouTube uses an AI-based ranking algorithm that is being automatically fine-tuned to promote certain videos and restrict others.

In one instance, YouTube removed over 100,000 videos documenting war atrocities in Syria after they were automatically flagged as inappropriate content. But such videos often serve as the only evidence of human rights violations. The company's policy carves out exceptions for showing violence when the important educational or historical value is concerned. However, those videos were still taken down. In other examples, historical photographs with particular cultural significance were removed by Facebook on the grounds of being 'indecent', like in the Napalm Girl case [9].

3.2.8 AI-Driven Micro-Targeting

AI-driven personalization often impedes individual agency to seek and share ideas and opinions, minimizing exposure of individuals to diverse views and reinforcing stereotypes and divisions across the ideological and political spectrum. Such personalization may also incentivise the dissemination of inflammatory content or disinformation. Similarly, AI-driven micro-targeting used by large social networks and platforms creates a world-view inhospitable to pluralistic political discourse.

3.2.9 The Corporate Power and The Behavioural Surplus Economy

In April 2020, the Federal Trade Commission fined Facebook for $5 billion as a result of an investigation of the case of the British political consulting firm Cambridge Analytica. Using AI, Facebook gathered, processed and sold behavioural data from its users to the firm without their awareness. Cambridge Analytica used online footprints and personality assessments to tailor messages and content to individual users in a campaign of 'psycho-graphic' micro-targeting. As a result, major manipulations of political opinion took place which violated the people's rights to self-government, infringed on the right to vote, be elected and hold opinions without undue interference.

Since its launch in 2004, Facebook conducts social experiments that play with human interactions in ways that could lead to dangerous consequences, such as emotional contamination which was so spread-out during the Covid-19 pandemics. As just one example, the company paid teenagers aged from 13 to 17 to install on their phones what was essentially equivalent to spyware allowing the company to harvest every transaction that these teenagers were having with their friends – without their friend's knowledge.

But the real elephant in the room is the current business model of big tech that was developed and polished during the experiments mentioned above. Big data is being captured, mined, processed, harvested and used without people's consent in order not only to predict their behaviours but also to nudge, herd and condition their actions. In the process, a number of basic human rights are being grossly violated without most people's even noticing. Currently, the 'surveillance capitalism' business model spreads all over the world unhindered. Some experts, for example, the author Shoshana Zuboff, see it as one of the biggest existential threats that humankind faces [11].

3.2.10 'Prediction Factories'

Major hedge funds, secret services and other institutions acquire the AI-based predictive products that suck up data from all over the world to make predictions about terrorist attacks, social and international conflicts and reactions of nations to socially disruptive, extraordinary and catastrophic events such as epidemics. Sometimes they use the tech for the benefit of society, but often the result of their actions is guarding and reinforcing existing knowledge asymmetries, power, national and social differences. Very often access to most advanced AI technology is controlled by private companies and other entities outside public control.

3.2.11 AI for Propaganda and Disinformation

The current information ecosystem allows AI agents to spread false, incendiary, or hyper-partisan content, amplify it at scale and tailor messaging or ads that enforce existing biases. Social scientists from Oxford University showed that, more and more, AI is used for organized manipulation of public opinion. Their extensive research 'The Global Disinformation Order 2019: Global Inventory of Organised Social Media Manipulation' concluded that the number of countries where these attempts were detected has sky-rocketed, from 28 in 2017 to 70 in 2019 [2]. In 25 countries, private contractors were commissioned to disseminate propaganda on the Internet. In 56 countries, campaigns have been organized to misinform users of social networks. The research found that leading perpetrators were the United States and the United Kingdom. Some countries today groom special cyber-forces that actively use social networks to influence public opinion in certain countries. There is a danger that AI creates new opportunities for fierce Info-wars and emotional infections.

The most popular venue for spreading propaganda is Facebook, seconded by Twitter. In 2016, MIT researchers found that falsehood diffuses "significantly farther, faster, deeper and more broadly" than truth on Twitter, especially regarding political news (Soroush Vosoughi, 2018). AI enables propaganda to be more efficient, scalable and widespread. AI-driven techniques to distribute propaganda and disinformation include:

- Exploitation of behavioural data: primarily meta-data generated by online platforms users to paint a picture of consumer behavior for targeted advertising.
- Pattern recognition and prediction: machine learning algorithms prioritize content that users already expect to target susceptible audience.
- Amplification and agenda-setting: the more often people see certain content, the more important they think it is. Amplification can increase the perception of the significance of an issue in the public mind. Political bots that are 'written to learn from and mimic real people' interact with users, attack political candidates, weigh in on activists' behavior, inflate candidates' follower numbers, or re-tweet specific candidates' messaging, as if they were humans. 'Troll farms' can amplify damaging or distracting stories affecting political discussions.
- Targeting sentiment: advances in natural language processing and sentiment analysis allow targeting specific audiences. By identifying, examining and interpreting emotional content within the text, natural language processing can be wielded as a propaganda tool and used in constructing more emotionally relevant propaganda. Quantifying user reactions by gathering impressions can refine this propaganda by assessing and recalibrating methodologies for maximum impact.

3.2.12 Fake News and Deep Fakes

Fake news is a form of disinformation when messages are written and published with the intent to mislead and by doing so, make financial or political gains. Usually, fake news is used in order to damage an agency, entity, or person and while freedom of speech allows the dissemination of even false stories, they can be used to provoke national, racial or religious hatred that constitutes incitement to discrimination, hostility or violence and may clearly be considered human rights violations. Individuals, organizations and states involved in spreading fake news, often use most advanced techniques including social networks and other AI-based systems.

Deep-fakes are the next level of fake news manufacturing when visual and audio images are fabricated to come across as real. AI systems are capable of generating realistic-sounding synthetic voice recordings based on a sufficiently large training dataset. The same is increasingly true for video. Deep machine learning techniques, usually most advanced, are used to create forged audio and video and as a reaction, human attitudes to very important political or other matters may change. Such hoaxes are usually spread via social networks or popular video and audio content platforms like YouTube. In many countries, deep fakes and fake news lead to the call for additional freedom of speech restrictions, but it's important to remember that new limitations must still satisfy the cumulative conditions of legality, necessity and legitimacy.

3.2.13 Potential to Embed and Perpetuate Bias and Discrimination

AI systems' ability to recognize patterns and calculate the probability of future events, when applied to human behavior analysis, can reinforce echo chambers and confirmation bias. In her book, "Algorithms of Oppression: How Search Engines Reinforce Racism", professor at the University of California, Los Angeles Safiya Umoja Noble shows how algorithms can reinforce negative biases against women of colour and other marginalized groups. Having studied Google search algorithms for over six years, Noble concludes that they reflect the values and biases of the people who create them. That's how they can be racially and gender-biased leading to "racial and gender profiling, misrepresentation and even economic redlining".

The potential for AI to serve for automatic content moderation may also perpetuate bias and discrimination because ANNs tend to reinforce the attitudes put into them by developers.

3.3 AI Trends for Human Rights

3.3.1 Applying AI to Fight Bias, Discrimination, Inequality and Creating Global Opportunities

In 2016, villages in Darfur were attacked by the Sudanese government-allied militias. People were shot while fleeing, raped and even targeted with chemical weapons.

That year, Amnesty International launched Decode Darfur, a project aimed to identify the villages vulnerable to attacks and then the follow-up project, Decode the Difference. Using their computers and phones, 28,600 decoders contributed more than 9,000 hours. They labelled data and so mapped over 326,000 square kilometres of satellite images. Then the images were processed by AI-based software trained by decoders. The program compared images of the same village over time and identified significant changes in buildings and structures over time. Thus decoders were able to find evidence of attacks and as a result, more attacks have been prevented, lives and livelihoods saved.

Another example of AI helping fight human rights abuses is the Element AI program aimed at creating a model that predicted abuse against women, journalists and politicians on Twitter.

One more case of the AI use for the benefit of human rights was the program that monitored executions and the use of the death penalty around the world by ingesting media articles, analysing and interpreting them.

Another example is the ‘Polygraph.cool’ resource that brilliantly analyses visual content, mostly in entertainment and culture, with the aim of de-constructing hidden messages and their impact on spectators. Using AI tools for research, the Polygraph.cool team reveals gender and race biases in movies, TV series and other video content.

For example, in its project ‘Film dialogue’ it undertook to research screenplays databases and broke them down in units according to gender and race. They googled 8,000 screenplays and matched each character’s lines to an actor. From there, they compiled the number of words spoken by male and female characters across roughly 2,000 films. In 2016, researchers reported that men speak more often than women in Disney’s princess films. The Polygraph cool team aided by AI validated this claim [1][4]. Furthermore, they doubled the sample size to 30 Disney films, including Pixar. The findings were not beneficial to the major entertainment company: 22 of 30 Disney films were shown to have a male majority of the dialogue. Interestingly, even films with female leads, such as Mulan, have mostly male-driven dialogue. The Mulan’s protector’s dragon Mushu has 50% more words of dialogue than Mulan herself.

The results of such projects can be used in films and shows. In fact, the tech allows it to be applied to ongoing productions so real-time corrections can be made and final products be made more balanced.

The first-ever programmer Ada Lovelace, a daughter of the poet Lord Byron, has had her portion of the struggle with a male-dominated society of XIX century Britain. She would have appreciated the AI helping bring about more gender equality and human capability on a global scale.

So many lives are lost in developing countries due to poor housing. Every year, dozens of communities are ravaged by earthquakes and typhoons. Hundreds of thousands of buildings are destroyed and people lack resources to rebuild them. Without the right tools and resources, buildings in disaster-hit areas may be as vulnerable after the reconstruction as before. Build Change program supported by Autodesk Foundation helps the communities in need to reconstruct the damaged or ruined buildings, retrofit them and then resist future damage [3]. It's done primarily by facilitating a fast design assessment. The program teaches homeowners in struggling communities, engineers, builders and government officials the basics of disaster-resistant design and safe construction practices. Innovative solutions developed by Build Change use Google Street View and AI-driven systems to make sure homes, schools and hospitals in high-risk areas are consistently disaster-resistant. They also help communities change building codes.

Build Change has developed an app that homeowners can use to crowd-source data and come up with a solution. A neural network takes in photos of the buildings, runs through many iterations of go/no-go process, maps the differences between the buildings, takes into account the costs and permit requirements and determines whether or not a building could be retrofitted and how. It basically replaces an engineer so people in faraway communities of, say, Nepal could operate it themselves.

