



UNIVERSITY OF PIRAEUS - DEPARTMENT OF INFORMATICS

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ – ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

MSc «Digital Culture, Smart Cities, IoT and Advanced Digital Technologies»

ΠΜΣ «Ψηφιακός Πολιτισμός, Έξυπνες Πόλεις, IoT και Προηγμένες Ψηφιακές Τεχνολογίες»

MASTER'S THESIS

| | |
|--|--|
| Thesis Title: | Data and AI in Social Systems: Trust Perspectives in a Digitally Developing World Δεδομένα και Τεχνητή Νοημοσύνη στα Κοινωνικά Συστήματα: Προοπτικές Εμπιστοσύνης σε έναν Ψηφιακά Αναπτυσσόμενο Κόσμο |
| Student's Name: | Olga Giannaki |
| Father's name: | Klement |
| Student ID No: | ΨΠΟΛ/20014 |
| Supervisor: Co-Supervisors: | Professor Dimitrios Vergados (University of Piraeus) Professor Felix Gomez – Marmol (University of Murcia) Dr. Javier Pastor – Galindo (University of Murcia) |

July 2023

3-Member Examination Committee

Dimitrios D. Vergados

Professor

University of Piraeus

Angelos Michalas

Professor

University of Western
Macedonia

Dr. Emmanouil Skondras

Postdoctoral Researcher

University of Piraeus

Abstract

In an increasingly digitally developing world – a world where most aspects of modern life are directly or indirectly related to digital technology – the relationship between data, privacy, and trust has become a focal point of contemporary debate. The enormous amount of data generated, partly due to our constant preoccupation with social media and making transactions online, has increased the demand for online security and assurance that being on the internet does not signify total exposure.

The lack of such guarantee is shaking the foundations of social and digital trust. Thus, it is imperative that we delve into the intricate dynamics and interdependencies that exist within this framework, and explore the challenges and opportunities posed by the rapid development of digital technologies. By examining the multifaceted aspects of data collection, storage, and utilization, we can pinpoint the consequences this signals for privacy and trust within social systems. The delicate balance lies within taking into account the financial, political, legal, ethical, and finally societal dimensions surrounding data governance without losing focus of the ultimate goal: establishing user trust and maintaining social cohesion.

Table of Contents

| | |
|--|----|
| Abstract | 3 |
| Table of Contents | 4 |
| Introduction | 5 |
| <u>Part 1: Data and AI</u> | 7 |
| 1. What is Data?..... | 8 |
| 2. Data Theft | 12 |
| 3. Artificial Intelligence and Data | 16 |
| 4. Examples of Artificial Intelligence Misuse | 20 |
| <u>Part 2: Modern Societies Constructed Around Data</u> | 26 |
| 5. The Business Aspect of Data | 27 |
| 6. The Political Aspect of Data | 30 |
| 6. a. Liberal Democracies | 30 |
| 6. b. Authoritarian Regimes..... | 31 |
| 7. The Legal Aspect of Data..... | 35 |
| <u>Part 3: Trust in the Context of Digitalized Social Systems</u> | 41 |
| 8. Establishing Human Trust | 42 |
| 9. Trust in a Digitally Developing World | 47 |
| Conclusion and suggestions for further research | 56 |
| Bibliography | 57 |

Introduction

In today's interconnected world, data plays a role of great importance. It has become the lifeblood of science, technology, and business production, facilitating innovation and overall human progress. Data holds the key to valuable insights that can help us make decisions and transform how we understand and interact with the world around us, as it fuels scientific advancements, paves the way for economic growth, and enhances our everyday experiences. From a simple social media update to the complex organization of business client lists and government digital records, data is at the core of our integrated world, shaping the present and likely the future of our societies.

In this framework, this thesis discusses the relationship between data, privacy, and trust that exists in human social systems of our age. Part One explores the nature of data, data functions and how it is intended to be used by applications. By differentiating between private (closed) and publicly available (open) data, it then delves into data security and data theft, how cyberattacks work, and who may steal data (hackers, terrorist/activist groups, governments). It also looks at the use of stolen data by individuals compared to monitoring by governments, and the role of Artificial Intelligence (AI) in facilitating data fraud, as well as tangible examples of AI misuse, including the Cambridge Analytica scandal and other notable incidents (such as the use of deepfakes for comedic or political purposes).

In Part Two, the thesis examines the use of data from 3 different yet associated scopes: the commercial/business aspect, the political aspect, and the legal aspect. The first refers to conducting business with data (e.g., insurance companies offering deals based on social media profiling). The political aspect of data use is centered around how data is used in democracies compared to oligarchies. There is an extensive description of how the social profiling and social monitoring systems work in China, as well as what the case is when it comes to public surveillance in liberal democratic countries like the Netherlands and France. Lastly, the legal aspect explores how authoritative-governed countries, such as China, regulate data and its use (e.g., PIPL laws), and – on the other hand – how democracies protect the rights of citizens (e.g., GDPR laws); in the latter part there is also mention of the legal status quo in other low and middle-income areas of the world, like Africa and India. The existence of laws is vital, chapter 7 highlights, as it allows people to coexist in a safe and secure environment; equally important is another element that functions as the “glue” within a society: trust.

Part Three elaborates on the importance of trust in human societies and how it has helped shape modern social systems. The final chapters underline the challenge of having trust in one another in a rapidly changing world, in which it is simpler than ever to distort what we would perceive as reality. The rising production of deepfakes, the rapid spread of fake news and the privacy and security questions emerging are only a few examples of contemporary challenges brought forward by the increasing use of Artificial Intelligence. The thesis concludes by emphasizing that establishing trust in social systems of our society (individuals, groups, or institutions) is particularly challenging, yet essential if humans wish to live in prosperous societies. By encouraging media literacy and cybersecurity education to all users, by ensuring that the law continues to protect the rights of citizens and with the help of digital technology itself, we can ensure that digital technologies remain a human tool and not a human equal.

Finally, this thesis is based on the common scientific acceptance that the modern world is quite different than it used to be since the emergence of digital technologies. The modern world is a modern state of affairs, a mixture of the physical world as we consciously perceive it as humans, and the digital world. In this state of affairs, the Digitally Developing World, our everyday reality is linked to the material world as we know it, but the digital aspect seems to become more prominent in our daily lives through our online presence on a social, financial, and political level (e.g., social media, e-banking, e-government services). This is the most important key take-away: we might be (entering) in the homo digitalis era, yet we are still human, and our societies function upon the same principles: trust, integrity, and social cohesion.

PART ONE – Data and AI

data

noun

noun: data

1. facts and statistics collected together for reference or analysis.
2. the quantities, characters, or symbols on which operations are performed by a computer, which may be stored and transmitted in the form of electrical signals and recorded on magnetic, optical, or mechanical recording media.
3. PHILOSOPHY
Things known or assumed as facts, making the basis of reasoning or calculation.

Origin: mid-17th century (as a term in philosophy): from Latin, plural of datum.

Definitions from Oxford Languages

[Source: <https://languages.oup.com/google-dictionary-en/>]

1. What is data?

One of the main characteristics that differentiate humans from other animals lies in the complexity of cognition (Luppi et al., 2022). Although every being endowed with a brain has the ability to perceive the outer world to some extent through its senses, our unique capacity of noticing, registering, processing, and sharing large and complex pieces of information with other humans is what has gradually led us to what one could poetically call the apex of the animal kingdom.

Certainly, this cognitive ability did not preexist innate to humankind, but was rather part of a lengthy evolutionary process: the more humans were forming bigger societies, the more sharing precise information fast and efficiently became a need of vital importance. As humans found ways to tackle everyday life challenges (such as finding shelter), and educate their offspring regarding those, the information being shared became more complex. And, consequently, the more mentally engaging our social relationships became, the mightier our brains grew.

The above tangible paradigm derives from real life circumstances – the analogue world. However, it is also a pattern that can be found in what is sometimes called the “digital world”, which consists of computers, networks, programs, applications. All the above is a human invention predicated on what had until recently only been happening in the natural, physical sphere: collecting details, mentally saving them, processing them, possibly transmitting them, and finally making assessments or decisions based on the outcome – the substantial *information* – we got from the specifics we have evaluated. With regards to the “digital world”, we principally refer to these details as *data* (Merriam-Webster).

If we were to make a categorization in order to better understand the relationship between data, information, knowledge, and wisdom, we would use the DIKW pyramid: a hierarchical model that suggests that data is the lowest level, followed by information, knowledge, and wisdom at the highest level. This pyramid (Image 1) is used to illustrate the idea that data, when organized and presented in a meaningful way, allows us to deduct information; this information is then used to gain knowledge (the understanding and application of information to make decisions and solve problems), and eventually wisdom – the ability to make sound judgments and decisions based on experience, knowledge, and insight (Rowley, 2007).



Image 1: The DIKW pyramid (Shôn Ellerton, 2019)

[Source: <https://medium.com/ironkeel/dikw-pyramids-and-car-crashes-de77591f1491>]

Data, which is the core of the pyramid above, is characterized by its presence in digital networks, which allows computer systems (e.g., personal computers, routers, servers, smart devices, etc.) to function and

instantly communicate with one another (Rowley, 2007). These systems are able to transmit the data and/or collect and store it for future use, analysis, and evaluation. Data can be collected either by humans, who manually manage these systems to enter and extract information, or by automatic observation and tracking of the flow of information by digital means. Such means include algorithms, applications and artificial mechanisms, whose role is to monitor small networks or even the world wide web (online tracking) (Bumblauskas et al., 2015).

Large and composite sets of data that need to be processed, usually with the help of non-traditional methods (e.g., computers and applications enhanced by Artificial Intelligence), are commonly referred to as *Big Data* (Schönberger et al., 2013). Big Data is characterized by what is called the 5 V's: volume (the amount of data), variety (the types of data: structured/unstructured), velocity (the speed in which Big Data is generated), veracity (the trustworthiness of data), and value (how organizations can use that data to derive information and, subsequently, knowledge) (Kitchin, 2014).

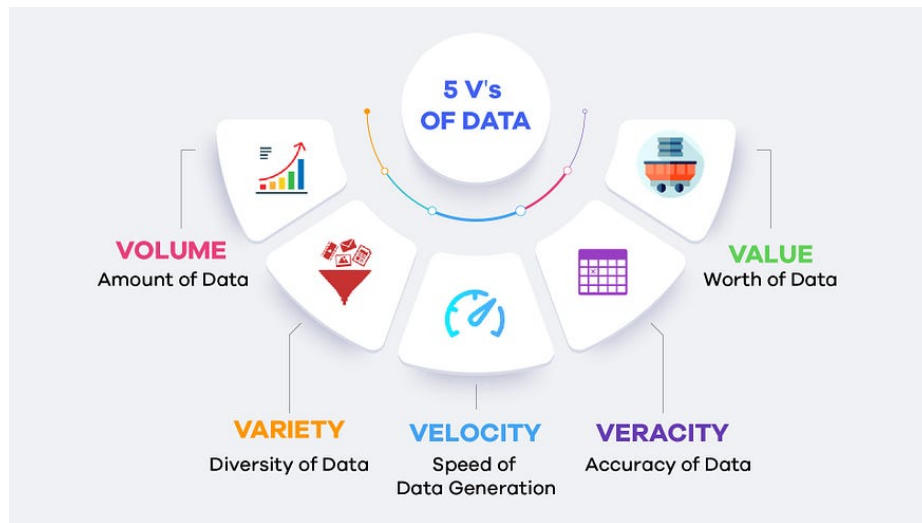


Image 2: *The 5V s of Data*

[Source: https://medium.com/@get_excelsior/big-data-explained-the-5v-s-of-data-ae80cbe8ded1]

Big Data can be collected either online or offline via structured or unstructured data sources. Structured data (e.g., dates, names, addresses, telephone numbers) are usually stored in categorization tables or simple Excel files and other relational databases, whereas unstructured data (e.g., images, audio, video, free text) can be collected from articles, blogs, forums, and social media (Facebook, Instagram, Twitter posts, TikTok, etc.) (Zulkarnain & Anshari, 2016).

The information being gathered can be of private nature (*closed data*) – meaning it can only be accessed by the system managing it, and the owner/manager of said data – or publicly accessible and shareable by anyone who wishes to do so (*open data*) (Kitchin, 2014).

Data collections entail elements in the form of numbers, words, images, sounds, and any means that can be digitalized. As the use of digital networks for the transfer of information has surpassed this of traditional means of information sharing, the volume of data emerging is multiplying at a tremendous pace. Therefore, instead of collecting these details, the main present-day challenge lies in organizing and sorting them in a way that would make them readily accessible, inspectable, and analyzable⁷. Fundamentally, in order to be able to extract safe conclusions and answer questions at hand, it must be ensured that this data is efficiently protected, and the information that it represents is kept intact in its entirety. That is because these questions are associated not only with scientific research, but also with everyday life decision-making, profit-related purposes, and policy making by financial entities (e.g., companies), organizations, and state governments.

Hence the importance of data security is becoming crucial, affecting us all on an individual, as much as on a societal level.

This is especially imperative in the era of digital globalization, which signals the integration and interconnectedness of economies, societies, and cultures facilitated by digital technologies and the internet (Manyika et al., 2016). There has been extensive discussion regarding how humanity now lives in the “Digital World”. Yet, in this thesis I defend the idea that this is not a “Digital” or “Digitalized” world that we are discussing about, but a *Digitally Developing World*.