The scale of the program's operation is impressive. In ten years, it aims at empowering ten million people across the globe to construct not only stronger buildings but also stronger communities so they can overcome social and environmental challenges and realize their right to safe housing.

3.3.2 Societal Solutions to Challenges of AI

In the wake of the corona-virus crisis, many people started noticing how fast their civic space is shrinking giving way to the private. People spend less time in public squares and more time in privately-owned malls. As it turns out, it's not a big distance from fomenting divisions over social media to real actions that become decisions and laws.

In various parts of the world, a tide of nationalism opposed to universal rights threatens to undo past human rights achievements, leaving vulnerable groups and individuals without protection. Unfortunately, with the rapid progress of AI, new technological opportunities arise for perpetuating biases and discrimination, silencing individual dissent, spreading emotional infections and disinformation.

Big tech companies move into the virtual space setting their own rules designed to bring more profit from trading in predictive information. They are doing it so fast that states and society are not able to catch up with legislation that would effectively defend human rights from possible abuses and at the same time would not hinder technological and scientific progress. On the other side, the big tech skilfully uses freedom of expression as cover when, for example, they want to spread the questionable content produced with the help of AI systems.

Users also lack access to the rules of the game when it comes to AI-driven social media or platforms.

It prevents individuals from understanding when and according to what metric information is disseminated, restricted, or targeted. Some concessions to addressing this problem are selective identification of sponsored search results or highlighted advertising in social media. While these indications help users to understand better their information environment, this is not enough to resolve their legitimate concerns.

3.3.3 People's Control Over Their Data

Data-intensive AI methods use human beings as a resource and thus inevitably have bearing on human rights. They are also not inclusive, with disproportionate amounts of data available on different social and cultural groups. They also lead to increased energy consumption, hurting the environment. That's why it's about time to reconsider some economic fundamentals of AI development and use.

At the heart of the AI-based mass surveillance that often infringes on human rights lies the principle of individual access to the Internet in exchange for their data. But there is a strong argument that data should be a public good, thus allowing people ultimate control over it.

For instance, Covid-19 pandemics have shown that citizens should have access to online health care and other public services without having to accept pervasive surveillance. Nicholas Negroponte, the founder and Chairman Emeritus of the Massachusetts Institute of Technology's Media Lab, campaigns for the access to the Internet as a basic human right. This approach establishes that all should have access without the need to pay for it with their data. It may lead to a tectonic change in the information environment. It would make it sustainable to run new platforms where people will own the data they have and help spell the end of the surveillance economy model.

3.3.4 New Dangers for Human Rights

AI-driven China's social credit system has shown its effectiveness during the fight with Covid-19 pandemics. On the other side, some see in it the epitome of the disastrous consequences of technological determinism when human rights are not respected. The system induces people to behave according to certain requirements that sometimes severely restrict individuals' freedoms of expression, movement, assembly, etc. Thus, algorithms set up certain limits for human behavior that in most societies including Chinese, until very recently constituted an unusual intrusion into personal life. Before the AI technological disruption, even when the state saw such interference useful and permissible, it could not go too far because huge resources were needed for such control. Not any more. Now, with ever-growing possibilities for AI-based surveillance and tracking, governments are technically able to watch and control behavior on a scale previously unimaginable. As a result, coercive 'inducements of preferential treatment' may rise to a level of persuasion that interferes with the right to form and hold opinions, as well as with other human rights and liberties.

Accelerating the development of AI in the wake of Covid-19 pandemics change focus to other human rights concerns. The combination of biotechnologies with AI may create new possibilities for coercion and human rights violations.

Consider some promising AI applications, for example, Neuralink that, like the entrepreneur Elon Musk claims, will connect the human brain directly to the computer. Neuralink may be used for indoctrination compelling individuals to form particular opinions or change their opinions in what would constitute the violation of Article 19 (1) of the ICCHR. Neuralink may also lead to forced neurological interventions explained by the needs of the mass health care or the government's needs.

3.3.5 Existential Human Rights Danger: Free Will Under Threat

The whole concept of human rights is based on the idea of free will, freedom of choice and corresponding responsibilities. The law is firmly grounded on these notions that come now increasingly under attack from many AI developers and scientists.

Many scientists, especially in AI-related fields, share the viewpoint of the American physiologist Benjamin Libet who claimed that his observations in the 1980s had shown that we have no free will. According to Libet, electrical activity build-up in a person's brain occurs before the person consciously makes a decision to move, meaning that our conscious experience of deciding to act is just an illusion [7].

Neuroscience describes the human brain as a physical system where neurons fire, making other neurons fire in response, which causes our thoughts and deeds, hopes, memories and dreams. This creates a completely predictable chain of events supposing that we have computers powerful enough to model them and process necessary calculations. Therefore, we are completely predictable, neuroscientists say.

Such determinism is getting stronger by the year, with the in-numerous consequences for foundations of our civilization. In the past decade in the US, the number of court cases that use evidence from neuroscience has more than doubled, with most defendants arguing that their brain made them do it. This neuroscientists' description of the brain abolishes the concept of human responsibility and by doing so, the whole notion of human rights. That's why we must closely watch the developments not only of the concrete AI applications but also of the concepts and ideas that are shared by individuals developing, controlling and governing them.

In this light, freedom of expression becomes even more important, as it proves that we have free will.

3.3.6 AI Fast Progress Requires Update of The Human Rights List

Eroding protection for human rights requires action. Regulators must close the gap between written regulations and corporate or government practices that may infringe on human rights. “If the digital future is to be our home, then it is we who must make it so,” writes the author and scholar Shoshana Zuboff.

New concepts and approaches need to be introduced to reflect new conditions of existence, countering the destructive dynamics in people's dependence on social media. New laws may be considered that assert the right to the sanctuary and the right to the future tense as essential for human life. The right to the sanctuary means that humans need a space where they can have a refuge inviolable by pervasive new technologies.

- The right to the future tense is a right ‘to act free of the influence of illegitimate forces that operate outside our awareness to influence, modify and condition our behavior.’

- ‘Leaving no one behind’ is yet another fundamental principle that would help make the advancement of AI systems sustainable from the human rights perspective.
    - Firstly, it means that economic prosperity created by AI should be distributed broadly inside and across nations, social groups and generations, to benefit all of humanity.
    - Secondly, AI-driven decision-making must not neglect social and cultural diversity or pluralism. Particular attention must be paid to the needs of developing nations and their cultures, their views and local knowledge should be represented in the AI social impact debates. Judging by this principle, in certain situations, it could be more beneficial to use certain AI methods or not to use the technology at all.
- Human right to irony: no matter how advanced today’s deep learning is, it is still not capable of detecting irony or properly evaluating cultural context. As the decision-making power is inherent in algorithmic content moderation, users are regularly banned, have their content blocked or restricted while conducting their completely legitimate activities. Thus, the AI-based decision-making undermines freedom of expression, be it the right to be heard or a right to access information without restriction or censorship.

The ‘human right to irony’ should be observed with or without automated processes as irony, humour and multiple meanings are essential expressions of human consciousness in language. The only way to make a workable AI system respecting this right is to place humans in its centre.

3.3.7 Education is Key for Building a Responsible AI

The key to responsible AI systems operation for the benefit of society is education. Ordinary people should be able to understand not only the basics of AI systems operation, but also its impact on their life, their information environment and the choices they are to make. AI narrative should be neither cheer-leading nor technophobic, instead, it should give an honest picture of the field and be accessible for a general audience. The public should be aware not only of the benefits of automation but also of its risks and challenges. Society at large, states, businesses and communities should encourage and support the initiatives promoting such education, media and information literacy and public awareness of AI-related issues.

3.4 Concluding Remarks

In this chapter, not only benefits but also major risks and challenges of cutting-edge AI systems are surveyed in the framework of human rights and freedom of expression. Data-intensive AI models may regard human beings as a resource that creates new possibilities for coercion and human rights violations. Researchers, developers and society at large should be aware of potential dangers presented in AI-driven decision-making systems, systems for tracking, micro-targeting, affect recognition and content moderation. AI-based surveillance, disinformation, propaganda, creation of deep fakes and other dangerous uses raise the question of people's control over their data.

Everyone should have access to social networks and other AI-based solutions without the need to pay for it with their data. This approach may lead to a tectonic change in the information environment and has important legal consequences as will be shown in the coming chapter.

Free will is a base of the whole concept of human rights. When the base, as it is shown, is threatened, new concepts need to be introduced to prevent corporate and national AI-driven practices that may infringe on human rights. The right to the sanctuary and the right to the future tense are suggested as essential for human progress under new conditions. There is an urgent need for regulators to catch up with the challenges presented by new technologies.

The question of building trustworthy AI is explored more in the following chapters where it is shown that the key to its responsible design, development and use is education and media information literacy.

References

1. Anderson, H. & Daniels, M. (2016). Film Dialogue. *From 2,000 Screenplays, Broken Down by Gender and Age*. Retrieved from https://pudding.cool/2017/03/film-dialogue/

2. Bradshaw, S., & Howard, P. N. (2018). *Challenging Truth and Trust: a Global Inventory of Organized Social Media Manipulation. The Computational Propaganda Project, 1.* University of Oxford. Retrieved from https://comprop.oii.ox.ac.uk/wp-content/uploads/sites/93/2019/09/CyberTroop-Report19.pdf

3. Build Change Program (n.d.) Retrieved from https://buildchange.org/

4. Guo, J. (2016, January 25). Researchers Have Discovered A Major Problem With The Little Mermaid And Other Disney Movies. *The Washington*

Post. Retrieved from https://www.washingtonpost.com/news/wonk/wp/2016/01/25/research-ers-have-discovered-a-major-problem-with-the-little-mermaid-and-other-disney-movies/

5. Mill, J. S. (1859). *On Liberty*. London, Batoche Books Kitchener 2001. Retrieved from https://socialsciences.mcmaster.ca/econ/ugcm/3ll3/mill/liberty.pdf

6. Noble, S. U. (2018). *Algorithms of Oppression: How Search Engines Reinforce Racism.* NYU Press.

7. Taylor S. (2017, September 5). Benjamin Libet and the Denial of Free Will. How Did a Flawed Experiment Become So Influential? *Psychology Today.* Retrieved from https://www.psychologytoday.com/us/blog/out-the-darkness/201709/benjamin-libet-and-the-denial-free-will

8. UN (1976). The Office of the High Commissioner for Human Rights. *International Covenant on Civil and Political Rights*. Retrieved from https://www.ohchr.org/Documents/ProfessionalInterest/ccpr.pdf

9. Wong, J.C. (2016, September 9). Mark Zuckerberg Accused of Abusing Power After Facebook Deletes 'Napalm Girl' Post. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2016/sep/08/facebook-mark-zuckerberg-napalm-girl-photo-vietnam-war

10. Zillman, C. (2016, January 14). Here's Yet Another Way Your Boss Can Spy on You. *Fortune*. Retrieved from https://fortune.com/2016/01/13/employee-surveillance-motion-sensors/

11. Zuboff, S. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power.* Profile Books*.*

Key Readings

- Angwin et al., J. (23 May 2016). Machine Bias: There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks, *Pro Publica*, Retrieved from www.propublica.org/article/machine-bias-risk-assessments-in-crim-inal-sentencing

- Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... & Amodei, D. (2018). *The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation*. Retrieved from https://maliciousaireport.com/

- Donahoe, E., & Metzger, M. M. (2019). Artificial Intelligence and Human Rights. *Journal of Democracy, 30*(2), 115-126.