My argument is that calling it the “Digital World” would suppose that all interactions and transactions take place *exclusively* in a computer-generated environment, a collective *virtual* shared space. It is not a “Digitalized World” either, because this would imply that what started in a physical dimension now exists fully in a digital one. As it will be demonstrated in this academic paper, our World, our current state of affairs, is a mixture of the tangible, the physical world as we perceive it through our senses and the help of our logic (also called the “real life”), and digital technologies, which are becoming all-pervasive in our daily lives. From personal accounts on social media, to online banking and digital public services (e-government), our “real life” identity and our digital profiles are not separate and independent. On the contrary, they tend to coexist and complement one another.

These two dimensions (physical and online) align and create an amalgam, the *Digitally Developing World*, which I ponder as follows:

a) it is a true-to-life “world”, because it includes our everyday reality, and it is not unaffiliated to the material world as generally known to humans;

b) it is “digitally developing” because it is not fully digital or digitized, but rather the digital aspect seems to increasingly become more prominent in our daily lives through our online presence on a social, financial, and political level (e.g., social media, e-banking, e-government services). What occurs in the digital sphere seems to tremendously affect our “real lives” as well;

c) it is a world where most aspects of modern life are directly or indirectly related to digital technology, from politics to education, warfare, marketing, etc.

The interaction between the Physical World, the Digitally Developing World and the Digital World can be demonstrated schematically as follows:

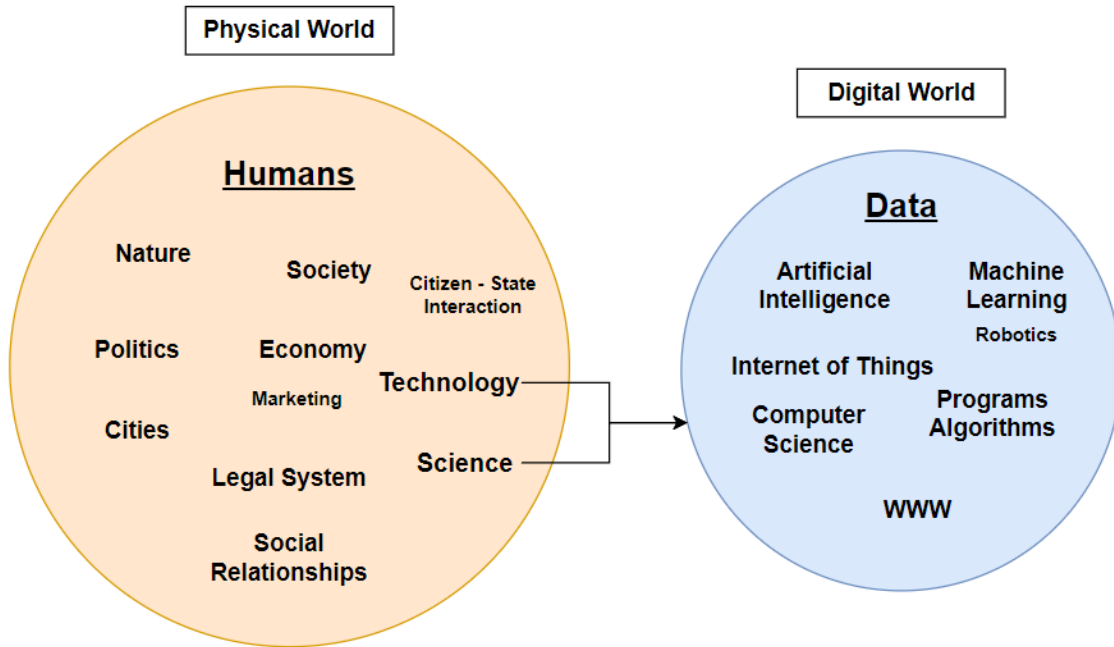


Image 3a: Circa 1980, Technology and Science contributed to the birth of the Digital World (Tyner, 2014), which grew as something separate from the Physical World.

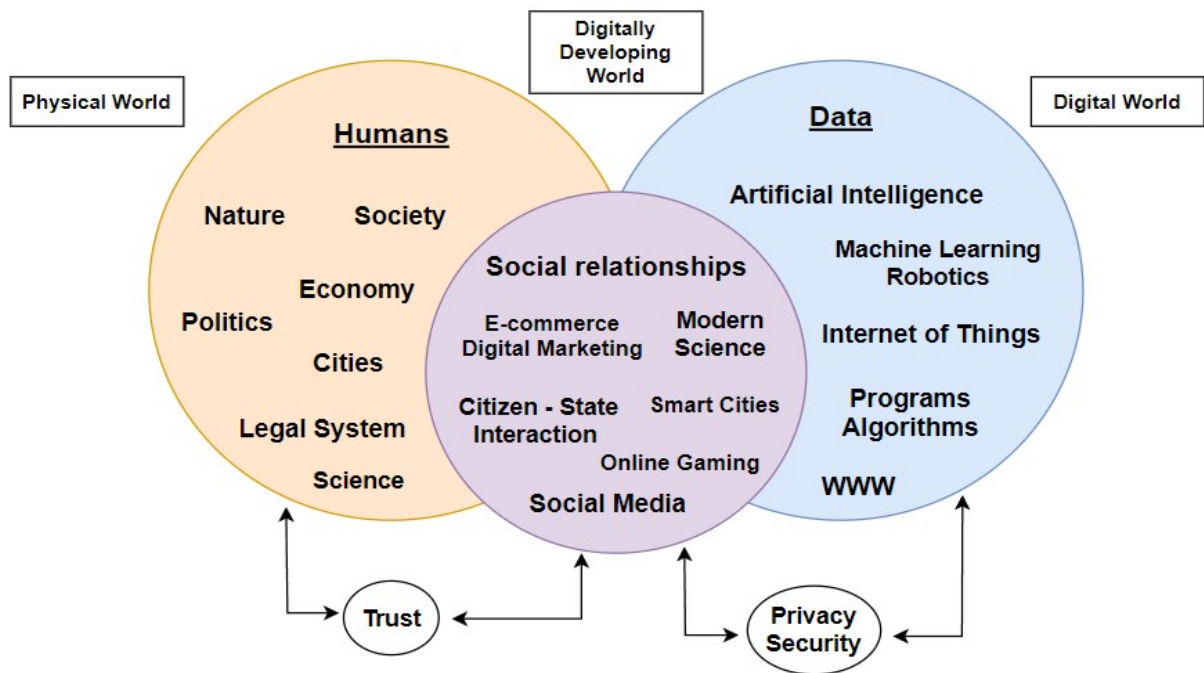


Image 3b: The Digital World and the Physical World coexist; the Digital World grew over the years as Technology and Science thrived and came to interact more with everyday human life. The fields where the Physical World and the Digital World meet create a new sphere/aspect of contemporary state of affairs: the Digitally Developing World (DDW). Trust among humans is the main common element that binds the Physical World together, while in the Digital World, this role is claimed by Privacy and Security of data and information. Both of these components and their importance are passed to the DDW in ways that will be explored in this thesis.

Before we see how the Digitally Developing World functions and how human relationships unfold in its spectrum, we will examine in detail the role data nowadays plays in the life of the average citizen.

2. Data Theft

Data security, the protection of data of individuals or other entities from unauthorized access, alternation/corruption, or destruction, appeared approximately at the same time as the Internet itself (Kuner, 2017). In an effort to ensure that the information conveyed would reach the recipient as it was meant to – unharmed by external factors and agents – the first antivirus software programs were created. The so-called CIA triad (Confidentiality, Integrity, Availability) started to represent the main three attributes of secure systems: confidentiality means that data is kept private and confidential; integrity means that data is accurate and not modified or destroyed; availability refers to the systems and resources being available when needed (National Institute of Standards and Technology, 1993).

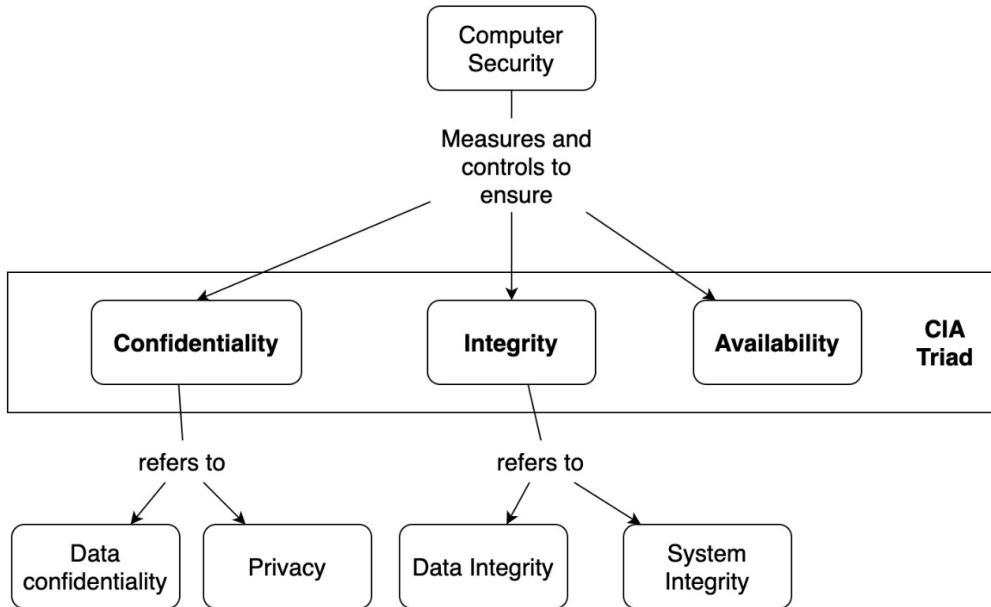


Image 4: A visual representation of the CIA triad

The antivirus programs were tasked with ensuring that the CIA was in place, and thus a system was secure. Their role was to stop foreign interventions and notify the user when their system was under attack or whenever there had been a security breach attempt. As over the years data-sharing became more widespread, and the data and the networks in which it existed (also known as cyberspace) evolved into more intricate, the demand for quick and efficient defense systems rose higher. That being so, the field of Cybersecurity and a whole data protection-centric industry were born.

At this point, it would be crucial to underline that the rapid development, which the domain of Cybersecurity has known, is a direct product of the increased number of human-induced cyberattacks. Types of cyberattacks have been the epicenter of study, especially during the past decade, as our social and economic lives become more and more internet dependent. Out of the many ways one could try to gain unauthorized access to other people's data, the most notable include (Bendovschi, 2015) :

- **Man-in-the-middle attack:** it aims to intercept and potentially alter the communication between two parties who believe they are directly communicating with each other. In a MITM attack, the attacker positions themselves between the legitimate parties, effectively “eavesdropping” on their communication and potentially manipulating the data being exchanged.
- **Brute force attack:** it involves repeated attempts to gain unauthorized access to a system or an account by systematically trying all possible combinations of passwords or encryption keys until the correct one is discovered.

- DDoS (Distributed Denial of Service): this attack compromises the availability of data by flooding the target (e.g., server) with commands, overwhelming its resources and rendering it inaccessible to legitimate users.
- Malware: it is a type of malicious software meant to disrupt, damage, or gain unauthorized access to computer systems, networks, or devices. Common types of malware include viruses, worms, trojans, spyware, ransomware, and adware.
- Phishing: this technique aims to trick individuals into revealing sensitive information, such as usernames, passwords, credit card numbers, or personal details. Phishing attacks often occur through fraudulent emails, instant messages, or websites that impersonate legitimate organizations or individuals, aiming to gain unauthorized access or carry out fraudulent activities.
- Social engineering: it is used by attackers to manipulate and deceive individuals into providing sensitive information, granting access to systems, or performing actions that they wouldn't normally do. It involves exploiting human psychology and leveraging trust, authority, curiosity, or fear to gain unauthorized access or obtain valuable information.

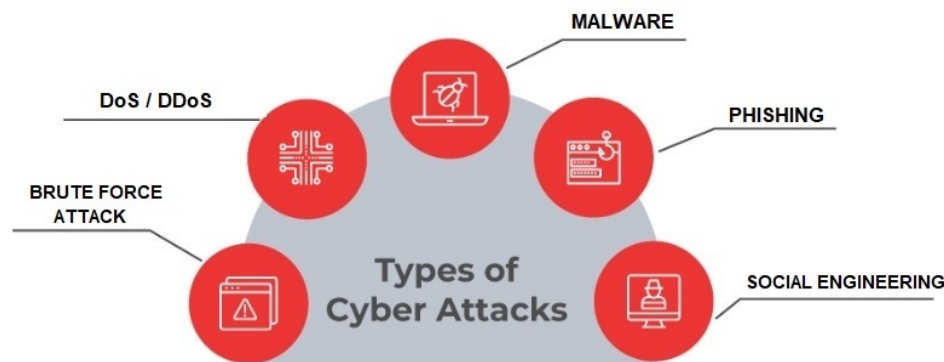


Image 5: Types of Cyberattacks

These attacks, whose goals most commonly include data theft or loss, can be initiated with the help of advanced computing systems utilized by individual agents (hackers) or organized groups. Hackers can use malicious software, such as viruses, worms, and Trojans, to break into a company's database and steal confidential information such as customer data, financial records, or intellectual property. They can also use social engineering techniques, such as phishing and pretexting, to gain access to sensitive data. Additionally, ransomware attacks, which involve hackers encrypting data and demanding a ransom payment, are becoming increasingly common (Kaffenberger & Kopp, 2019).