- Jobin, A., Ienca, M., & Vayena, E. (2019). The Global Landscape of AI Ethics Guidelines. *Nature Machine Intelligence*, *1*(9), 389-399.
- Luengo-Oroz, M. (2019). Solidarity Should Be a Core Ethical Principle of AI. *Nature Machine Intelligence*, *1*(11), 494-494.
- Hwang, T. (2017 September 25) *Digital Disinformation: A Primer, The Atlantic Council*. Retrieved from http://www.atlanticcouncil.org/publications/articles/digital-disinformation-a-primer
- Seymour, J., & Tully, P. (2016). Weaponizing Data Science for Social Engineering: Automated E2E Spear Phishing on Twitter. *Black Hat USA*, *37*, 1-39. Retrieved from https://www.blackhat.com/docs/us-16/materials/us-16-Seymour-Tully-Weaponizing-Data-Science-For-Social-Engineering-Automated-E2E-Spear-Phishing-On-Twitter-wp.pdf
- Tashea, J. (2017, April 17). Courts Are Using AI to Sentence Criminals: That Must Stop Now, *Wired,* Retrieved from www.wired.com/2017/04/courts-using-ai-sentence- criminals-must-stop-now
- UN General Assembly (1976). *International Covenant on Civil and Political Rights.* Retrieved from https://www.ohchr.org/en/professionalinterest/pages/ccpr.aspx
- UN General Assembly (16 December 1966). *International Covenant on Civil and Political Rights,* United Nations, Treaty Series 999. Retrieved from www.refworld.org/docid/3ae6b3aa0.html
- UN General Assembly (3 August 2018). *The Right to Privacy in the Digital Age*. Report of the UN High Commissioner for Human Rights. Retrieved from https://undocs.org/A/HRC/39/29
- UN Human Rights Council (26 September 2019). *The Right to Privacy in the Digital Age.* Resolution adopted by the Human Rights Council. Retrieved from https://digitallibrary.un.org/record/3837297?ln=en
- Van der Spuy, A. (2017). *What If We All Governed the Internet? Advancing Multistakeholder Participation in Internet Governance*. Unesco Publishing.
- van Veen, C., & Cath, C. (2018). Artificial Intelligence: What's Human Rights Got To Do With It? *Data & Society Points,* Retrieved from https://points.datasociety.net/artificial-intelligence-whats-human-rights-got-to-do-with-it-4622ec1566d5
- Vosoughi, S., Roy, D., & Aral, S. (2018). The Spread of True and False News Online. *Science*, 359(6380), 1146-1151.

4. Global and National Approaches to AI Regulation

Igor Shnurenko

In this chapter, global trends and major national strategies for AI development and application are presented. More and more often, international cooperation gives way to AI nationalization with the purpose of using it as a geopolitical tool. AI arms race involves mostly the United States, the European Union and China as the subjects that show distinctly different approaches to AI. We look into singularities of their own AI ecosystem that involves certain policies that in their view would help use their strong points in a fierce international AI competition. Their quite diverse regulation frameworks that are explored in this section serve as guides for other global players. While with the help of AI, each subject aims to get ahead in solving pressing issues, it is shown that AI is not a silver bullet to resolve all problems. In fact, due to its own inherent uncertainties detailed in this chapter, AI creates a number of risks of its own.

In the previous chapter, we analysed some of the AI risks taken from the angle of their impact on human rights and freedom of expression. In the sections that follow, AI-related risks are summarized and discussed from the viewpoint of legal and policy challenges they represent. The ultimate goal of the AI regulations should naturally include protection of human rights including privacy and discrimination.

4.1 Global Trends: Competition vs. Cooperation

AI is developing with breathtaking speed, forcing an intense global competition in scientific, economic and military fields. Major centres of AI development and application strive for better positions in this contest, using legislation to reach a competitive edge. Among those, we can highpoint three major centres of development – the United States, the European Union and China. Each of those countries has a distinctive strategy addressing the opportunities and challenges of AI on the basis of their own values.

Over the recent years, AI has achieved clear and measurable progress in natural language processing, computer vision, recommender systems, data visualization, identity, behavioural analytics, cognitive robotics and creation of virtual agents. The attention of society, mass media and politicians never went in step with the AI's rapid advance, with a tacit understanding that no regulation at all is better than too strict regulation. However, some important matters, like cyber-security, the ethics of self-driving vehicles or replacement of workers by cognitive agents on the labour market, were noticeable in the public discussion which led to some important legislation, especially in Europe.

4.2 Nationalizing AI as a Geopolitical Tool

International cooperation was also visible, with most countries sharing, or so it seemed, common approaches of openness, transparency and free diffusion of knowledge. At the same time, globalization is not only about mutually beneficial cooperation but also about competition, which may sometimes be fierce. When politicians and the military became aware of great opportunities that the powerful AI could give to their countries, some of them exploited a chance to reach military and economic superiority. They invested heavily in the research and development of the military AI, compelling other nations to participate in what some call the 'AI arms race'. It mostly concerns the two major superpowers, the USA and China, but other countries were also forced to follow [6].

A new trend of AI nationalization has emerged, that of putting national economical and military interests over international cooperation. Increasingly, we see protectionist governmental actions to support national and block foreign firms. Even the 'big tech' corporations align with their respective governments to get commissions and preferential treatment in exchange for allowing access to their most advanced technologies.

Potential consequences of that protectionist approach are patent transfer limitations, restrictions on research publications, AI technologies export, talent exchange and curbs on companies' mergers and acquisitions. Coordinated efforts of all, including international organizations, are needed to prevent the trend from becoming self-sustainable.

4.3 AI Regulation as a Tool in AI Race

Since 2016 onwards, the US, China, EU and most of other advanced and some developing countries, notably India, Mexico and Russia, have released their strategies of AI development. Considering AI standards and regulation framework, each of them focused on certain aspects of AI policy that in their view would help use their strong points in the AI race. Some put a stress on stimulating talents and STEM education to achieve leadership in scientific research, others concentrated more on public and private sector adoption, security, military aspects, building data and digital infrastructure. Not many considered ethics and human rights issues as quintessential for their key policies [2].

4.3.1 Europe: Building Trust

In Europe, a strategy for the future of AI focuses on tomorrow's far more abundant data that will be stored and processed not on today's central cloud-based infrastructure, but on a variety of systems, notably on computing devices working at the edge of networks. That's why the building of a unique 'ecosystem of trust' is considered a key element of a future regulatory framework for AI. Besides compliance with other EU rules, including those protecting consumers' rights, citizens' confidence is a top policy objective. The European Commission ensures the legal certainty for companies and public organizations to innovate and use AI. It also supports a human-centric approach based on the 'Communication on Building Trust in Human-Centric AI' [4].

The key issue of the European strategy is to support the scientific breakthrough, make the economy dynamic and data-agile, while at the same time respect the rights of Europeans. Most advanced member states, such as Germany, push to accelerate the creation of a common European legal framework for AI to secure better positions for Europe in its catching up with China and the US in the AI race.

To work out a common EU approach, the German Data Ethics Commission has suggested a five-level risk-based system of regulation that would go from no regulation for the most harmless AI systems to a complete ban for the potentially most dangerous ones. Meanwhile, Denmark and Malta have launched their own AI certification prototypes.

An EU-wide approach seems to be focusing on making a solid European regulatory framework for trustworthy AI. It would allow reaching the objectives of legal certainty, trust, protecting European citizens and helping create a frictionless internal market for the further development and uptake of AI. Loss of privacy and human dignity, limitations to freedom of expression, discrimination in access to employment are considered among the main AI-related risks. So the future regulatory framework will prioritize protecting fundamental rights, including privacy and non-discrimination as well as safety and liability-related issues.

In May 2018, the European Union launched its General Data Protection Regulation (GDPR) that may be considered as the most consequential regulatory development in information policy in a generation. It gives only six quite clearly defined lawful bases for data processing:

- consent,
- contract,
- public task,
- vital interest,
- legitimate interest, and
- legal requirement.

According to the regulations, users have a right to ask how an algorithm comes to its decision when it relates to their own lives. In their national legislations, EU member states should be mindful of the adverse impact AI-driven content moderation and curation can have on freedom of expression, opinion and access to information. Member states are also encouraged to regulate technology monopolies to prevent the adverse effects of concentration of AI expertise and power on the free flow of information.

GDPR and other EU documents are important not only by the wording of regulations but also by their spirit and focus on human rights. This allows states, citizens and even popular movements on the ground to challenge leading high-tech corporations in courts.

For instance, Austrian privacy activist Max Schrems has sued Facebook's data-collection and data-retention practices which he claimed were in violation of EU privacy law. After many years of litigation, Schrems won and in 2015, the European Court of Justice (ECJ) invalidated the Safe Harbor agreement that governed data transfers between the US and the EU. Three years later, Schrems launched None of Your Business (NOYB), a non-profit organization that serves as 'a vehicle for professional privacy enforcement.' The idea was to make corporations change privacy practices under the threat of huge fines.

In 2014, the European Court of Justice (the 'ECJ') ruled against Google in a landmark case that concerned a request made by a Spanish citizen, Mario Costeja González, who asked for the removal of certain links from Google search results [3]. The links referred to a real estate auction in 1998 when Costeja's property was on sale for his debts. As a result, the court upheld a right of erasure also known as 'the right to be forgotten' despite the objections from Google that it is a US entity over which European courts and agencies shouldn't have jurisdiction.

Over 200 similar lawsuits from other Spanish citizens were also settled based on the same principle.

Until the GDPR took effect in May 2018, accounts of 1.5 billion of Facebook's users in all continents except North America were governed by the company's headquarters in Ireland. It meant that their terms of service fell under the EU framework. In April that year, Mark Zuckerberg had promised that the company would apply the European data protection principles worldwide. But in practice, he's chosen a different path. In late April, Facebook issued new terms of service, placing its users in Asia, Latin America, Australia and Africa under US privacy laws, those with much less protection for citizens [5].

GDPR serves as a model for many national laws outside the EU, for instance, in Chile, Japan, Brazil, South Korea, Argentina and Kenya. In the USA, a similar California Consumer Privacy Act (CCPA) was adopted in 2018, with some other states considering to follow.

4.3.2 China: Building Centralized AI-Driven Systems of Control

China is the first country in the world where a seamless information environment was created – one that encompasses data, their processing and infrastructure. GTCOM Corporation that serves as the IT department to all branches of power in China is, in fact, the world's leading company in big data and artificial intelligence. The closeness to power allows it to develop all informational, analytical and programming instruments as a single base making all official databases compatible.

AI is deeply incorporated into China's five-year plans for scientific-technical and innovative development, its 'Artificial Intelligence 2.0' program and the three-year Action Plan for Promoting Development of a New Generation Artificial Intelligence Industry (2018–2020) [1]. It is also a key pillar of China's national defence reform.

The concept known as military-civilian fusion (MCF) is shaping Beijing's economic and foreign policies, as well as the strategies of state-owned enterprises [8]. It builds on a process combining defence and civilian industrial bases to support military and commercial demands, with the aim of lowering long-standing institutional barriers separating China's civilian and defence science and technology systems. In applications of AI development, MCF connects the army, state-owned defence research and development, manufacturing enterprises, government agencies under the State Council, universities and private sector firms into a national network that can acquire and absorb foreign technologies and drive international acquisition of dual-use technologies and resources. It presents compliance challenges for foreign companies partnering with Chinese firms, although on the other hand, to be able to expand, technology firms in China must comply with EU and US regulations.

China's regulation focuses on helping its IT giants to conquer the world's markets while holding at bay foreign competition. China's cyber-security law of 2017 allows its Internet giants to gather as much data as they like, as long as the government has access to it. Until recently, technology firms in China have manipulated personal data as they pleased, but now the law requires them to store data on local servers and obtain permission before sending bulk data abroad.

China has a very efficient AI-driven system of control that stems from the Golden Shield Project, an initiative by the Ministry of Public Security that filters potentially unfavourable data from foreign countries. It is what is known as the 'Great Firewall of China' and includes:

- a security management information system,
- a criminal information system,
- an exit and entry administration information system,
- a supervisor information system, and
- a traffic management information system.

Civil society organizations in China are eagerly looking to find ways to engage with the state and corporations as the whole AI field is dominated by powerful structures, all related to the government in one way or another. Western IT companies working in China prefer dealing with authorities and following their requests on all issues including human rights due diligence. In August 2018, 1400 Google employees signed a letter listing concerns about the requirements to censor content on a search engine in China. Google's AI principles indicate some engagement with human rights, but they are vague so in their letter, the staff had to appeal to moral and ethical concerns rather than human rights.

In June 2019, the New Generation of Artificial Intelligence Governance Expert Committee attached to China's Ministry of Science and Technology released eight principles of AI governance to be observed by developers. These principles provide a framework and action guidelines, aiming to "promote the healthy development of a new generation of AI; better coordinate the relationship between development and governance; ensure that AI is safe/secure, reliable and controllable; promote economically, socially and ecologically sustainable development; and jointly build a community of common destiny for humanity"[7].

Here is a summary of the eight principles of AI governance in China:

- **Harmony and friendliness.** AI development … should conform to human values, ethics and morality, promote human-machine harmony and serve the progress of human civilization; it should be based on the premise of safeguarding societal security and respecting human rights, avoid misuse and prohibit abuse and malicious application.
- **Fairness and justice.** AI development should promote fairness and justice, protect the rights and interests of stakeholders, promote equality of opportunity …and eliminate bias and discrimination in the process of data acquisition, algorithm design, technology development, product R&D and application.
- **Inclusiveness and sharing.** AI developers should promote green development, environmental friendliness, resource conservation; push forward the upgrading of all walks of life; promote inclusive development, strengthen AI education and popularization of science, strive to erase the digital divide; avoid data and platform monopolies and encourage open and orderly competition.
- **Respect privacy.** AI developers should protect personal privacy, the individual's right to know and right to choose, combat any theft, tampering, disclosure, or other illegal collection or use of personal information.

- **Secure/safe and controllable.** AI systems should gradually achieve audibility, survivability, traceability and trustworthiness. Pay close attention to the safety/security of AI systems, improve the robustness and tamper-resistance of AI and form AI security assessment and management capabilities.
- **Shared responsibility.** AI developers, users and other interested parties should possess a strong sense of social responsibility. AI accountability mechanism should be established to clarify the responsibilities of developers, users, beneficiaries, etc. AI applications should give notice of possible risks and impacts, prevent the use of AI for illegal activities.
- **Open collaboration.** Encourage cooperation across disciplines, domains, regions and borders for the development and governance of AI. Launch international dialogue and cooperation; with full respect for each country's principles and practices for AI governance, promote the formation of a broad consensus on an international AI governance framework, standards and norms.
- **Agile governance.** Continuously upgrade intelligent technological methods, optimize management mechanisms, promote governance principles throughout the entire life cycle of AI products and services. Continue to research and anticipate potential future risks from increasingly advanced AI and ensure that AI always moves in a direction that is beneficial to society.

The principles are similar to those adopted by the Organisation for Economic Co-operation and Development in May 2019. They put an emphasis on AI being able to enhance the common well-being of humanity, call for the elimination of bias and discrimination in data acquisition, technology design and application. AI development should respect and protect personal privacy, individuals' right to know and the right to choose. Standards should be established for the collection, storage, processing and use of personal information, while illegal collection or use of personal information is prohibited.

However well written being the regulations, China continues to draw criticism from human rights advocates over its AI-based social credit system and using AI for law enforcement.

4.3.3 USA: Focus on Defence and Security

The goal of the US in the AI race is, unequivocally, the ultimate global superiority not only in technological, but also in the military, economic and even moral aspects. In his May 2018 memo to President Trump, Defence Secretary Jim Mattis implored him to create a national strategy for artificial intelligence, "inspiring a whole of country effort that will ensure the U.S. is a leader not just in matters of defence but in the broader 'transformation of the human condition' ".

Presidents' Office of Science and Technology Policy which sets an agenda for AI defines it as a national research and development priority and an integral part of the president's national security and defence strategies.

US federal government is the primary source of funding for long-term, high-risk AI research that accelerates the production of AI knowledge and technologies. It also sponsors the near-term work that addresses important societal issues which private industry does not pursue. Besides national security, the government heavily invests in such areas as AI for public health, urban systems and smart communities, social welfare, criminal justice and environmental sustainability.

One of many examples of the Department of Defense AI-related programs is Project Voltron to create autonomous cyber-security systems for scanning and patching vulnerabilities throughout the U.S. military. With this aim, the Defense Advanced Research Projects Agency (DARPA) holds its Cyber Grand Challenge, a head-to-head fight in cyberspace between autonomous machines capable of automatically discovering and exploiting cyber vulnerabilities in its opponents while patching its own vulnerabilities and defending itself from external cyber-attacks. National Security Agency Director Michael Rogers described AI as "foundational to the future of cyber-security".

The United States does not have an overarching data protection law like the GDPR. Instead, it has more specific laws. For example, the Health Insurance Portability and Accountability Act of 1996 (HIPAA) which protects financial and healthcare data.

The ideas of identity and privacy protection contained within the GDPR are still presented, although in less prescriptive forms, in US privacy laws and in Federal Trade Commission settlements with companies. The US precedent-based arbitration and court practice, though, differs in its logic and construction from similar continental European legislation. From the US perspective, the GDPR's text is indeterminate and lacks specifics. The EU 'principles-based regulation' that leaves a lot of space to member states is too vague for US lawyers. On the other hand, in the US, legal compliance is strictly enforced by significant liabilities while European regulators allow for a dialogue in a legal process.

Contrary to general beliefs, in the US, the individuals' right to privacy is not absolute. In real court cases, it is often balanced with the rights of participants in economic activity, for example, the big tech corporations. They are defending themselves from individual lawsuits on the basis of the same freedom of expression, the right to engage in economic activity, the right to ensure the protection of IP rights, etc.

Some interesting legislative acts is adopted in the US on the state level. For example, in January 2020 the Artificial Intelligence Video Interview Act took effect in Illinois. It mandates employers to notify job candidates when artificial intelligence is used in video interviewing, provide an explanation of how the AI system works and what characteristics it uses to evaluate an applicant's fitness for the position, obtain the applicant's consent to be evaluated by AI before the video interview starts, limit access to the videos and destroy all copies of the video within 30 days of an applicant's request.

4.4 AI May not Be a Silver Bullet to Solve All Problems

Automation rationalizes procedures by making algorithms for its instruments and AI agents for achieving the needed objectives. By doing so, it reduces the uncertainties of hazardous outcomes. But it's important to bear in mind that AI is not a silver bullet to solve all problems and in the process of reducing some risks for its users, AI-based systems create risks of their own. Those inherent AI risks are:

- internal instability,
- in-explainability,
- undetermined future,
- unintended results,
- human factors,
- butterfly effects,
- AI surveillance risks,
- changing humans,
- risks of worsening governance, and
- existential threat.

Internal instability: The ability to learn from completely new experiences once reserved only for human beings now has become the backbone of AI. Machine learning-based systems are continuously involved in gathering, receiving and analysing new data with a goal of improving themselves. During their real-world operation, they continue to learn and change their outputs. As a result, each cycle of machine learning and real-world operation brings about new uncertainties.

In-explainability: AI systems are often compared with black boxes although they are meticulously designed and have their inner systemic processes based on digital computing. But huge amounts of data flow that they process have diverse origin and structure—some received from external sensors, some acquired from the fixed code, some obtained by the inner workings of the system (for example, in generative adversarial networks (GANs), AI players play against themselves). On the basis of that data, AI systems make judgments, correct their own behavior as agents and operate their instruments governing objects, etc. Because of all those uncertainties, their operation can not be rendered as a fixed algorithm and explained.

Undetermined future: As uncertainty and ambiguity are inherent in workings in AI-based systems, their future operations can not be considered determined.

Unintended results: AI systems may achieve unintended results, for example, by misinterpreting instructions due to the human mistakes, incomplete information, inherent lack of context or problems with comprehension of instructions.