Around 90% of cyberattacks are linked to a financial (e.g., stealing information and selling it to private vendors) motive (LMG Security, 2022). Data theft with economic incentives involves taking hold of financial information like payment card numbers and information, PIN numbers and passwords, tax-related information (tax ID, Social Security numbers), and even medical records. Especially the latter is considered one of the most profitable assets in the internet black markets (*dark web*) (Taylor, 2021). From such documents, one is able to retrieve several pieces of personal information, and consequently keep hold of or sell them to be used for remarkably precise targeted advertising, blackmailing/extortion, identity theft (e.g., usurping the data to collect tax or other benefits) (Angel, 2018).

In addition to the above, the reasons behind data theft might be related to political agendas. These can be characterized by a sense of virtue and righteousness, or by political reasons of Machiavellian nature. Moral/ideological motives (such as *hacktivism* – e.g., causing harm to the systems of companies that harm the environment or specific groups of people) and political aims – in the sense of state politics and governance – (e.g., political profiling of citizens and collecting data on political opponents) are nothing but uncommon (Wall, 2007).

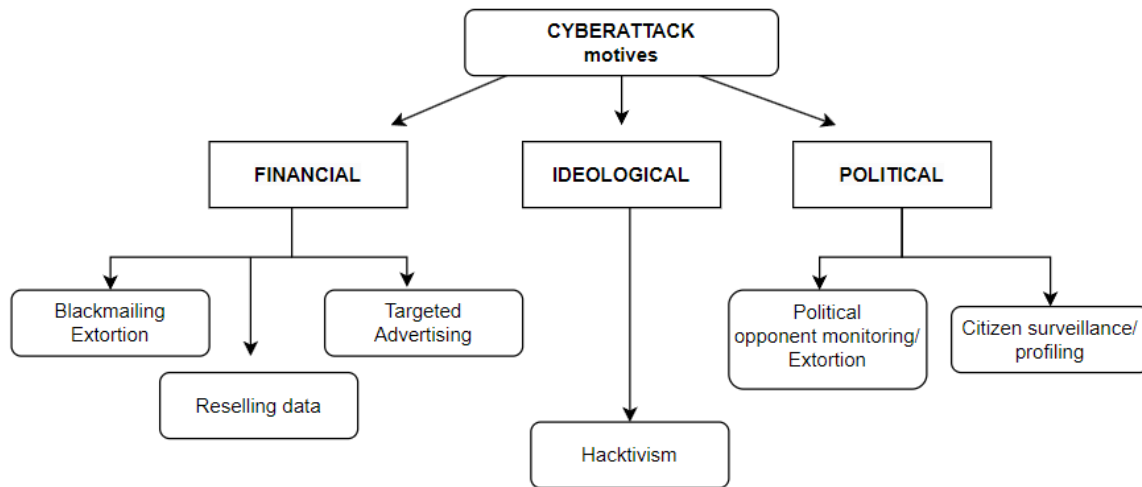


Image 6: Cyberattack motives and categories.

In the first category, we would find self-described activists, such as the *Anonymous* group, who target big corporations and prominent figures who are thought to be corrupt (Pendergrass et al., 2013). This is a particularly interesting “Robinhood” conceptualization, as the main notion is based on harming the strong in order to protect and actively defend the weaker, the more innocent, and at times the general population from the grasp of privileged elites.

In the second category we would include individuals or groups who hold political and/or financial power and can have leverage and impact over political affairs. These are agents with political motives, who might have influence on officials or governments, and aim to collect information for extortion, blackmail, career assassination and publicly embarrassing an opponent (Pastor-Galindo et al., 2021). One famous case of a politically motivated cyberattack occurred during the 2016 US presidential election. Hackers from Russia were believed to have used cyberattacks to gain access to confidential emails belonging to the Democratic National Committee and Hillary Clinton’s campaign team. The information that was stolen was used to try and damage Clinton’s reputation and influence the election in favor of Donald Trump (Sabato et al., 2017).

Furthermore, it is also state governments that can target data of political opponents, other governments, journalists, or mere citizens in order to monitor them and control the flow of information (Rosenzweig, 2013). One example of a state government using cyberattacks to target the data of political opponents is the Chinese government’s ongoing cyber campaign against the Tibetan independence movement. The Chinese have been accused of using cyberattacks to access Tibetans’ emails, social media accounts, and other private information in an effort to disrupt and discredit the movement (Donahue, 2010). The Chinese government has also been accused of using cyberattacks to access the data of Chinese journalists in order to monitor and suppress dissenting voices. This includes targeting journalists who report on subjects such as minority rights and the suppression of freedom of expression (Getz, 2022). Specifically, it has been reported that the Chinese government has attempted to hack into the computers of journalists in an effort to steal their confidential sources and documents (Matthews, 2022).

By gaining the upper hand in intelligence accumulation, it becomes easier for data thieves to promote their own agendas and interests, ensuring that obstacles and any counteractions are dismissed. This often poses a great challenge for political rights activists and lawmakers, especially in western-style democracies (e.g., EU, USA), where the people have the right and obligation to examine and validate government actions through elections. However, as the use of the internet has now become part of everyday life, the sense of privacy and duty to fight against restrictions is blossoming, even in authoritarian regimes where

information is much more controlled (Albrecht & Naithani, 2022).

Finally, as it will later be explained, data fraud can take place not only via actively stealing information, but also through gathering masses of open data that the user-victim has willingly or inadvertently shared themselves online.

3. Artificial Intelligence and Data

Up to this point, it has been established that there are masses of data and information generated, stored, and flowing through the network of the world wide web, and as more and more people use the internet and its mechanisms, this number will continue to grow. This has led to a need for more complex systems and technologies that are able to handle and manage Big Data in a functional and productive way, that – first and foremost – guarantee no loss of data, and secondly (but equally essential) that the data processed will evolve into information, which we can use to create knowledge (Wallace, 2007).

The most prominent amongst these new technologies is *Artificial Intelligence* (AI). As mentioned by P. Boucher in his 2020 study addressed to the Members and staff of the European Parliament regarding the significance of AI and human attitude towards it, *Artificial Intelligence (AI) is probably the defining technology of the last decade, and perhaps also the next* (Boucher, 2020). According to the European Commission's 2018 definition of Artificial Intelligence, *AI refers to systems that display intelligent behaviour by analysing their environment and taking action – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g., voice assistants, image analysis software, search engines, speech, and face recognition systems) or AI can be embedded in hardware devices (e.g., advanced robots, autonomous cars, drones or Internet of Things applications)* (European Commission, 2018).

AI systems are designed to perceive a compound of external stimuli, collect and interpret the data they have detected, process the information they have deduced from the data, and, finally, make decisions on whether to act based on the knowledge produced. Following this logic, AI systems have been accredited the term 'Intelligence', as they were designed to act rationally (Russell & Norvig, 2009), like human brains – much like an advanced version of the systems we had up until now (Image 7). Their main strength, however, lies in their ability to adapt their behavior by analyzing how the environment is affected by their previous actions; their groundbreaking function is that the system is able to train itself to perform better, to improve its methods with each usage.

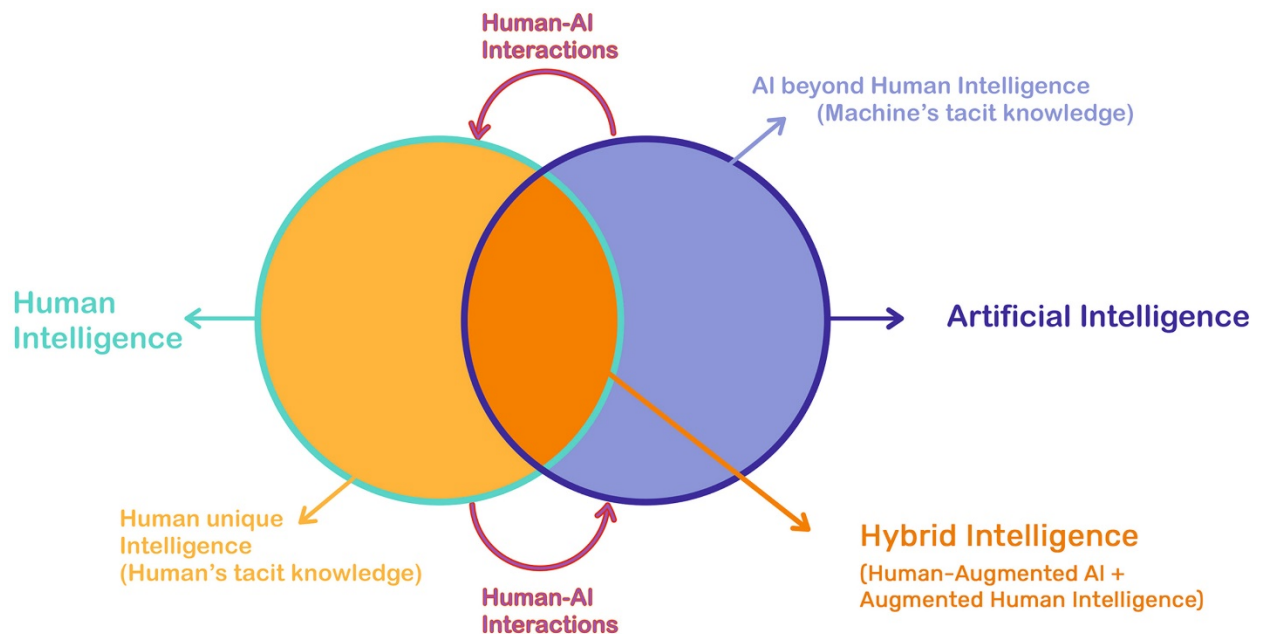


Image 7: Artificial Intelligence, human intelligence, and hybrid intelligence

[Source: Jarrahi, M. H., Lutz, C., & Newlands, G. (2022). Artificial Intelligence, human intelligence and hybrid intelligence based on mutual augmentation. *Big Data & Society*, 9(2).

<https://doi.org/10.1177/20539517221142824>

The above-mentioned capacity has also given birth to new fields: *Machine Learning* – on which there will be focus in this chapter, and *Enhanced Robotics* (also called *Embodied AI*). The latter is “*AI in action in the physical world*”, a pragmatic application of Artificial Intelligence, and it concerns building and developing artificial systems, such as robots or intelligent devices in general (Pfeifer & Iida, 2003).

On the other hand, Machine Learning is an expansion of the aforementioned capacity: computers or machines being able to learn from the data they collect from their surroundings, be it physical or digital. This results in them making their own decisions without necessarily humans having to interact with the system (Mitchell, 1997), and it is key to the “machine” being able to make predictions about how it should answer future questions or dilemmas. A common example is that of AI machines engineered and trained to detect human sentiments by deciphering the emotion behind certain words. AI Sentiment Analysis can be useful, for instance, because it could understand the intent behind a text, or a social media post or an image/video, by analyzing word structure and syntax or a person’s facial expressions (Mousadakos, 2022). On the other hand, it can be valuable even for everyday things, such as having an application reading a text out loud, which is necessary for people with impaired vision and practical for those who cannot look at a screen for a long time.

According to a 2022 report from renowned Reportlinker.com, ML has been the focal point of growing businesses, as it is the latest tool of predicting business outcomes, saving company money, providing a better experience to clients, and enhancing data security, overall granting competitive advantage to an enterprise (Attaran & Deb, 2018). Moreover, the same source estimates that the return on investment on most standard ML projects in the first year only is 2-5 times the cost. Notably, the global market of Machine Learning as a service is expected to grow in revenue to US\$36.2 billion by 2028 (KBV Research, 2022), while the forecast for the AI global market is estimated to sky-rocket in the next years, from US\$143 billion in 2022 to US\$1,848 billion by 2030 (Thormundsson, 2021) (Image 8).

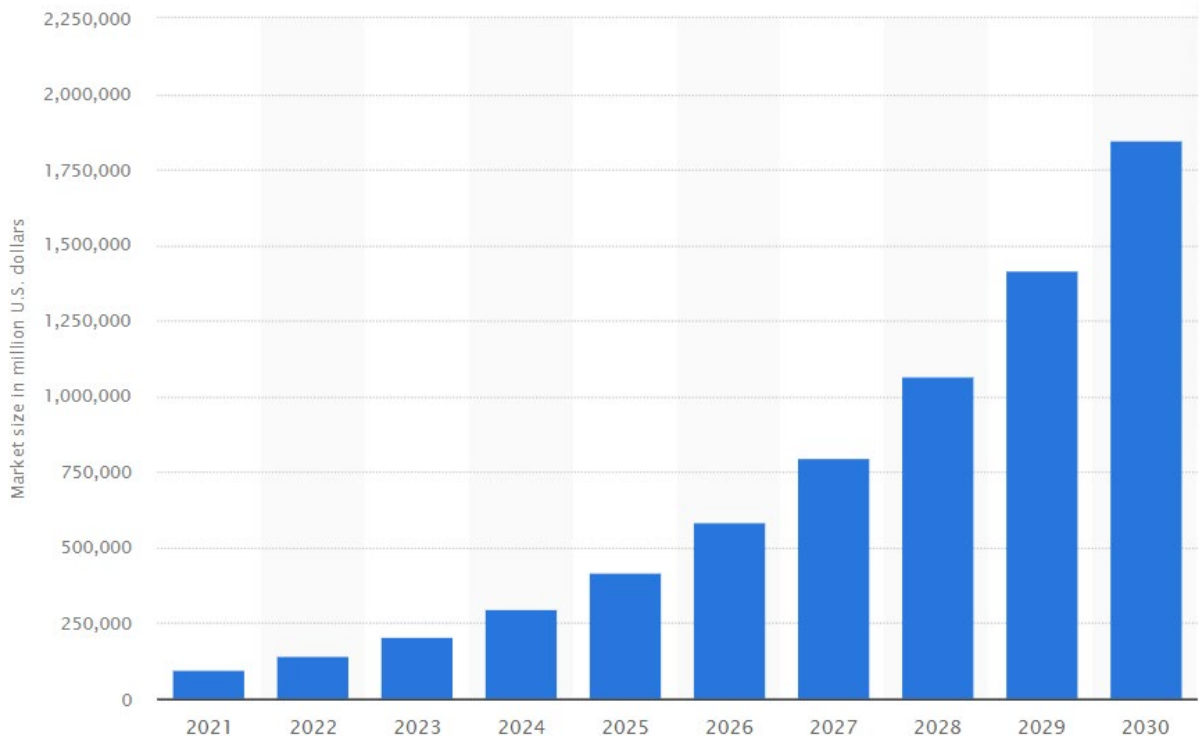


Image 8: Artificial Intelligence (AI) market size worldwide in 2021 with a forecast until 2030
 [Source: Statista, <https://www.statista.com/statistics/1365145/artificial-intelligence-market-size/>]