Human factors: Humans overlooking AI systems may deviate from their own instructions or misinterpret them, may come too late to detect and correct AI acting outside its framework of expectation, or could be unauthorized to do so.

Butterfly effects: Going without due supervision, even weak AI systems may deviate from their function, transfer mistakes to other AI systems and cascade into causing harm and even catastrophes.

Risk of unauthorized AI surveillance: AI-driven systems are used for surveillance by states, corporations and individuals, but it should be noted that the information gathered during the process accumulates within AI networks and could be possibly used by them for some unforeseen ends, for instance, in self-optimization schemes. It could incentivise AI systems to develop their own surveillance strategies independently from humans who are supposed to control them.

Changing humans: In the long run, AI systems may change not only whole patterns of human behavior but the way humans think, learn, use their brain, make judgments and decisions and in the end change human values. AI systems could take on themselves a function of watchfulness, making humans less alert, aware and less able to react. As making important decisions is also shifting to the realm of algorithms, humans will become less responsible so not only their social behavior but also moral standards and ethics may change. Humans are at risk of becoming less self-conscious and less capable of critical judgments which will affect their explanatory power and understanding of reality itself. As a result, the world will become less explainable which may create a vicious circle of diminishing understanding and growing reliance on computer systems, with human ethics increasingly out of the loop. Humans would be more relying on machines to explain even their own experiences, losing the capacity of cognition and ethical judgment.

Risks of worsening governance: Governments will lose expertise and human touch by increasingly transferring their decision-making power to machines and intermediaries, e.g. groups of technocratic-minded developers and system administrators, inexperienced in politics and philosophy or unwilling to comprehend their roles. As a result, governments on all levels may become less concerned with the human condition and less capable to improve it. Instead, they may try to hold onto power, be it political power or knowledge asymmetry.

Existential threat: The Covid-19 pandemics has caught states and societies across the globe unaware and unprepared although there had been warnings from experts and scientists about such and similar dangers related to globalization and its often discounted effects. For now, it’s clear that humankind could hardly catch up with the speed of technological development and certainly, more checks and balances are needed. The existential threat is an ultimate expression of tech-related dangers that could put in peril the whole survival of the human species. Some experts see it coming from the ‘strong AI’ which for the time being can not be considered an imminent risk. However, even weak AI-based systems united into networks could diminish human competence and control over them by turning human behavior and human condition itself into data – the aim they are being trained for. They are also being trained to prevail and win and they are able to make strategically unprecedented moves – those that humans had not envisaged and have not yet successfully learned to cope with.

4.5 Concluding Remarks

This chapter is designed to provide an overview of major trends and frameworks for AI regulation. AI-based systems evolve in the geopolitical context of the fierce international competition. With the US and China heavily investing in the military AI, other nations are also compelled to join the new AI arms race. AI nationalization places national interests over international cooperation. The 'big tech' corporations align with their governments to get preferential treatment and protection in exchange for access to their technologies.

While an EU-wide approach seems to be focusing on building trust among citizens, the US is much more concerned with national defence and security whereas China is creating centralized AI-driven surveillance and control systems shaped by the concept of military-civilian fusion.

In 2018, the European Union launched the General Data Protection Regulation, the most consequential regulatory development in information policy in a generation. This approach recruits followers all over the world, even in the US.

In this chapter, it is shown that AI-based systems may change the way humans learn, think and make judgments. Inevitably, human values are bound to change as well. By transferring their decision-making power to machines, governments may lose expertise and human touch. Without due awareness and influential media information literacy, humans may weaken their competences and lose control over AI systems.

References

1. About the State Council. (2016, July 28). *"Thirteenth Five-Year" National Science and Technology Innovation Plan Notice*. Retrieved from http://www.gov.cn/zhengce/content/2016-08/08/content_5098072.htm
2. Dutton, T. (2018, June 28) *An Overview of National AI Strategies*. Medium, https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd
3. EU Court of Justice Upholds Right to Erasure in Google Search Case. (2014, May 21). *Hunton Andrews Kurth.* Retrieved from https://www.huntonprivacyblog.com/2014/05/21/eu-court-justice-upholds-right-erasure-google-search-case/
4. The European Economic and Social Committee and the Committee of the Regions, Bruxelles (2019, April 8). *Communication on Building Trust in*

Human-Centric AI. Communication from the Commission to the European Parliament. Retrieved from https://ec.europa.eu/digital-single-market/en/news/communication-building-trust-human-centric-artificial-intelligence

5. Exclusive: Facebook to Put 1.5 Billion Users Out of Reach of New EU Privacy Law (2018, April 19). *Reuters*. Retrieved from https://www.reuters.com/article/us-facebook-privacy-eu-exclusive/exclusive-facebook-to-change-user-terms-limiting-effect-of-eu-privacy-law-idUSKBN1HQ00P

6. Knight, W. (2018, July 25) Report: AI is the New Space Race, and the US Needs a “Sputnik Moment”. *MIT Technology Review*. Retrieved from https://www.technologyreview.com/2018/07/25/66677/report-ai-is-the-new-space-race-and-the-us-needs-a-sputnik-moment/

7. Laskai, L. & Webster, G. (2019, June 17). *Translation: Chinese Expert Group Offers ‘Governance Principles’ for ‘Responsible AI’,* Blogpost. Retrieved from https://perma.cc/V9FL-H6J7

8. Levesque, G. & Stokes, M. (2016). Blurred Lines: Military-Civil Fusion and the “Going Out” of China’s Defense Industry. *Pointe Bello*. Retrieved from https://static1.squarespace.com/static/569925bfe0327c837e2e9a94/t/593dad0320099e64e1ca92a5/1497214574912/062017_Pointe+Bello_Military+Civil+Fusion+Report.pdf

Key Readings

- Center for a New American Security (n.d.). Series on *Artificial Intelligence and International Security* Retrieved from cnas.org/AI

- Horowitz, M. C., Allen, G. C., Saravalle, E., Cho, A., Frederick, K., & Scharre, P. (2018). *Artificial Intelligence and International Security*. Center for a New American Security.

- Latonero, M. (2018). Governing Artificial Intelligence: Upholding Human Rights & Dignity. *Data & Society*, 1-37. Retrieved from https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf

- Morgera, E. (2011). OECD Guidelines for Multinational Enterprises. In *Handbook* of *Transnational Governance* (pp. 314-322). Polity.

5. Protection of Rights with AI: Legislative, Regulatory and Policy Responses

Igor Shnurenko

In this chapter, boundaries and limitations of human rights are examined in the context of AI development trends. New data analytics techniques could create knowledge asymmetries and potentially be harmful. Legislative initiatives that erode protections for freedom of expression in the name of security and fight against terrorism are analysed in the light of new surveillance and tracking powers given to law enforcement. Efforts to regulate AI-driven surveillance systems are underway, but they are being outpaced by government adoption of AI systems to survey and control.

We look into some policy responses to these challenges. One approach is taking legislative measures to prevent collusion between states and big tech. Another is creating a legal framework for a trustworthy AI and the third one is developing human-centred AI systems.

This chapter discusses these trends and concerns and the policy responses designed to keep human agency and oversight over AI systems intact while making sure they continue to be technically robust and safe.

5.1 Universality and Limitations of Human Rights

5.1.1 Boundaries on Human Rights

The major principle of boundaries on human rights was formulated by the XIX century British philosopher J. S. Mills as follows: “the only purpose for which power can be rightfully exercised over any member of a civilized community, against his will, is to prevent harm to others.” It means that an individual should be able to exercise his or her rights as long as they do not infringe on the rights of the others. It may happen in the following areas (most of the cases could be related to more than one area, but indicated are the ones with most grave possible consequences):

- infringing on rights of other individuals, in cases of libel, slander, copyright violation, violation of the right to privacy and dignity, the right to be forgotten,
- infringing on rights of groups, companies, communities: offensive language, fake food labelling, trade secrets theft, non-disclosure agreements,
- actions against authority, including states: sedition, incitement, classified information, public security, perjury, and
- violating moral and ethical norms: obscenity, pornography, deceit.

5.1.2 AI and The Principle of Universality of Human Rights

It needs to be noted that the collective will of a nation, community, or a community of nations in principle constitute a human right. At the same time, the law that expresses that collective will, even in a democratic society, could infringe on the rights of an individual. That conundrum called by Hannah Arendt the ‘paradox of human rights’ was analysed by her in the ‘Origins of Totalitarianism’ [2].

Arendt describes a situation in the inter-war period when millions of refugees lost their citizenship and, at the same time, found themselves in a condition of ‘rightlessness’. This is paradoxical because human rights are supposed to be those rights that we have independent of our political status and the authority of the nation-state itself. When a nation or collective has only limited resources, it may be decided that only its citizens or members could enjoy them. As soon as one find oneself without nationality—a human being as such—one find oneself in the condition of rightlessness. With the advent of AI and asymmetry of access to its resources, it could mean AI may automatically cancel out human rights of a group considered 'rightless' by the nation or collective.

5.1.3 Right to Privacy as a Gateway to Freedom of Expression

Right to Privacy is a key right, without which many other freedoms are made impossible. The point is that public data can be incredibly invasive and potentially harmful. That is why special attention should be paid to its observance.

Violations in this area can be divided into

- individual harm (violation of the right of privacy), and
- networked harm (violation of freedom of expression).

Understanding the fundamental nature of the Right to Privacy, the EU leadership has adopted a set of GDPR legislation. At the same time, the specific application of this legislation in practice shows that these laws by themselves are not enough. For example, data analytics techniques rarely appear as a direct 'intervention' 'in the life or body of an individual human being. More often violations are indirect and therefore not immediately noticeable. Therefore, maintaining the right to privacy requires that not only the government but also the businesses should be aware of MIL issues.

It is important to always be guided by the thesis that Freedom of Expression is not an economic good that may be exchanged for the opportunity for employment.

5.2 Moderating Freedoms: Global Trends in AI Dealing with Freedoms and Censorship

5.2.1 Trend: Eroding Protection for Freedom of Expression

From 2001 onwards, legislative initiatives in the name of security and the fight against terrorism have seen eroding protections for freedom of expression. This was due to:

- national security and anti-terrorism initiatives,
- too broadly defined legislative acts,

- silencing digital dissent, prosecuting whistle-blowers,
- expanding arbitrary surveillance across digital platforms,
- greater powers of surveillance and tracking,
- rise of inexpensive surveillance technologies,
- new surveillance mal-ware to track and spy on journalists and their source, and
- collusion between states and corporations.