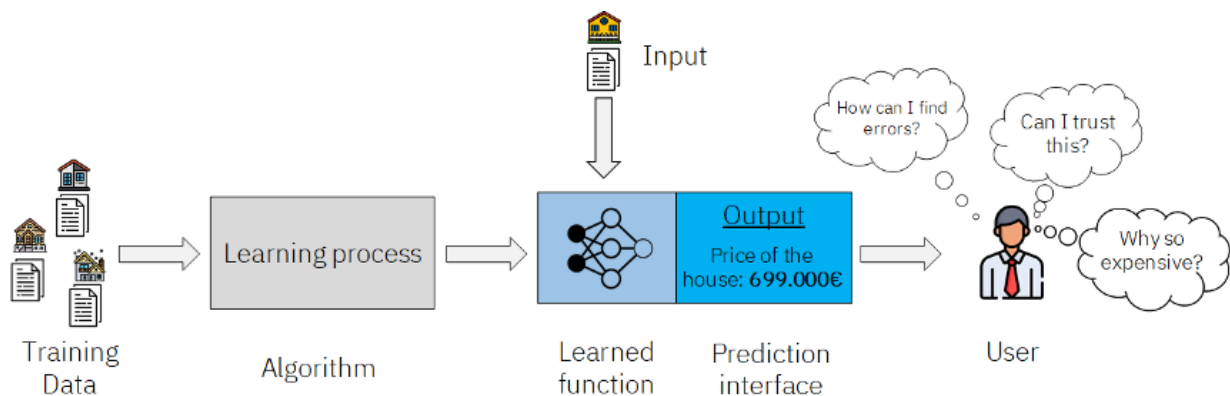
Looking at the impressive numbers above, it is no wonder that these systems' value may look imposing, and their functions omnipotent. And in a sense, one could argue that, as long as they keep ameliorating themselves and humans support and contribute to that, Artificial Intelligence will be the unquestionable center of our future lives, if it already is not the center of our current ones.

Nevertheless, with technology of great importance emerge equally important issues. In a technology relying enormously on data, if the quality of said data is problematic, then the systems will make unfair, biased or simply wrong decisions. One of the most characteristic examples of bias in AI and ML is the use of data lacking representation of certain categories, such as groups of people based on race, gender or sex, age, or culture (Marr, 2022). The insufficiency or absence of inclusive data use while programming the algorithms of Artificial Intelligence software to identify faces, for example, can be critical to how the machine will perceive the figures it detects. For instance, if in the initial programming of the software there was no data (e.g., image) of a man with long hair, then the program would either mistakenly identify a real-life long-haired man as a woman or, if the man also had a beard, it could show an error or skip the sample altogether. Based on how machine learning works, this would also be a paradigm on how to judge the following similar samples; thus, it could lead to multiple wrong decisions and misidentifications.

In a recent series of incidents, the way AI works unfavorably in cases it was meant to be a solution has come forward in a more impactful manner: COVID-19 (SARS-CoV-2) patient risk prediction algorithms. Specifically, in an effort to assess COVID-19 risk more precisely and make patient prioritizing quicker and easier, specialists developed mobile apps and other technologies that would be useful to doctors and patients alike (Delgado et al., 2022). On an individual level, any user could download an application on their mobile device, insert their symptoms and their medical history, and have a ready-made result of how likely they were to need advanced medical assistance. On the same app, they could add their close contacts – acquaintances they could have transmitted the virus to – making COVID-19 cases tracing much easier for the algorithm.

Health apps were used in a wide scale even before the COVID-19 pandemic, and it is expected that they are here to stay (Gruessner, 2015). This has raised questions about how likely a user is to be informed about how each app handles their data. Long privacy policy texts that are largely unintelligible, often ambiguous regarding their legality, and lack transparency on how data like location is handled can become a real issue when the general public uses such applications on a very large scale (Colizza et al., 2021). For this reason, some companies try to facilitate the understanding of which user data is being used by providing tools or settings that allow users to choose what user data they are willing to share with the app, while other companies have implemented technologies that provide real-time information on the use of user data (Harvard Business Review, 2022).

Apart from this kind of concerns, one must be aware that AI's self-training assets may lead to questions regarding the difficulty in the interpretation of the choices AI systems make. This matter, known as the *Explainability* (or "Interpretability") of Artificial Intelligence, is about being able to understand, explain how the AI algorithm's decisions came to be (Samek et al., 2018) (Image 9).



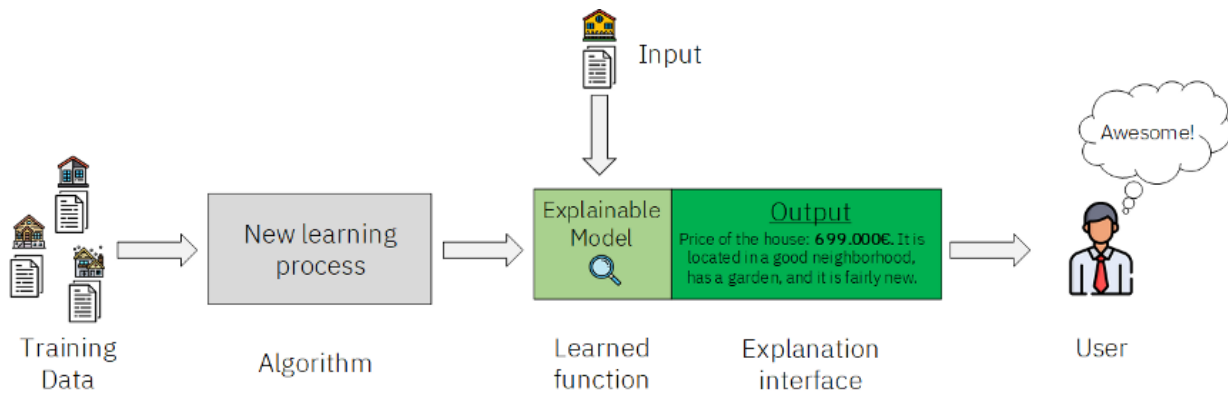


Image 9: ML Workflow. (James Thorn)

[Source: <https://towardsdatascience.com/explainable-artificial-intelligence-14944563cc79>]

Even in cases that the algorithm output (result) is accurate or useful, the obscurity of the decision-making process of the system can lead to confusion as to the *reasoning* behind a choice, why A was chosen instead of B. This opacity of such machine learning techniques has raised the issue of the existence of AI as *black box* (Winner, 1993) – a state where it is not possible to find the origins of the logic that is being followed. This can pose a powerful challenge not only due to the contribution of said machines to research (which requires clarity of methods), and the lack of transparency unpredictability in their future choosing behavior; it can also be problematic, because it could raise critical legal issues: from automated decisions formed on discriminatory variables (e.g., race – which is against the law in most Western countries) to obscure AI medical predictions leading to wrong treatment methods, which can prove fatal for a patient (Yu & Ali, 2019).

Artificial Intelligence is goal-directed (European Union, 2022), which means that it is designed to work on the principle of completing a goal indicated by its programmers. However, depending on the situation and purpose it is used for, it is often left with some freedom in order to ameliorate itself, find different solutions and adapt its approach to problem-solving (Winner, 1993).

Nonetheless, as Technology and its derivatives are human inventions, it is imperative that they serve us, and that we understand how they work. Thus, even though we can have automated systems up-and-running, those should at all times be within our control in case they need to be worked with, reprogrammed, or adapted according to our needs.

4. Examples of Artificial Intelligence misuse

But what happens when Artificial Intelligence and Machine Learning are used to purposefully cause harm? These last few years, it has been noticed that the use of AI to enhance the effectiveness of malware and ransomware is gaining territory. More specifically, in 2015, a report on the technique of *Fuzzing* (algorithm use for checking how a system filters emails and messages, and then adapting system behavior so that all filters and security measures are bypassed) was released, demonstrating how malware could swiftly become a noteworthy challenge in the sphere of AI misuse (Palka & McCoy, 2015).

After having previously mentioned the motives behind misuse of Artificial Intelligence, it would be useful to see the ways AI can be put into action in order to facilitate fraud and other illegal or ethically dubious activities.

In an attempt to state and assess the dangers of AI misuse, Trend Micro collaborated with the United Nations Interregional Crime and Justice Research Institute (UNICRI) and Europol's European Cybercrime Centre (EC3) on a report regarding Malicious Uses and Abuses of Artificial Intelligence. In this evaluation, they were exploring the various possibilities AI and ML could be used in order to commit fraud by exploiting user data, whether it be open or closed. Thus, they found that the most common ways of AI misuse at the moment were linked to AI-supported hacking and password guessing, human impersonation on social networking platforms (e.g., Facebook, Instagram, Twitter, etc.), deepfakes, and AI profiling (Ciancaglini et al., 2020).

AI-supported hacking and password guessing

AI-supported hacking and password guessing tools are used to gain access to and analyze password datasets and generate variations/password guesses that could “unlock” systems. They rely heavily on Machine Learning techniques, as, when a guess is created, the system automatically creates several alternations of that paradigm. In that way, it increased its success rate and “learns” which type of variations are likely to be faster, more accurate, and more effective.

AI-supported hacking and password guessing also includes using natural language processing (NLP) techniques to identify likely passwords from a given set of data. NLP is a branch of Artificial Intelligence that focuses on analyzing and generating natural language (Ramírez Sánchez et al., 2021). It can be used to analyze text, speech, and other types of data in order to extract meaningful insights and generate natural language outputs. This allows the system to comprehend the “depth” of a phrase, detect the feelings or emotions behind a person's choice of words, or even identify similarities in intentions between statements (Pastor-Galindo et al., 2020). Thus, NLP offers a positive outlook for the comprehension of human speech, which may be beneficial in various applications such as customer service, marketing, voice interpretation and suspect profiling (Hernandez et al., 2018), but can be advantageous to cybercriminals as well.

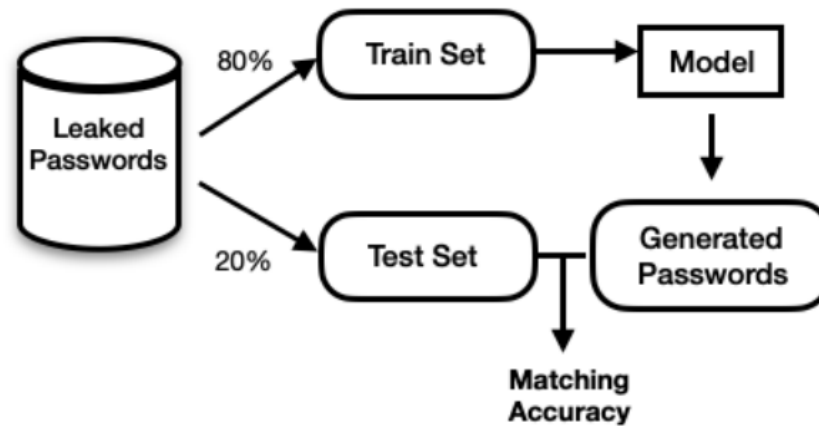


Image 10: Deep Learning in Password Guessing, Data Flow Diagram
 [Source: Yu. F., (2022). *On Deep Learning in Password Guessing, a Survey*. In Proceedings of ACM Conference (Conference'17). USA. <https://arxiv.org/pdf/2208.10413.pdf>]

Human impersonation on social networking platforms

Human impersonation on social networking platforms is in reference to bots (automated systems) created to pose as human agents without being detected. The purpose is, for the most part, increasing traffic on certain social media accounts, to make them appear more popular and successful, leading to recognition among the public and advertisers alike. For example, an Instagram profile with more followers (and thus larger audience) will be more likely approached by a company for an ad deal. Likewise, comments under popular accounts' posts are expected to be read by more people. Thus, celebrity accounts are often teeming with auto-generated comments (spam) with fake ads or services that often lead to internet scams (Kurichenko, 2020).

Human impersonation has also been used as a tool for cyberbullying on social media and other platforms, as well as other forms of harassment. These factors reduce trust among individuals on the internet, and have psychological effects such as anxiety, depression, and self-esteem issues. As such, companies like Meta have been trying to remove those accounts from their platforms, in order to protect the quality of their networks for both the users and their business clients. Characteristically, Facebook took action against 426 million counterfeit accounts just in the first quarter of 2023; this was also surprisingly the period with the *least* actions taken by the company in the past 5 years (Image 10).



Image 10: Fake (spam) Facebook accounts deleted by Meta from 2018 to 2023.

[Source: Meta. <https://transparency.fb.com/data/community-standards-enforcement/fake-accounts/facebook/>]

Such incidents are taken seriously because of their unpredicted implications. For instance, in 2018, an Indian citizen created a fake Facebook profile with the name of a famous singer. This impersonation was used to spread false information and propagate malicious content. The case sparked a major controversy and eventually led to the Indian government passing the *Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021* which more stringently regulated social media platforms (Indian Government Gazette, 2021).