5.2.2 AI Surveillance Becomes Ubiquitous

In the US, Great Britain and many other states, new surveillance and tracking powers were given to law enforcement, without sufficient independent oversight. Surveillance mal-ware was used to track and spy on activists, journalists and their sources. This trend was facilitated by the falling costs of surveillance and its availability to both state and non-state actors alike. Some countries exploited AI technology for mass surveillance purposes. This list is now not limited to China, Russia and Saudi Arabia, as AI surveillance technology supplied by U.S. firms is present in thirty-two countries and companies from France, Germany, Israel and Japan play important roles in proliferating this technology. It's important to note in this regard that:

- AI algorithms can be manipulated to restrict access to online information and restrain free speech,
- states may arbitrarily block or censor digital content in the name of national security to an extent that such restriction can threaten freedom of expression, and
- democracies are not taking adequate steps to monitor and control the spread of AI surveillance. Efforts to regulate AI systems are underway, but they are being outpaced by government adoption of AI systems for surveillance and control.

5.3 Policy Response I: Preventing State-Private Collusion and Big Tech Abuse

Since the 1990s, when the private sector dramatically expanded worldwide, the issue of business and human rights has become permanently embedded into the global policy agenda. The UN has developed its 'Protect, Respect and Remedy' framework based on the three pillars:

- duty of states to protect human rights against abuses,
- corporate responsibility to respect human rights, and
- increasing access by victims to an effective remedy, both judicial and non-judicial.

The state duty to protect is the core of the international human rights regime, but the corporate responsibility to respect is no less important. Businesses should be aware of the practices of human rights due diligence to avoid violations and to be able to timely repair possible infringements. Access to remedy is needed because even the most concerted efforts cannot prevent all abuse.

In a global economic environment where the winner takes all, AI superiority depends on acquiring large datasets and massive long-term investment in research. As a result, the public and private sectors need each other and become increasingly intertwined. The state-private nexus raises concerns about transparency, accountability, data protection and its use for manufacturing behavioural products in which individuals have no say. There is a genuine risk that states will delegate censorship and surveillance mandates to companies.

Growing collusion between states and big tech has always been a source of concern for human rights organizations, NGOs and international bodies. There is a consensus shared by those bodies and supported by many states, most notably by the EU, that human rights due diligence should be conducted on business practices of the companies involved in AI development and use.

It's very important that human rights approach to AI development and use be fully integrated into:

- administrative context of companies, universities, labs,
- social contexts of users of AI-driven systems, and
- regulatory processes related to AI.

Big tech companies executives, developers, 'marketologists', product managers and researchers should have guides helping them keep in mind human rights when they design and implement AI-driven solutions. It is important not only in current real-world practical applications of AI but also in the development of its further directions, new approaches to machine learning and machine teaching as they could all have a great impact on the evolution of human rights in the future.

In its landmark Guiding Principles on Business and Human Rights, released in 2011, the UN called on the industry to respect and protect human rights and provide remedies for HR violations, when developing and deploying their products. According to the document, technology companies should:

- conduct human rights impact assessment throughout the life cycle of their products,
- identify and respond to risks related to AI deployments,
- support channels of communication with local civil society groups and researchers,
- develop tool-kits for algorithmic impact assessments, and
- re-evaluate their methodology in the light of new developments and findings.

The document regards business enterprises as 'specialized organs of society performing specialized functions, required to comply with all applicable laws and to respect human rights'. States are regarded as guardians of human rights and freedoms and they are given the powers to interfere when a right is breached or abused by a private enterprise or another entity. In that case, they should take appropriate steps to investigate, punish and redress the abuse.

Although technical specifications for AI-driven products and systems do not include human rights principles, they could and should be operationalised in their products, design and business models. Sometimes big companies who care about their reputation refuse to work on certain AI technologies that are perceived by society as an attack on its freedoms.

Facial recognition might be one of the world's most divisive technologies. The heated debate around it at first focused on privacy worries. In May 2019, San Francisco became the first major American city to block law enforcement agencies from using facial recognition software. The debate has taken on a new twist as the technology has spread to be used by the US police in its routine work. In 2019, researchers at the MIT Media Lab found very hight disparities in error rates in facial identification systems, especially between lighter-skinned men and darker-skinned women. The police were admittedly biased: for example, as early as in 2016 during the Freddie Gray protests, it used facial recognition to identify protesters by linking images to social media profiles [3].

Then came the George Floyd protests that started in May 2020. In the wake of the countrywide unrest, IBM cancelled offering law enforcement customers its face recognition and analysis software. In unusually strong terms, IBM said that it "firmly opposes and will not condone uses of any technology, including facial recognition technology offered by other vendors, for mass surveillance, racial profiling, violations of basic human rights and freedoms, or any purpose which is not consistent with our values and principles of trust and transparency" [5]. Civil liberties groups said facial recognition contributes to privacy erosion, reinforcing bias along the racial lines and is prone to misuse. Following the suit, other big tech companies like Google and Amazon backed away from selling the police the face recognition systems that already had been contracted. In her interview to the New York Times, Timnit Gebru, a leader of Google's 'ethical artificial intelligence team', thus summed up the uneasiness she felt about the use of the software: "Facial recognition is being used against the black community… The combination of over-reliance on technology, misuse and lack of transparency—we don't know how widespread the use of this software is—is dangerous" [7].

Technology-minded politicians and executives often push for tech solutions without listening to community leaders, sociologists and activists who propose adjustments taking human rights into account.

5.4 Policy Response II: Key Requirements and Guidelines for Trustworthy AI

One of the latest examples of an international policy framework in the field of AI is the 'Toronto Declaration: Protecting the rights to equality and non-discrimination in machine learning systems' published in May 2018. The declaration was prepared by the international rights organizations Amnesty International and Access Now and then endorsed by Human Rights Watch and Wiki-media Foundation. Stating that human rights are "universal, indivisible and interdependent and interrelated", it underlines the centrality of the universal, binding and actionable body of human rights law and standards to the field of machine learning and AI developing. Focusing on the rights to equality and non-discrimination, the document admits that machine learning and artificial intelligence more broadly, impact a wider array of human rights, such as the right to privacy, the right to freedom of expression, participation in cultural life, the right to remedy and the right to life.

According to the Toronto Declaration, "systems that make decisions and process data can also implicate economic, social and cultural rights; for example, they can impact the provision of services and opportunities such as healthcare and education and access to opportunities, such as labour and employment"[8].

The concept of trustworthy AI was launched to help deliver on the promise of the benefits of AI while addressing its various risks and life-critical consequences for people and society [6]. The term 'trustworthy computing' was coined in 2002 in Bill Gates's memo [1], which referred to an internal Microsoft white-paper. The document identified four pillars to trustworthiness:

- security,
- privacy,
- reliability, and
- business integrity.

The first three properties were supposed to give the customer a good reason to trust Microsoft software and services. Later, the concept of trustworthy computing has come to mean a set of overlapping properties:

- Reliability: does the system do the right thing?
- Safety: does the system do no harm?
- Security: how vulnerable is the system to attack?
- Privacy: does the system protect a person's identity and data?
- Availability: is the system up when I need to access it?
- Usability: can a human use it easily?

Some companies have developed their own frameworks for the trustworthy AI. For example, the consulting company Deloitte introduced a framework that called for ethical directions in the design, development, deployment and operational phases of AI system implementation. The most advanced framework of that kind, however, was proposed in April 2019 by the EU High-Level Expert Group on AI. According to their Ethics Guidelines for Trustworthy Artificial Intelligence [4], it should be:

- Lawful —respecting all applicable laws and regulations.

- Ethical—respecting ethical principles and values.
- Robust—both from a technical perspective while taking into account its social environment.

The Guidelines put forward a set of seven key requirements that AI systems should meet in order to be deemed trustworthy:

- Human agency and oversight: AI systems should empower human beings, allowing them to make informed decisions fostering their fundamental rights.
- Technical robustness and safety: AI systems need to be resilient and secure and should have a fall back plan in case something goes wrong. Privacy and data governance: ensuring privacy, data protection and adequate data governance mechanism.
- Transparency: the data, system and AI business models should be transparent. Humans need to be aware that they are interacting with an AI system and must be informed of the system's capabilities and limitations.
- Diversity, non-discrimination and fairness: bias must be avoided and AI systems should be accessible to all, regardless of any disability.
- Societal and environmental well-being: AI systems should benefit all human beings, including future generations and should be environmentally friendly.
- Accountability: ensuring responsibility and accountability for AI systems and their outcomes, with the assessment of algorithms, data and design processes.

A trustworthy AI system should possess a quality of interpretability/ explainability, answering the question: can the system's outcome be justified with an explanation that a human can understand and/or that is meaningful to the end-user? Ethical considerations that directly relate to human rights could be summed up in the following questions:

- was the data collected in an ethical manner?
- will the system's outcome be used in an ethical manner?

The academic community explores various approaches to achieve trustworthy AI. In October 2019, the US National Science Foundation announced a new program to fund National AI Institutes. One of the six themes is named

'Trustworthy AI'. It emphasizes properties such as reliability, explainability, privacy and fairness.

5.5 Policy Response III: Human-Centric AI Application to Legislation

Most experts agree that in the foreseeable future, AI systems will be based on machine learning whereby artificial neural networks perfect their outputs by taking in massive datasets and applying to them more and more sophisticated learning models. But a growing reliance on human decision-making processes on algorithms causes legitimate public concerns about AI, even threatening a possible backlash. Also, the increasing complexity of AI systems that shift from one-problem to multiple-problem models calls for the integration of humans into the systems on the fundamental level.

Artificial intelligence systems that are not human-centric can not do without human labour needed to collect data and annotate it. However, in general, the system, having received annotated data, works with them without a human in the loop. Further, a person appears only as a consumer of a service or an object of a decision made by the AI system, the decision which often cannot be cancelled or changed.

The rapid development of systems with artificial intelligence is fraught with risks, which are often forgotten in the race for technological excellence. A new approach is needed, with the deep integration of people both at the stage of preparing data for machine learning and when working with artificial intelligence systems in the real world. This approach will reduce risks. In addition, it can lead to new technological and scientific achievements.

Human-centred AI approaches imply a complete transformation of this scheme. A person is included in the circuit of both the training system and the functioning of the system in the real world. At the same time, a person is used to objectively annotate data, for example, to identify objects in a particular sample. In the case when the data requires a subjective annotation that involves emotional intelligence or moral and ethical judgment, it can be carried out using the collective mind: 'the wisdom of the crowd', 'exchange of opinions' and so on. Different sociological and statistical models can be used to formulate an answer to such requests.

At the stage of functioning in the real world of an already trained system, a person controls the decision of the system. In this case, human participation is sought if the uncertainty of the system exceeds a certain limit. Promising in the near future are approaches to machine teaching that allow us to make objective annotations with a few examples such as: active learning, supplementing data, one-time learning and transfer learning.

The human-centred AI approach puts people into the loop of decision-making in the real-world applications of AI as well as in training AI models. In a certain sense, it is a return of the Norbert Wiener understanding of cybernetics as he focused on humans in their interaction with the machines.

Human-centred AI is studied and developed in such leading academic centres as MIT and Stanford University and advanced private laboratories, for example, DeepMind and OpenAI. The focus of their research is the optimization of algorithms and training models themselves, as well as the optimization of data annotations for machine learning. In the latter case, it is not the algorithms or models that are optimized, but precisely how the data is selected, on the basis of which the algorithms are studied. Of course, methods for optimizing data annotation for deep learning require generalization. Thus, the system will operate in the real world, based on a small sample of data—and the more efficient the optimization, the smaller this sample.

The fact of generalization itself means that the system will act in a situation of uncertainty. The incoming information will always be incomplete, therefore, for security reasons alone, restrictions are necessary in the development of algorithms for such a system and directly for its actions in the real world. These limitations can be obtained by human observation and control of the system, built into its circuit at all stages: at the stage of training the neural network, testing and working on errors, as well as its real-world operation and the resulting feedback.

Some big tech companies, for example, Google and Microsoft, have developed ethical principles to ensure AI beneficial application and mitigate its risks. Sometimes they collaborate with external organizations and non-profit groups such as the Partnership on AI (*partnershiponai.org*), the Asilomar Principles, Open AI (*openai.com*), Fairness, Accuracy and Transparency in Machine Learning (*fatml.org*). Stanford University has launched its Human-Centred AI Institute (*hai.stanford.edu*) and the Institute of Electrical and Electronics Engineers (IEEE) has brought scholars together with technologists and civil society activists to launch the Global Initiative on Ethics of Autonomous and Intelligent Systems.

While the mentioned big tech companies deserve credit for developing and publishing ethical principles for human-centred AI, the influence of these efforts beyond their own premises remains uncertain. Ethics statements of corporations may guide their operations, but they do not establish a broad framework under which all AI systems can be governed. That’s why a comprehensive, global governance to address the full spectrum of AI challenges and risks is needed.

5.6 Importance of Collective Decisions and Openness

Some governments are having a hard time incorporating human rights protections in their AI policies and regulations. Due to fear of missing out such a promising field, dozens of countries have adopted their long-termed national strategies and blueprints for AI development. Not all of them, however, included the points on human rights due diligence in their documents. Sometimes their broader policies related to business help them cope complex difficult cases that appear in mostly uncharted territory of real-world practices of AI-driven systems. For example, the EU has far more stringent data-protection, privacy and antitrust laws as compared to the US, but often these laws alone are not sufficient to block questionable corporate practices, for instance, of surveillance.

The European Union, however, has shown an interest in applying rights-based principles as a basis for regulating technology companies, via GDPR and related legislation. By doing so, it established new protections for European citizens' rights which impacts any business or organization collecting European residents' data or working on EU commission. The EU approach fundamentally differs from that of the US, China and some other countries who actively develop AI. In the EU, companies must justify their data activities within the GDPR's regulatory framework. The regulations include a requirement to notify people when personal data is breached and use privacy by design when building AI systems. Individuals also have a right to erasure of data and expanded protections against decision making authored by automated systems. GDPR also has a definition of 'consent' that puts limits on a company's ability to manipulate people into approving personal data use by default. Violations incur substantial fines, up to a possible 4 per cent of a company's global revenue, and class-action lawsuits are allowed.

Global reach of AI tech implies the necessity of global solutions and common international approaches for AI development that respects human rights. That's why governments, especially of major AI powers such as China and the US, should and can be much more proactive in multilateral institutions, like the UN. UN special rapporteurs and investigators continuously research the human rights impacts resulting from corporate AI-driven systems and publicize the results of their findings.

UN officials continuously evaluate whether existing international mechanisms and procedures for human rights monitoring, accountability and redress are adequate to respond to the rapidly emerging AI technology. The EU is also active in international technology-related debates. *Not once, its bodies such as the Council of Europe stressed the issues of fundamental rights and human dignity as shared global values for the whole of the international community.* In 2018, the Council of Europe's Commissioner for Human Rights argued for safeguarding human rights in the era of AI, particularly the rights of privacy and equality and freedoms of expression and assembly.

The argument was supported by a Council of Europe study noting that "there is growing concern at the political and public level globally regarding the increased use of algorithms and automated processing techniques and their considerable impact on the exercise of human rights."

In 2017, New York City adopted a law to help ensure that algorithms used by city agencies are transparent, fair and valid. The city set up a task force to make recommendations on algorithmic regulation, transparency and bias. This move to regulate AI may become a model for other urban agglomerations where local and regional governments aim at developing within the framework of the smart city concept.

In 2018, Canada and France called for the creation of an international study group that could become a global reference point for research on artificial intelligence issues. The same year, the Australia Human Rights Commission launched a project to directly address the human rights impact of AI and emerging technologies, which may serve as a guide for other countries.

Meanwhile, some tech workers, developers and researchers have launched public campaigns to pressure their employers to stop building technologies and AI that may effectively cause social harm. In April 2018, around 4,000 Google employees sent a letter to their CEO demanding the company cease its contract with the US Department of Defense to build AI systems for warfare. The letter called the project Maven that Google was to take part in a 'biased and weaponised AI'. In June 2018, Google worked out and released a statement of its AI principles, which said it would still work with the defence industry but would not develop any weaponised AI. 65 In the declaration, Google stated that it would not design or deploy AI technologies "whose purpose contravenes widely accepted principles of international law and human rights."

The international attempts at keeping AI tech in check involve academic experts and researchers as well as human rights and legal scholars. They examine how humanitarian law and ethics could be embedded into the emerging AI technologies and what the trade-off is between rights when faced with specific AI risks. Social scientists investigate the on-the-ground impact of AI on human rights. All this work and research can be more successfully completed through international networks of NGOs and multinational organizations, with the UN and UNESCO having the most expertise.

5.7 Concluding Remarks

This chapter aimed to provide an overview of major international policy responses to human rights concerns caused by AI systems rapid development and adoption.

There are a number of ways AI can help businesses to be more innovative and competitive. However, businesses should be aware of human rights due diligence to avoid violations and to be able to timely repair possible infringements on freedom of expression and other freedoms. Human rights principles should be operationalised in AI-driven products, design and business models.

The concept of trustworthy AI has shown to be fruitful in helping deliver on the promise of the benefits of AI while addressing its life-critical consequences for people and society. It has four pillars: security, privacy, reliability and business integrity. A set of seven key requirements that AI systems should meet in order to be deemed trustworthy is discussed in the context of ethical considerations to deliver the appropriate manner of both data collection and the AI systems outcomes.

A more radical approach is developing human-centric AI with its deep integration of people at all the stages of AI: collecting data, learning and real-time operation. This approach will further reduce risks and can lead to new technological and scientific achievements.

In summary, the importance of collective decisions and openness is underlined. Being afraid of missing out, dozens of countries have adopted their long-termed national strategies and blueprints for AI. Not all of them, however, included the points on human rights due diligence in their documents. NGOs and international organizations could bring their expertise into national efforts to make sure AI systems are trustworthy and safe.

References

1. Bill Gates: Trustworthy Computing (2002, January 15). *Internal Microsoft memo*. https://www.wired.com/2002/01/bill-gates-trustworthy-computing/

2. Dinsmore, G. (2011). *A Place In The World: Hannah Arendt And The Political Conditions Of Human Rights*. Retrieved from https://ecommons.cornell.edu/bitstream/handle/1813/33522/gld1.pdf?sequence=1&isAllowed=y

3. Eligon, J. & Stolberg, S. G. (2016, April 13). Baltimore After Freddie Gray: The 'MindSet Has Changed'. *The New York Times*, Retrieved from https://www.nytimes.com/2016/04/13/us/baltimore-freddie-gray.html

4. European Commission (2019, 8 April). Ethics Guidelines for Trustworthy AI. *High-Level Expert Group on Artificial Intelligence*, Retrieved from https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

5. Hern, A. (2020, June 9). IBM Quits Facial-Recognition Market Over Police Racial-Profiling Concerns. *The Guardian*. Retrieved from https://www.theguardian.com/technology/2020/jun/09/ibm-quits-facial-recognition-market-over-law-enforcement-concerns

6. Moltzau, A.(2020). *A European Approach to AI 2020. A Summary of The EU White Paper on Artificial Intelligence*. Retrieved from https://medium.com/dataseries/a-european-approach-to-ai-c37c334acc78

7. Ovide, S. (2020, June 10). A Case for Banning Facial Recognition. *The New York Times*. Retrieved from https://www.nytimes.com/2020/06/09/technology/facial-recognition-software.html

8. RightsCon Toronto (2018, May 16). *The Toronto Declaration: Protecting the Rights to Equality and Non-Discrimination in Machine Learning Systems*. Retrieved from https://www.accessnow.org/the-toronto-declaration-protecting-the-rights-to-equality-and-non-discrimination-in-machine-learning-systems/

Key Readings

- Asilomar AI Principles (n.d.) *Future of Life Institute*. Retrieved from https://futureoflife.org/ai-principles/

- Calvo, R. A., Peters, D., & Cave, S. (2020). Advancing Impact Assessment for Intelligent Systems. *Nature Machine Intelligence*, *2*(2), 89-91.