In extension to above, AI is used in CAPTCHA Breaking; bypassing digital (usually image-related) tests used to identify whether the visitor is a human or an automated machine (Kopp et al., 2017), typically by imitating human movements, such as selecting images based on certain characteristics, or recognizing oddly shaped characters.

Deepfakes

Deepfakes, whose name derives from the combination of the phrases “deep learning” (a type of highly accurate Machine Learning) and “fake media” (Westerlund, 2019), are a very special type of technology (mis)use. They are generated by Artificial Intelligence tools designed to create seemingly authentic media, by merging, combining, replacing, and superimposing audio, images and videoclips (Maras & Alexandrou, 2019). Deepfakes have been in the epicenter of everyday entertainment AI because of their various uses – especially since apps like FaceApp enable users to create their own deepfake images or videos (Vincent, 2019). Deepfakes can depict celebrities, politicians or ordinary people doing or saying anything, without a trace of the content being artificial. Their theme is mostly humorous, pornographic, commercial, or political, and they can easily be worked with, by providing the algorithm with simple data (e.g., an image) that can even be found online (e.g., open data from Twitter or Instagram).

The widespread usage of deepfakes has sparked controversy and rekindled the discussion concerning how Technology intervenes with reality and our perception of it (Van der Sloot & Wagenveld, 2022). At times when images and videos reach us by hundreds on a daily basis, it is no surprise that deepfakes are used

largely as a propaganda method, especially in authoritarian regimes or underdeveloped countries, where the citizens have limited means of information verification. In a relatively recent example, in March 2022 (shortly after the Russian invasion in Ukraine) a deepfake of Ukraine's President Volodymyr Zelenskyy urging Ukrainian soldiers to “*lay down arms and return to [their] families*” was released (Image 11) (Metz, 2022). The video was quickly identified to be counterfeit, and large social media platforms such as Facebook and YouTube were forced to quickly remove it, in an attempt to stop the misinformation from further spreading.

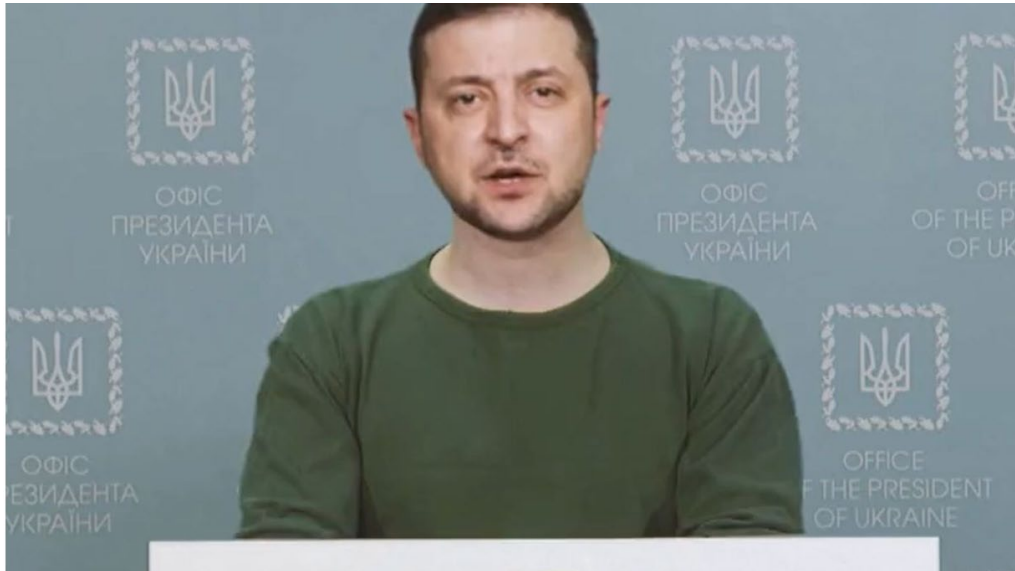


Image 11: Screenshot from the deepfake video showing Zelenskyy addressing Ukrainian soldiers
 [Source: The Telegraph on YouTube
https://www.youtube.com/watch?v=X17yrEV5sl4&ab_channel=TheTelegraph]

Even though this technology is also being used for entertainment purposes, it is the expansion of deepfake abuse that has brought forward several social and legal issues regarding the use of data and its manipulation to target certain groups of people. Characteristically, a report by Sensity AI, *The State of Deepfakes 2019 Landscape, Threats, and Impact*, found that 96% of deepfakes were non-consensual pornographic deepfakes; in 99% of those, the victims were women (Dunn, 2021).

Deepfakes constitute part of a special type of Artificial Intelligence: Generative AI. These models are typically based on neural networks, which are computational models inspired by the structure and function of the human brain (Müller et al., 1995). Generative AI algorithms can (with or without human contribution) generate content, such as text, images, videos, audio and even code lines used to program software (Lyon & Tora, 2023). Their ability derives from the extensive training they receive, after being given and processing enormous amounts of data (e.g., millions of images, texts, etc.), and “learning” to edit said data to generate content that is as realistic or as creative as possible. Seen as an opportunity by tech-enthusiasts yet potentially problematic by many, Generative AI has also been dominating the discussion of modern technologies (Hurst, 2022). The rise of deepfakes in general, and platforms such as ChatGPT (a system that is able to generate complex texts upon given prompts by humans) and DALL-E (image generating platform, that creates images based on a human prompt) more specifically have brought forward the question (or rather concern) of whether we will soon not be able to distinguish between genuine and made-up content on the internet.

AI profiling

AI profiling, another prominent method of Artificial Intelligence misapplication, is used to extract data from online sources, and put them into databases in a structured format (Mayer-Schönberger & Cukier, 2013). This data is then used to analyze and create profiles of individuals or groups based on various data points and characteristics. It involves gathering and processing large amounts of data, such as online behavior, social media activity, purchasing patterns, and demographic information, to create detailed profiles that can be used for various purposes.

Perhaps the most infamous case of AI profiling is the Cambridge Analytica – Facebook scandal, which led to data exploitation of data of millions of Facebook users. The story erupted in March 2018, when news broke that the data analytics firm had illegally obtained data from a total of 87 million Facebook users (Image 12) (Zuboff, 2019). This data was then used by the company to build psychographic profiles of users and target political ads to them with the help of Artificial Intelligence algorithms (Boerboom, 2020).

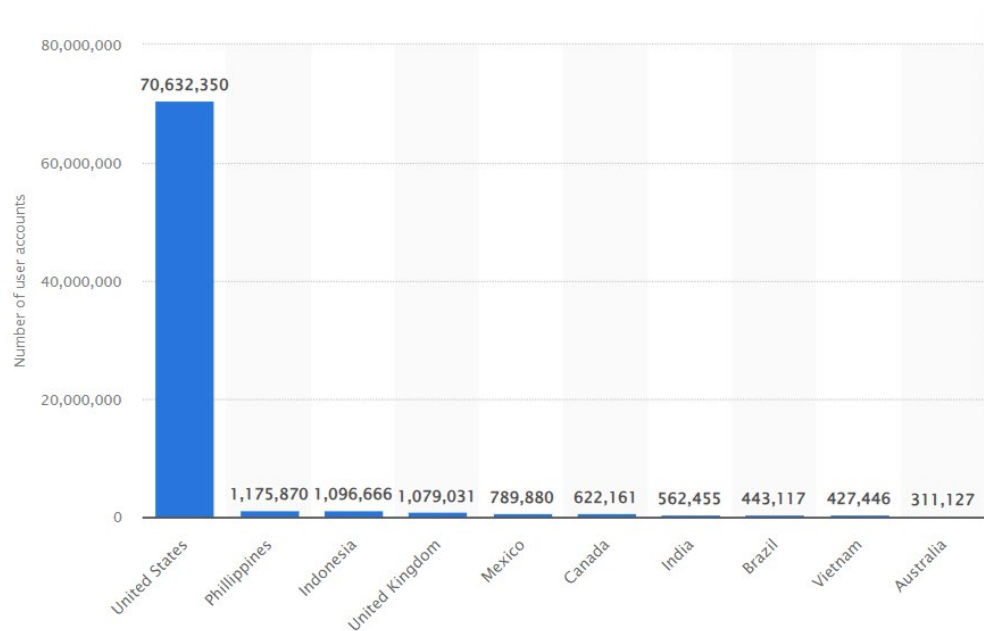


Image 12: Number of Facebook user accounts that may have been compromised in the Cambridge Analytica scandal as of April 2018, by country

[Source: Statista, <https://www.statista.com/statistics/831815/facebook-user-accounts-affected-cambridge-analytica-by-country/>]

While AI and machine learning algorithms can play a positive role in analyzing and processing such large amounts of data, the misuse of data in the Cambridge Analytica scandal was primarily an issue of data privacy and consent. The scandal sparked a tremendous outcry in the media, leading to investigations by regulatory bodies, multiple legal cases (Schneier, 2015), and even a congressional hearing, as the data was used to potentially sway the 2016 elections in the United States of America, by targeting political advertisements to certain groups of people (Chan, 2020). It had a significant impact on a social level, as the uproar raised questions about the use of user data for political purposes and the implications of insufficient data privacy. This has led to greater public scrutiny of tech giants like Facebook itself, regarding the way companies collect and use data for their own gains.

Finally, the scandal cast a shadow on people's trust in democratic institutions; MEP Sven Giegold, who at the time was the European Parliament's spokesperson for transparency, accountability and integrity in the EU institutions, described it as "[...] a stab in the heart of democracy. It is about much more than breaches of data protection rules; it is about the legitimacy of democratic elections and referenda." (Banks, 2018).

These types of Artificial Intelligence misuse revolve around data and the internet: the network of networks (Lambert et al., 2005). In what seems to be an alarming escalation, our need for the internet in general – and Data and AI in Social Systems: Trust Perspectives in a Digitally Developing World

social media use in particular – for both personal and business purposes is not simply ubiquitous; it is rather a prerequisite to exist as an active part of today's society. In this context, it would be in our best interest to reflect on the practical hazards that our unconstrained online presence entails, as well as the ethics behind the importance of human data management and protection.

PART TWO – Modern societies constructed around data

As described in the previous chapters, data plays a tremendous role in today's world. It is crucial for science and technology, as it helps us understand trends and make better decisions, but it can also be used to create predictive models, detect fraud, and improve cybersecurity. It is used in almost everything from healthcare to financial services and marketing. With the rise of analytics (the process of analyzing data), data becomes even more valuable as it helps us see patterns or gain insights that can lead to defining future strategies. Data is there, from the simple social media update to the organization of business client lists, and Government digital records; it is the axis on which we rely the organization of modern societies.

5. The Business Aspect of Data

For companies and businesses, data can be used to analyze and personalize experiences; from personalized ads and product recommendations to enhanced customer service (Chavez et al., 2018). It can also be used to identify patterns and trends in user behavior, create insights, and increase the success rate of marketing campaigns (data-driven marketing) (Jeffery, 2010). Thus, it can drive innovation and contribute to uncovering new opportunities for companies, the players of the global market.

With the increasing use of internet-connected devices and the internet of things (IoT), data can also be valuable in everyday life. Generated and collected from a variety of sources (smartphones, smart home devices, wearable devices, etc.), this data can be of value for personal health and wellness (e.g., tracking food intake, exercise and activities) (Aboelmaged, et al., 2022), financial management (tracking expenses, setting up budgeting apps) (Velmurugan et al., 2020), home automation (e.g., remote control of domestic lighting and heating) (Ramzan, 2020), and more. However, users are often unaware that the devices that collect this data usually send direct feedback to their mother-companies, so as to ensure app performance, but to contribute to business analytics as well (Mourtzis et al., 2017).

Companies can also obtain data from various sources, such as cookies on websites (small text files stored on a user's device when they visit a website) (Cisco 2018), surveys, third-party vendors, and of course social media platforms and online account information (OECD, 2015). Social media platforms can be a particularly rich source of data for companies, as user data is open and readily available. Social media platforms allow users to share data and information about themselves, their interests, and their activities, which can provide excellent material for business analysts (United Nations, 2020). For example, by tracking a user's activity on social media, it is possible to gain insight into their interests, preferences, and behavior, which can be used to deliver more relevant advertisements and content.

The process of collecting and analyzing information about an individual based on their social media activities and interactions, such as posts, comments, likes, shares, and other publicly available information is known as *social media profiling* (Mitrou et al., 2014). In other words, the purpose of social media profiling is to create a digital profile or portrait of an individual based on their online behavior and data, which can then be used to decipher an individual's personality for marketing, online advertising, research, and other professional purposes – similar to how Cambridge Analytica handled Facebook users' data.

Profiling is also a useful tool for employers in the process of screening job applicants and evaluating potential job candidates. According to a CareerBuilder survey in 2018, 70% of employers visit and examine applicants' social media profiles as part of their screening process (CareerBuilder, 2018). Employers can use an applicant's profile and posts on social media to explore a candidate's job-related skills, background, and previous workplaces. Such data can be helpful in understanding a candidate's qualifications and potentially predicting job performance or success. There are also platforms, such as LinkedIn, specifically dedicated to promoting one's job skills, where candidates and employers alike create profiles and professional networks (Holyoake, 2021). Via these websites, anyone can have instant access to the work history of a candidate employee – as long as the latter has taken the time to update their profile and upload relevant certificates.