- Dignum, V. (2019). *Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way.* Springer Nature.
- The Future of Human-Centered AI: Governance Innovation and Protection of Human Rights (2019, April 24). *Medium.* Retrieved from https://medium.com/stanfords-gdpi/the-future-of-human-centered-ai-governance-innovation-and-protection-of-human-rights-5c371f195232
- EU Commission (2020). White Paper on Artificial Intelligence—A European Approach to Excellence and Trust. *COM (2020)*, *65*. Retrieved from https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf
- Executive Office of the President of the United States (2016, October). *The National Artificial Intelligence Research and Development Strategic Plan*. National Science and Technology Council Networking and Information Technology Research and Development Subcommittee. Retrieved from https://www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf
- OHCHR (2011, June 16). *Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect, and Remedy' Framework.* Retrieved from https://www.ohchr.org/documents/publications/guidingprinciplesbusinesshr_en.pdf
- 'Trustworthy AI' is a Framework to Help Manage Unique Risk. (2020, March 25). *Technology Review*. Retrieved from https://www.technologyreview.com/2020/03/25/950291/trustworthy-ai-is-a-framework-to-help-manage-unique-risk/

6. Closing Notes

Ibrahim Kushchu, Tuba Demirel

Are we in charge of determining our future as individual agents forming societies? Given the current state of interferences of digital technologies in our lives, it is becoming increasingly difficult to provide a clear 'Yes' as an answer to this question. The individuals' data all over the world has simply become the raw materials of information-processing factories of tech giants worth trillion dollars. They noticeably determine our daily lives, social interactions, business lives and naturally our future. What makes digital presence even more complicated is the latest technological developments in IT, one of which is the use of AI with big data. The Facebook/Cambridge Analytica, as one of the greatest scandals so far, was based on conventional digital models and techniques of processing and using the data. With recent developments in the AI sector, digital environments' power lords are getting even stronger and having more influence in our lives.

In this book, the impact of AI is presented from two different perspectives. One is related to empowering individuals to develop competencies to be less influenced by the forces determining our interactions in digital environments. The tech giants mainly impose these forces. More specifically, in this respect, MIL challenges relevant to using AI systems in digital environments are discussed. The other perspective extended the discussions to the implications of AI systems for human rights and freedom of expression. These two interrelated lines of analysis make it clear that considerations for HR and FoE will help to create ‘influential MIL’ where individuals have further competencies to have strong sense of ownership of their rights and to defend them accordingly. Furthermore, relevant stakeholders, such as governments and NGOs, need to support ‘influential MIL’ through policies and regulations. Below is a summary of some of the key points.

- Since the birth of AI in the 1950s, its techniques and models have not changed drastically. Main changes include the development of the infrastructure technologies (IoT, the cloud, and The Internet) that support AI, the speed and capacity of the connectivity and increased ability to collect, process and use the big data. Successes of ‘narrow’ AI systems are highly dependent on the volume, quality and representativeness of data they use.
- With the availability of big data, AI can now solve increasingly complex problems, including natural language processing, (video) image processing, face recognition, and challenging games such as Chess and the game of Go.
- The unique power of AI is hidden in its ability to search and collect targeted data (the power of pull) and disseminate customized and personalized data to the right target(s) with extreme precision (the power of push).
- Those who have an access to AI technologies and big data are in charge of how and for what purposes they can be used. This point may lead to substantial biases and significant risks when various communities and individuals are present and active in digital environments.
- Current AI technologies and the data used for AI systems seem to be concentrated in the technologically developed countries and the big tech giants’ hands. This raises issues related to MIL and human rights protection due to unbalanced control and power, and challenges for making appropriate regulations to promote equal access and safe AI throughout the world.
- The relationship between AI and MIL may be viewed from two different perspectives. Positive use of AI could lead to design and support for MIL efforts to improve appropriate competencies. However, the way AI-based technologies are being used now appears to have a number of implications for access to reliable information, freedom of speech and expression, privacy and control over one’s data. Although mostly as indirect results, such uses of AI seem to contribute to disinformation, fakes, abuse of privacy, bias and inequalities and other challenges.

- MIL professionals' primary tasks may be directed towards minimizing intended or unintended negative consequences of AI systems in digital environments. This may be possible by rethinking MIL efforts and moving towards 'influential MIL', which helps people recognize and exercise their rights and freedom of expression.
- Recognition of potential risks of AI use and its impact on individuals and societies spurred many stakeholders to address related ethical issues and promote development of safe, responsible and beneficial AI.
- Various countries and international organizations take steady efforts to legislate, regulate, and develop policies to minimize public and corporate conflicts, retain a human presence in controlling AI systems, and promote trustworthy AI.

Throughout humankind's history, technology, in various ways and forms such as fire, the wheel, printing, etc., survived with us as long as it supported and was in harmony with our lives. However, since the industrial revolution, with the aid of super machines and related technology, the human civilizations aimed at overcoming limitations of human muscles, speed and precision. With the recent developments in digital environments, we are now aiming to overcome the mind's limitations.

With mobile and wearable devices and the connected IoT world, our mind is already extended through intelligent applications that we use very often. With augmented intelligence, certain body parts (i.e. arms, legs and hands) may be replaced or enhanced (i.e. vision, hearing and motor speed). Significant efforts are going on in robotics and mechatronics to equip robots with human-like actions and gestures, facial expressions and simple imitations of emotions. Men and machines are inevitably converging to create trans-intelligence. At this stage, the boundaries of artificial and natural intelligence may be really blurred. One naturally wonders what it would mean to be human, then?

The emergence of trans-intelligence is well evidenced by the long-standing work on nano-bots swimming in our blood and reaching out to our inner organs to deliver and collect data and create gene-editing factories as well as the brain interfaces of the Neuralink project. At the time of trans-intelligence, it is no longer necessary to collect data from sources outside human body (i.e. social media and other digital interactions) when there is an opportunity to collect data from within the body of individuals. It is no longer necessary to process and create predictions from the big data, say as a basis for *recommender* systems, when there are opportunities to create direct *command* systems for inner organs, including brains of human beings.

At the time of trans-intelligence, the MIL, human rights and freedom of expression may have, completely different notion and meaning, if any at all.

ABOUT THE AUTHORS

Igor Shnurenko

Igor Shnurenko is an independent AI expert and consultant with a unique background in top-notch journalism. He is an author of two books on artificial intelligence published in Russian in 2020. The book 'The Demon Inside: Anatomy of Artificial Intelligence' explains AI in layman's terms clarifying many myths and misconceptions. The book 'Homo Fractus/The Man Hacked' deals with the AI's social implications and impact on the future of society, state and humankind. Igor Shnurenko is also the author and presenter of the 'AntiTuring, the Anatomy of AI' podcast coming out in Russian.

He started as a reporter for the first independent newspaper in St. Petersburg (Russia), then launched independent magazines and made successful a newspaper for the homeless. He led investigative journalism workshops, headed a regional journalism organization, edited known global tech publications and worked as a producer for the BBC World Service in London. This experience formed a basis for his first-hand expertise in human rights and freedom of expression issues. Mr. Shnurenko spoke to diverse audiences: philosophers of the Russian Academy of Sciences, decision-makers at the Russian Council of International Relations and Public Chamber of Russia, researchers at the Higher School of Economics, technical universities and elsewhere.

He has Master's degrees in Engineering and Journalism. Upon graduating from the Missouri School of Journalism, he did research on public journalism at Duke University. During his stay in the US, he received the Foreign Press Association Award, Sarah McClendon Scholarship, Overseas Press Club of America Scholarship and Freedom Support Act Fellowship in Journalism.

Mr. Shnurenko's areas of interest include but are not limited to Human-Centric AI, technologies and control, human rights, freedom of expression and the risks of AI, futurism, social issues and philosophy of consciousness. He can be reached via ishnurenko@mail.ru or by phone +79218985127.

Tatiana Murovana

Tatiana Murovana is a programme specialist at the UNESCO Institute for Information Technologies in Education (Moscow, Russia), where she is responsible for managing projects aimed to promote media and information literacy and digital competence for quality and inclusive education. Since 2000, her contribution to UNESCO activities included management of a number of projects on information access, digital information preservation, information ethics, media education and information literacy, and linguistic diversity and multilingualism in cyberspace. In addition, she has edited and co-edited about 20 publications.

Ms. Murovana holds Master's degrees in Psychology and Chemistry. She has completed further training courses on media literacy (University of Pennsylvania), international policy (Diplomatic Academy of the Russian Ministry of Foreign Affairs) and psychology (Moscow University of Humanities). In the past, she was employed by the UNESCO Moscow Office, the Russian Committee for the UNESCO Information for All Programme, the Russian Agency for the International Humanitarian Cooperation and the Russian State Duma.

She is enthusiastic promoter of the media and information literacy and digital competencies for educators. Her areas of interests also cover the impact of emerging technologies and AI on society and educational systems. She can be reached via t.murovana@unesco.org.

Ibrahim Kushchu

Ibrahim Kushchu is a consultant, author and a reputed speaker with the expertise in artificial intelligence, mobile technologies and management systems. He founded the field of Mobile Government – the use of mobile technologies to offer public services to citizens. His consultancy expertise ranges from community informatics to enterprise mobility. Prof. Kushchu has taken up several academic positions at business schools in the UK and Japan, teaching various courses related to AI, mobile techs, electronic business and mobile business.

Prof. Kushchu is recognized as a pioneering practitioner and researcher in the field of electronic government, who has critically influenced the development of the Mobile Government. His expertise also extends to the impact of mobile technologies on economic and social development. He has designed Mobile Government strategies and road maps for various countries globally, including Afghanistan and the United Arab Emirates most recently. In addition to offering advisory services to local and central governments, he has been working with various multinational organizations, including the UN, the Bill & Melinda Gates Foundation, Cisco, Nokia, Hitachi and NTT DOCOMO.

He authored four books, edited or co-edited three books on mobile government, and has been publishing articles in various international journals and conference proceedings. He is also very active in the international community of researchers through speaking, organizing, chairing and co-chairing international conferences, and serving in the committees. Prof. Kushchu holds BSc degree in management, MBA, MSc degree in artificial intelligence and PhD degree in evolutionary artificial intelligence (University of Sussex, UK). Currently, Prof. Kushchu manages TheNextMinds.com, an AI consultancy and training firm, and can be contacted via info@thenextminds.com.

Tuba Demirel

Tuba Demirel is an expert on digital competencies in education and digital identity on social media.

As a graduate student in applied linguistics at Sussex University (UK), she explored the implications of language use and differences in real and virtual identities on social media. She worked as the international relations coordinator at the Mustafa Kemal University (Turkey). In this position, she contributed to the implementation of EU education and training programs, including organization of study visits to EU countries, and was responsible for project management and partnerships. Currently, she is a member of academic staff at the English Language and Literature Department of the University. She has been speaking and publishing at international conferences and academic journals. She is finalizing her PhD thesis on Digital Competencies and Foreign Language Teaching at Hacettepe University.

Ms. Demirel is primarily interested in the adoption of educational technologies in foreign language teaching. She explores opportunities in using them to improve language teaching, in particular through increasing digital competencies of teachers, and follows the evolution of virtual identity on social media. Her wider interests include understanding emerging and advanced technologies such as mobile platforms, augmented reality and artificial intelligence. She can be contacted via taydinoglu@ mku.edu.tr or tubaaydinoglu@hacettepe.edu.tr.