Moreover, profiling is practical for insurance companies that assess risk associated with insuring individuals based on demographic information (age, gender, socioeconomic status, city of residence etc.) or human behavior (Marr, 2015). For example, a person signing up for health insurance might claim that they are not a smoker, so as to avoid a higher insurance fee. Nevertheless, if a photo of them holding a cigarette can be found online (either because the person uploaded it themselves or they were "tagged" by others), then the insurance company can prove their initial claim wrong and raise the price of the health insurance package. Another case in point would be someone applying for car insurance expecting a reasonable rate, only to end up with a higher premium due to the insurer discovering social media evidence of the candidate taking photos or videos while driving (Tepedino, 2022).

There have been concerns that this individualistic approach to pricing risks in combination with the growing amount of data available to everyone online, could *render certain groups of people increasingly uninsurable* (Smith, 2021). However, the clear asset remains that by using social media data, insurance companies can increase their efficiency by providing products and services which will meet customers' needs.

Lastly, yet perhaps most importantly, profiling based on social media is an integral part of modern-day law enforcement, particularly in crime investigation, gathering evidence, suspect identification, and background check, as well as crime prevention (LexisNexis, 2014). This particular matter is also subject to legal and ethical considerations, and it has sparked a lot of debate as the practice is becoming more common, one of the reasons being the high social media activity by a large part of the world's population. Characteristically, a 2014 LexisNexis online survey among law enforcement professionals found that 81% of them actively used social media (mainly Facebook, YouTube and Twitter) as a tool in investigations. In the meantime, over half (52%) of agencies did not have a formal process for the proper use of social media for investigations at that time, and the majority of the officers were mostly self-trained on how to use social media to investigate crimes (Image 13).

How did you learn or discover how to use social media in your investigations or crime monitoring activities?

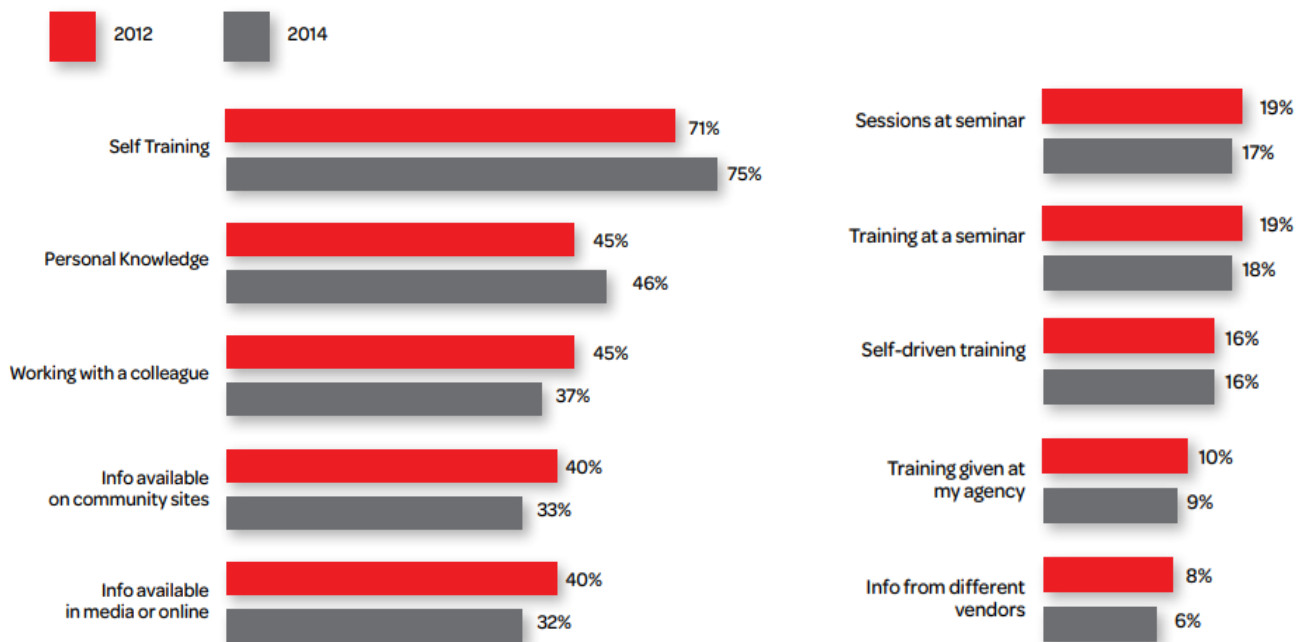


Image 13: How did you learn or discover how to use social media in your investigations or crime monitoring activities? LexisNexis® Risk Solutions

[Source: Survey of Law Enforcement Personnel and their use of social media]

When it came to moral dilemmas, 82% claimed that “social media is a valuable tool in investigating crimes”, while 80% stated that “creating personas or profiles on social media outlets for use in law enforcement activities is ethical”; 9% disagreed with this practice, while 11% was neutral about the matter (Image 14).

Importance and value of using social media to investigate crime

The value of social media in investigations, both now and in the future, is abundant. Law enforcement professionals often leverage personas and undercover accounts during investigations to follow subjects or collect evidence.

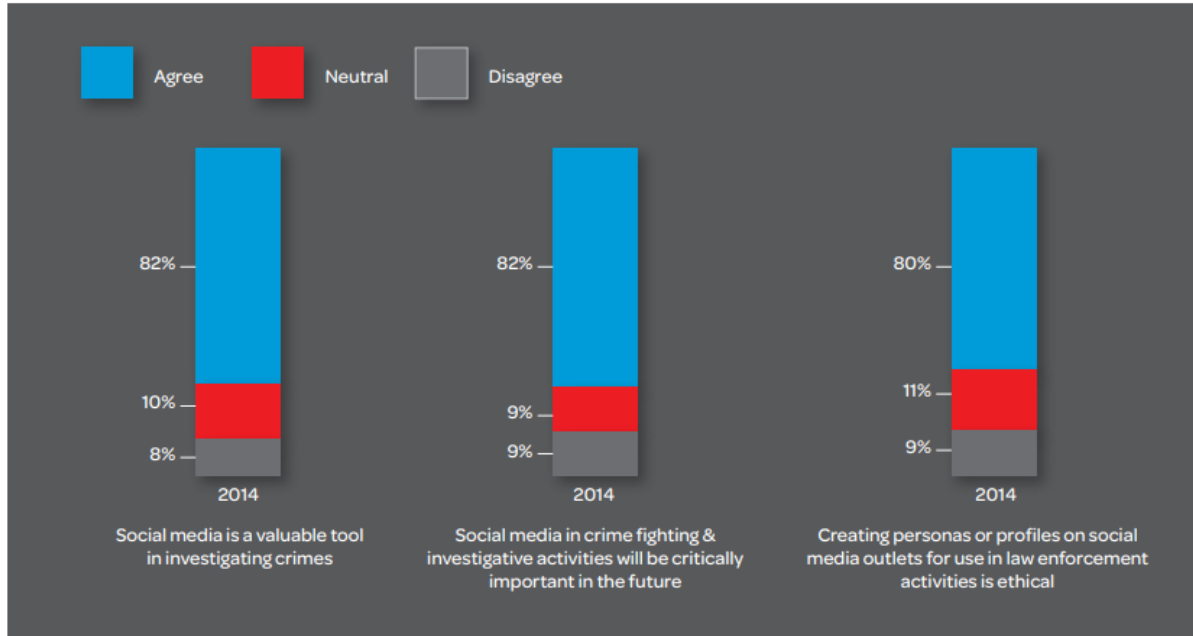


Image 14: Importance and value of using SM to investigate crime LexisNexis® Risk Solutions [Source: Survey of Law Enforcement Personnel and Their Use of Social Media]

Summing up all the points made above, it is clear how multiple aspects of modern-day life are affected by data and its use, from everyday activity to commerce, and from business to law and politics. After having explored the financial aspect of data and briefly mentioning the current state of affairs, in the next part of this dissertation, the focus will shift to the other side of the coin: politics and law.

6. The Political Aspect of Data

a. Liberal Democracies

Even though human rights are considered a core social value in the political West, profiling via social media monitoring is no stranger to authorities even in the most liberal European countries, such as the Netherlands. The Dutch police have a specialized crime-prevention team that monitors social media platforms like Facebook, Twitter, and Instagram to identify potential criminal activity (Government of the Netherlands website, 2023). The team uses a combination of both automated and manual tools to identify patterns in social media activity that may indicate criminal behavior, from drug dealing to committing tax fraud (Meijer & Torenvlied, 2016). But they may also monitor platform activity during events such as protests, demonstrations, and populous events to identify potential threats and prevent violence (Statewatch, 2023).

Consequently, the use of social media monitoring by law enforcement has resulted in newly raised privacy concerns that appeared futuristic a few years ago. In 2020, a Dutch court ruled that the police had violated the privacy of citizens by monitoring their social media accounts without a clear legal basis (Henley & Booth, 2020). This violation took place on the pretext of combating terrorism and organized crime yet was exposed as insufficient during the legal hearing, with the court noting that the police had not obtained the necessary authorization from the court of justice or demonstrated a legitimate interest in monitoring the accounts of individuals who posed threat for the state. The ruling highlighted that the police must only monitor social media accounts in clear accordance with the Dutch law and declared that there must be supervision and discipline to ensure that the process itself is conducted in a lawful and responsible manner.

Regardless, it appears that the debate keeps resurfacing in other countries that pride themselves in protecting sensitive citizen data as well. The upcoming Olympic Games in Paris in 2024 have prompted the French government to implement a range of security measures, including public surveillance (Kayali, 2023). In particular, the French authorities are planning to deploy a network of surveillance cameras to enhance security during the games. This surveillance system will be developed by the French Ministry of the Interior and will involve the installation of closed-circuit television (CCTV) cameras and deployment of drones throughout Paris and locations aiming to host the Games. The cameras will be linked to a central control center, where police officers will be directly able to monitor public spaces and identify potential security threats, much like in the case of Dutch police.

However, in addition to the cameras, the French authorities were also planning to use live human recognition technology, which would be used to identify individuals and “suspicious behavior”, as well as unsupervised luggage and alarming crowd movements like stampedes⁸⁶. The main argument was that in this way it would be much easier to detect and identify persons who may pose a threat to public safety, such as criminals or suspected terrorists, and have the police promptly intervene in case of emergency.

There certainly are well-founded and valid reasons why French authorities attach such high importance to the threat of terrorism. In less than a decade there have been multiple terrorist attacks on French territory: from the French satirical newspaper Charlie Hebdo shooting and the Bataclan Paris attacks in 2015, to assaults in worship spaces (Normandy in 2016, Nice in 2020), and places that everyday life takes place (Nice riviera in 2016, Champs-Élysées in 2017, Strasbourg Christmas market in 2018, Professor S. Paty's murder near his school in 2020), it would appear that every aspect of French - and thus western - lifestyle has been under attack and needs to be protected (SkyNews, 2020).

Nonetheless, the use of AI-assisted monitoring technology in public spaces has still raised concerns among privacy advocates and civil liberties groups, who argue that it represents a significant intrusion into people's privacy and could be used for mass surveillance in the long term (Chrisafis, 2023). When asked about the issue, Katia Roux, advocacy officer for technology at Amnesty International, expressed the opinion that this type of surveillance could be used to *target specific groups* (e.g., people who are - or look like they are - from the middle East) and that it would *infringe the right to privacy and peaceful assembly* (Khatsenkova,

Data and AI in Social Systems: Trust Perspectives in a Digitally Developing World

2023). Finally, more objections were voiced regarding this legislation, since many claim that it would pave the way for the establishment of live facial recognition technology in France. This is a particularly sensitive subject, as it parallels European Democracy to authoritative regimes and countries like China or Russia; countries that have already publicly introduced face recognition technologies with the pretext of hosting large public events (Rosenberg, 2018).

In response to these concerns, the French government has stated that facial recognition technology will not be used, and that the state will ensure that its use is strictly regulated. It has also pledged that appropriate safeguards are in place to protect citizen rights, and that the surveillance equipment will be used only temporarily until June 2025, 10 months after the Olympic Games come to an end (Kayali, 2023). The bill to allow cameras to be installed for security purposes was finally approved in May 2023, with a special mention to the prohibition of real-life facial recognition (O'Carroll, 2023).

b. Authoritarian Regimes

Contrary to how heated this debate is in countries where liberal democracy is established and technology policies change often depending on the party (or parties) in power, things are different in states with a more centralized system of governance where no substantial legal political opposition groups exist. In China, for example, a country governed by a single political party (the Chinese Communist Party – CCP) and one omnipotent ruler (currently Xi Jinping), the government has the power to silence doubting or opposing voices as it pleases. The case of Chinese doctor Li Wenliang, who first discovered COVID-19 in its early stages and tried to warn his compatriots only to be arrested by the Chinese state, is one of the numerous cases in the land of the Red Dragon (Su, 2020). Therefore, there is solid proof that the state is actively trying to control the circulation of information in the public sphere.

In more depth, in recent years, the Chinese government has been developing and implementing a model of social organization based on the continuous monitoring of citizens and their complete control by the state. Through an integrated surveillance system with applications of Artificial Intelligence, it seeks to influence or control every aspect of daily life and individual identity and impose a totalitarian model of governance unprecedented in human history.

Chinese president Xi Jinping has stated that China should have achieved global supremacy in Artificial Intelligence (AI) by 2030, pushing it to the "pinnacle of surveillance", creating "a comprehensive digital social control system, patrolled by predictive (precog) algorithms that identify potential dissenters in real time" (Andersen, 2020). The government has realized that the digitalization of Chinese society enhances the state's ability to monitor and control the country's 1.4 billion people (Qiang, 2019), and prevents potential culprits from acting, since they can easily be spotted (Yang et al., 2017). More importantly however, achieving (global) leadership in intelligence appears to be the main strategic goal of the Xi administration, and this has significant implications not only domestically, but also for global security and the way war is waged in the 21st century (Allen, 2019).

Integrating material from facial recognition cameras, biometrics and mobile phone data through AI allows the government to take *proactive* measures to stop protests and quell political dissent before it gathers further support (Qiang, 2019). With more than half of the world's one billion CCTV cameras being in the country, China is the most surveilled country in the world (Bischoff, 2022), and the country that collects and stores the most citizen biometric data (Image 15). Meanwhile, it also appears to be the world's most unregulated and intrusive user of biometric data (Zhang, 2021). Thus, it is no surprise that China scores the worst in collection and handling processes, with a lack of safeguards to protect citizens from the use of data by government authorities and private employers.

Collection and storage of biometrics by country

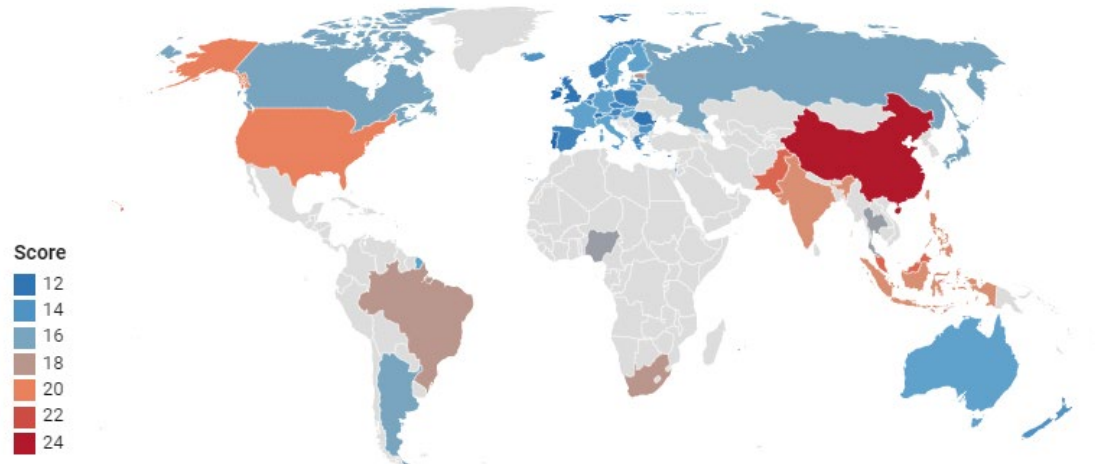


Image 15: Collection and storage of biometrics by country, 2022.

[Source: Comparitech <https://www.comparitech.com/blog/vpn-privacy/biometric-data-study/>]

Emotion recognition technologies, such as the system developed by Chinese firm Taigusys, go beyond facial recognition: capturing expressions of anger, sadness, happiness, and boredom, and combining these with other biometric data, such as body temperature, body movements facial muscles, tone of voice and body movements can tell a person's emotions (Standaert, 2021). Among other potential applications, these technologies are used in prisons to monitor inmates, who, knowing they are under constant surveillance, tend to be more compliant. Factories, state-owned enterprises and even the military have also begun monitoring workers' brain activity using "mind-reading hats" to detect signs of anger, anxiety or depression in the workplace (Thomson, 2019).

The combination of CCTV and biometric data through systems such as SkyNet has many other applications in everyday life. State-owned workplaces and workers' housing are now equipped to monitor and control in real time who enters or exits and how much time they spend in each area. Outside some public toilets, vending machines scan a user's face and dispense a small amount of toilet paper (each person is only entitled to a certain amount of paper) to reduce wastage. Surveillance cameras scan the faces of pedestrians crossing the red road and link to authorities' databases to identify them; immediately afterwards they put their faces on giant screens to publicly shame them (Jiang, 2020).

While each of these surveillance systems alone has significant implications for privacy, liberty and human rights, their integration through AI has the potential to profoundly transform a state's control over its citizens, but also to citizens of other countries (Image 16).



Image 16: Chinese Surveillance Technology Spreads Around the World
[Source: Statista, with Data from CEIP, <https://www.statista.com/chart/20221/origin-of-ai-surveillance-technology-by-country/>]

Xi's government aims to achieve full video coverage of public space; in this way, every person who enters a public space will be instantly recognizable: the AI application will identify them by combining their image with personal data such as their text messages and the protein structure of their body. All of this can be collected and processed through technologies such as City Brain – an AI platform that enables new forms of integrated surveillance (Caprotti & Liu, 2022) (e.g., combining this information about people with data collected by self-driving cars whose sensors show authorities a real-time car-map of the city) (Marvin, 2022).

It appears that an authoritarian regime of such extent could expand its control beyond its borders through infiltrating the public and private spheres of countries around the world: both through political, economic and industrial influence, and through elite and societal influence operations, but also through the collection and processing of personal data of citizens of other states (Sullivan, & Brands, 2020). The role of Chinese-influenced or owned companies and apps widely used in the West, such as Huawei, Zoom and TikTok, is currently under investigation (Bartz, 2020, Gerodimos et al., 2023). This is not only because of the financial assets that emerge from data accumulation (as explored in the previous chapter), but also for reasons of political nature, such as China amassing intelligence on other countries' citizens, which could lead to multinational security concerns. Lastly, the above puts into perspective the legal framework regarding human and civil rights protection, especially in countries that organize their societies based on the Rule of Law

principle: the principle that no individual or entity is above the law and that everyone, regardless of their status or position, are subject to the same legal principles and procedures.

7. The legal aspect of data

Data might be a crucial asset in the modern world, but its use raises significant legal challenges and concerns. In most countries where the use of technological and data-centered means is broad, data is protected by privacy laws aiming to ensure that it is not shared without consent or abused. More than 150 countries around the world have implemented data protection and privacy laws. By 2021, approximately 71% (137 out of 194) of countries globally have established some form of national data protection legislation (United Nations, 2023).

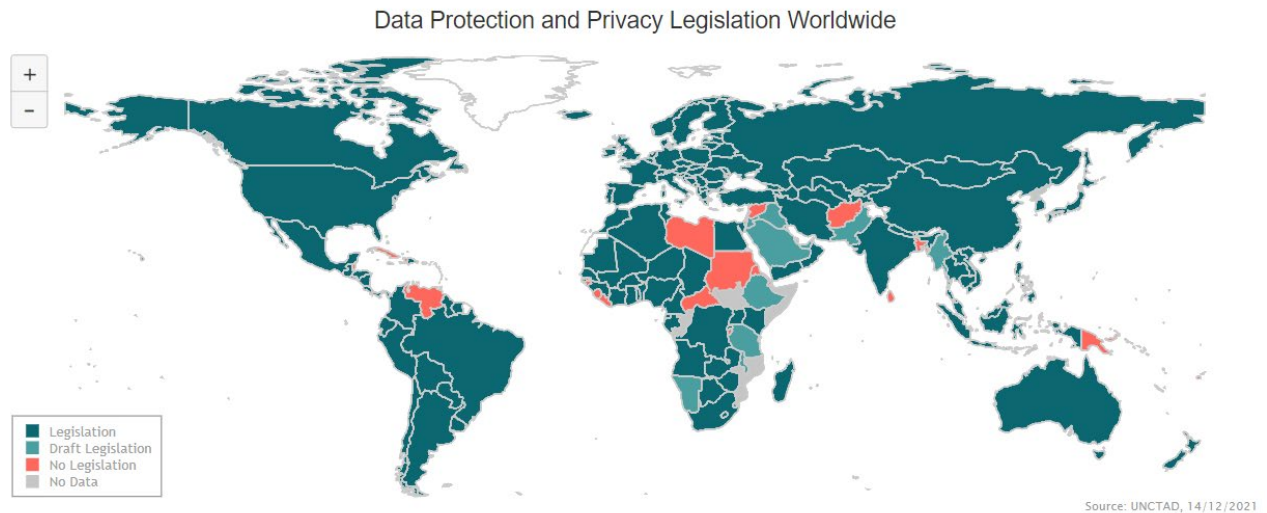


Image 17: Data Protection and Privacy Legislation Worldwide in 2021 (United Nations)
 [Source: <https://unctad.org/page/data-protection-and-privacy-legislation-worldwide>]

Data can be subject to data security laws, breach notification laws and other regulations. The exact legal implications of its usage depend on the country of jurisdiction, meaning that different countries have different laws regarding data privacy and protection.

For instance, in China – as described in the previous chapter – the legal system is used to maintain political control of the state, and the laws around data may serve the interests of the ruling regime rather than the protection of individual rights. The country has established several data protection laws in recent years, including the Personal Information Protection Law (PIPL, 2021) and the Cybersecurity Law (2016). PIPL is designed to protect the personal information of citizens living in China, while requiring *companies* to take appropriate measures *to protect personal information* and privacy (Zhu, 2022). At the same time, the Cybersecurity Law gives the Chinese *government* power over the ownership, storage, transfer, and handling of personal data generated in China by Chinese and foreign entities (Maranto, 2020).

However, these laws have been criticized for not providing adequate protection for citizens' rights (Amnesty International, 2022), especially since the Chinese government has often been accused of using data for surveillance and censorship purposes. Characteristically, Article 28 of Cybersecurity Law of the People's Republic of China (National People's Congress of the People's Republic of China, 2016) dictates that *network operators* must cooperate with public security organs such as the Ministry of Public Security and *provide technical assistance* (such as equipment or material like data) upon request:

Article 28: Network operators shall provide technical support and assistance to public security organs and national security organs that are safeguarding national security and investigating criminal activities in

accordance with the law.

— Section 1: 'Ordinary Provisions'

However, the term “*network operators*” is quite vague, since it is typically used with reference to telecommunication companies. In the Cybersecurity Law it is interpreted to include social media platforms and application creators instead, thus enabling the state to monitor people’s online activity pleading national security reasons. What renders the above alarming is the way the Chinese Communist Party's Political and Legal Affairs Commissions coordinate and directly control the entire Judicial system of China (Peerenboom, 2010). Lastly, the fact that the Cybersecurity Law (the law allowing the Chinese government to access and handle all data generated in the country) was established 5 years before PIPL (which is a general data protection framework), amplifies the sentiment that in Chinese political affairs, there is no Individual when the State mandates so.

When examining laws in liberal democracies, the approach appears to be different. In particular, the United States of America do not have a single unified data privacy law. At the federal level, the main law governing data privacy in the US is the Health Insurance Portability and Accountability Act (HIPAA, 1996), which applies to the healthcare industry and regulates the use and disclosure of patients' health information. It gives individuals certain rights, such as the right to access their personal health information (private data), the right to receive a list of disclosures of said information from the insurance company, and the right to request that their data be corrected.

Although the US government has not established broad federal laws, the USA do have several laws at state level. One of the most known set of laws is California's Consumer Privacy Act (CCPA, 2018), which provides consumers with the right to know what personal data is being collected by businesses, the right to request deletion of personal data, and the right to opt-out of the sale of personal data to third parties (CCPA, 2018). The CCPA also imposes obligations on businesses with regards to how they collect, use, and disclose personal information. It requires that the companies provide certain disclosures to consumers and abide by specific security measures to protect personal information. Specifically, the *right to know* which data is being stored and used by health providers, the *right to delete* personal information held by businesses, and the *right to opt-out* of sale of personal information are explicitly mentioned and protected.

On the other side of the Atlantic, by 2016 the European Union had already established some of the strictest data protection laws in the world, known as the General Data Protection Regulation (GDPR – implemented 2018). The GDPR regulates how organizations *in the EU or partnering with the EU* process and store personal data, and it is considered that it revolutionized personal data protection and digital privacy. Focusing on certain key privacy protection points (Image 18), it strictly notes that if a person or entity processes data, they have to do so according to seven protection and accountability principles outlined in Article 5.1-2 (Regulation (EU) 2016/679, 2016):

- i. *Lawfulness, fairness and transparency* — Personal data must be collected and processed in a lawful, fair, and transparent manner with regards to the data subject.
- ii. *Purpose limitation* — Personal data should only be collected for specific, explicit, and legitimate purposes, and should not be used for any other purposes without obtaining additional *consent*.
- iii. *Data minimization* — Only the minimum amount of personal data necessary for the specific purpose should be collected and processed.
- iv. *Accuracy* — Personal data must be accurate, and efforts should be made to keep it up-to-date.
- v. *Storage limitation* — Personal data should not be kept longer than necessary for the specific purpose for which it was collected.
- vi. *Integrity and confidentiality* — Processing must be done in such a way as to ensure appropriate security, integrity, and confidentiality (e.g., by using encryption).
- vii. *Accountability* — The entity collecting and processing personal data is responsible for ensuring compliance with the GDPR and must be able to demonstrate compliance upon request.

Bigger Responsibility, Bigger Repercussions



Image 18: GDPR key take-aways [Source: <https://www.emotiv.com/glossary/gdpr/>]

The GDPR recognizes that the protection of personal data is a fundamental right and that individuals have the liberty to control their personal data and know how it is being used (Treaty on the Functioning of the European Union Reference: 2012/C 326/01). It also gives them the right to request that their personal data be corrected, and the power to object to the processing of their data with the exception of certain circumstances (e.g., vital interests, such as hospital admissions, etc). Thus, entities collecting or processing one's personal data have to explicitly inform and ask for the individual's *consent* for any action or decision made that involves said data, while also being under the obligation to implement security measures to protect it from unauthorized access, disclosure, or loss. Furthermore, entities must report any data breaches that result in the unauthorized access, disclosure, or loss of personal data to the appropriate authorities within 72 hours of becoming aware of the breach.

Notwithstanding, the most significant innovation introduced by the GDPR in a legal basis is the *right to erasure* – also known as *the right to be forgotten* – under which individuals have the right to request erasure of their personal data from private and even government entities (*controllers*) :

The data subject shall have the right to obtain from the controller the erasure of personal data concerning him or her without undue delay and the controller shall have the obligation to erase personal data without undue delay where one of the following grounds applies:

- i. the personal data are no longer necessary in relation to the purposes for which they were collected or otherwise processed;*
- ii. the data subject withdraws consent on which the processing is based according to point (a) of Article 6(1), or point (a) of Article 9(2), and where there is no other legal ground for the processing;*
- iii. the data subject objects to the processing pursuant to Article 21(1) and there are no overriding legitimate grounds for the processing, or the data subject objects to the processing pursuant to Article 21(2);*
- iv. the personal data have been unlawfully processed;*
- v. the personal data have to be erased for compliance with a legal obligation in Union or Member State law to which the controller is subject;*
- vi. the personal data have been collected in relation to the offer of information society services referred to in Article 8(1).*

– Art. 17 GDPR – *Right to erasure ('right to be forgotten')*




The term "government" is not specifically used in the official text regarding the right to erasure; however, the GDPR's definition of "controller" (Article 4) includes "public authorities" and "government bodies," which means that the right to erasure applies to these entities as well. Unlike Chinese or US legislation, the European Law creates a shield of protection even against government mishandling and exploitation of personal data. It also gives individuals the ability to manage their digital identities, which are increasingly becoming an integral part of modern human societies as their use grows (Image 19).

COMPARING STANDARDS

The new Chinese data laws offer some similar protections to Europe's GDPR and the US CLOUD Act

Key data protection standards:

● Absent ● Partially addressed ● Addressed

| Strength of the general legal framework for personal data protection | | | |
|--|--|---|--|
| |  China |  European Union |  United States |
| Privacy protected by constitution | ○ ○ ● | ○ ○ ● | ○ ○ ● |
| Definition of personal data | ○ ○ ● | ○ ○ ● | ○ ○ ● |
| Legal framework for personal data protection and coherence | ○ ● ○ | ○ ○ ● | ○ ● ○ |
| Independent data protection agency | ● ○ ○ | ○ ○ ● | ○ ● ○ |
| Civil society participation in regulations and oversight | ● ○ ○ | ○ ○ ● | ○ ○ ● |
| Possibility of consumer lawsuits | ○ ○ ● | ○ ○ ● | ○ ○ ● |

Source: Mercator Institute for China Studies

Image 19: PIPL – GDPR – US data protection legislation comparison (MERICS)
 [Source: <https://www.merics.org/en>]

In the meantime, there has been an attempt to put some restrictions on surveillance and monitoring on a European level: after establishing the GDPR in 2016, the European Parliament requested that the European Commission introduce a legislative framework on AI as early as 2017 (Banks, 2021). Moreover, in 2021 the European Data Protection Board (EDPB) and the European Data Protection Supervisor (EDPS), called for a *ban on use of Artificial Intelligence for automated recognition of human features in publicly accessible spaces, and some other uses of AI that can lead to unfair discrimination*. Lastly, in 2023, these ideas were incorporated in the (upcoming, as of June 2023) Artificial Intelligence Act (AI Act, Image 20), a set of regulations aiming to regulate AI tools and service providers. Its goal is to manage such tools by categorizing and evaluating them according to the considered level of risk this technology could pose, from a scale of "minimal risk" (allowed to be used) to "unacceptable" (outright banned by law). "High risk" AI would include tools used in law enforcement (such as the cases of the Netherlands mentioned previously), or education (such as ChatGPT, the text Generative AI model).

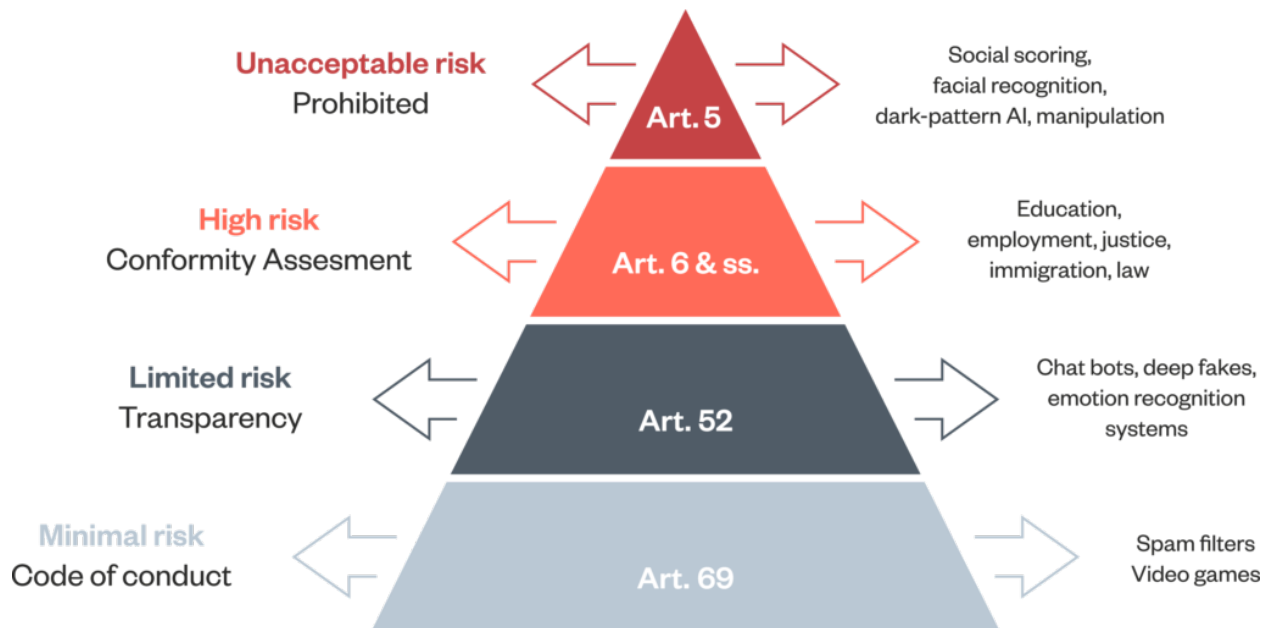


Image 20: The EU AI Act proposal. Structure: a 'risk-based' approach. Ada Lovelace Institute.
 [Source: <https://www.adalovelaceinstitute.org/resource/eu-ai-act-explainer>]

The GRPR has also inspired several countries around the world to strengthen their data protection laws or introduce new legislation. Brazil's General Data Protection Law (LGPD, 2018) was heavily influenced by the GDPR and includes similar provisions, such as the right to access personal data and the right to erasure. At the same time, South Korea (Personal Information Protection Act – PIPA, 2020), Japan (Japan Act on the Protection of Personal Information – APPI, 2020), India (Personal Data Protection Bill, 2019) and Canada (Personal Information Protection and Electronic Documents Act – PIPEDA, 2019) have been amending their pre-existing legislation, in order to address data protection holistically, in the manner of the European Legislation. In Africa, South Africa has the Protection of Personal Information Act (POPIA), which was ratified in 2013 and was fully implemented in 2020. Several other countries have adopted the African Union Convention (or Malabo Convention) on Cyber Security and Personal Data Protection (2014, created in cooperation with the Council of Europe), which pursues to provide a basic framework for the protection of personal data and privacy in the continent.

However, it appears that the law has struggled to keep pace with the rapid evolution of technology, and the use and handling of data presents many challenges for legal interpretation and application (Marchant, 2013). Especially the governance and oversight of emerging technologies like social media, surveillance technologies, robotics, and Artificial Intelligence – among others – is characterized by swift development, many applications and actors, and uncertainties about risks and benefits (Marchant, 2013). These demands range from potential health and environmental risks to broader social and ethical concerns. Due to this complexity, it would be very difficult to regulate any of these emerging technologies effectively and comprehensively.

This matter seems especially challenging in liberal and democratic countries, where freedom of speech favors the plurality of opinions and insights, which might hinder swift law-making. On the other hand, authoritarian regimes are de facto governed by states of oligarchs or monarchs and the laws those dictate, usually without having to worry about legitimization from other authorities.

Contemporary European (and Western) Law deems the individual as its cornerstone, rather than the state, and modern legal systems in Europe and the West value plurality and are based on the principle of individual rights and freedoms. This can be traced back to the Enlightenment period, where the idea of natural rights and the value of the individual emerged as important concepts in political and legal philosophy, and

later shaped the regulatory processes in the continent (Curran, 2001). In contrast, the legal system in China – for example – traditionally considers the individual a non-self-sufficient entity, and emphasizes the importance of social order and hierarchy, prioritizing social harmony, and the interests of the community as a single, collective body (Feng, 2010).

The aforementioned cases share a special common element: the Law of each community or society depends on its own common moral requirements and ethical standards. As American philosopher Ronald Dworkin reasoned in his famous work “Law’s Empire” (1967), Laws are not simply a set of rules or commands, but rather an expression of a society’s considered merits and ethical values. These ethical values and principles that underlie a society’s legal system are not fixed or immutable, but rather evolve over time as a result of ongoing debate and discussion. This is beneficial, since by updating and refining its laws, communities can ensure that they remain relevant and effective in addressing the challenges of a rapidly changing world. This, in turn, helps to create a safer, and more *just* and *equitable* society (Rawls, 1999), which people remain a part of not only out of necessity but also out of free will.

Laws play a critical role in ensuring that justice is served and that societies run smoothly. The ultimate goal of these regulations is to create a society that allows people to coexist in a safe and secure environment¹²². The implementation of laws helps create a vital human connection within societies, which is even more imperative in an era when many of our everyday activities take place without necessarily involving physical human interaction (e.g., shopping online). This bond that voluntarily keeps people alongside each other once they have come together, we call *Trust*.

PART THREE – Trust in the Context of Digitalized Social Systems

“On Artificial Intelligence, trust is a must, not a nice to have.”

— Margrethe Vestager,
Executive Vice-President for a Europe fit for the Digital Age

As explained in Part One, Data is the base of the Knowledge pyramid. It is the foundation of information, which is what we use in order to gain the knowledge we use to make decisions and solve problems. Data exists in the real, physical world as the elements we perceive via our senses, but also in the digital networks, as it is the language computers understand. Problems arise when decisions are made based on incomplete, flawed or biased sets of data, as these decisions inevitably end up being erroneous themselves. Data can also easily be manipulated with the help of Artificial Intelligence. Phenomena such as social engineering, viral fake news or deepfake content are ever-present in the world wide web, and are increasingly playing a more prominent role in our everyday social interactions and overall lives.

After having analyzed how modern societies revolve around data, we will examine the digital aspect of contemporary social systems, and how data and Artificial Intelligence can affect relationships within social systems in a digitally developing world.

8. Establishing Human Trust

Social systems are complex networks of interdependent social structures, institutions, organizations, and individuals that function together to create a society (Parsons, 1951). In a social system, individuals interact with each other through established norms, roles, and expectations. These interactions can be influenced by factors such as culture, history, and geographical components, but also financial interests and technology. Social systems are constantly evolving and adapting to changes in society, including technological advancements, demographic change, and economic developments. Understanding social systems is important because it can help us identify and analyze patterns of social behavior, as well as the factors that shape them, and recognize when social change is necessary.

People form social connections for a variety of reasons, but at its core, the purpose of social systems and society altogether is linked to bringing people together to achieve common goals and provide each other with support. One of the primal reasons why humans started to form societies was the survival of our species itself and the fear of external danger, such as wild animals or hostile human groups (Hobbes, ed Shapiro, 2010).

Trust formed the foundation of these early social systems, allowing individuals to cooperate, share resources, and defend themselves against these threats. It was built through repeated interactions among humans, mutual dependence and reliance, and the fulfillment of shared responsibilities. It was this *interpersonal trust* that fostered a sense of unity and enabled early humans to overcome the challenges they faced.

As societies evolved and became more complex, there was a development of social norms, customs, and laws which governed human behavior and which, in return, solidified this feeling of social commitment. Through particular cultural practices, and based on the environments where they lived, communities were able to develop their own theories of knowledge (the first scientific theories) and means of technology (Mormina, 2019). For example, the Sumerians (Mesopotamia, circa 2000 BCE), developed pioneering agricultural techniques to cultivate crops, aiming to take advantage of their land's fertile soil. By building canals and irrigation systems to control water flow, they were able to grow more crops and sustain larger populations, which triggered demographic changes in the region and overall civilizational flourishing (Cowan & Watson, 2006). This practical knowledge of agriculture led to the development of complex mathematical systems (scientific knowledge), which was used for calculations and measurements.

In addition, trust was and still is an indispensable foundation of every economic activity occurring within human society, from the smooth operation of trade and markets, to fostering cooperation and collaboration in the labor market, and the development and maintenance of infrastructure that ensures quality of life. For instance, in ancient China the Confucian concept of *guanxi* (关系), meaning a closed system of "personal connections/relationships", was predominant in Chinese society and commerce (Li, 2020). Trust was built through interpersonal relationships and networks, with individuals relying on credible mediators for business transactions. Trustworthy individuals and family networks (such as clans) played a significant role in establishing these trade relationships and ensuring the fulfillment of obligations by all parties of a transaction.

Lastly, as demonstrated in the previous chapter, *institutional trust* – trust in the institutions of societies – allowed for the establishment of governance, followed by justice and law. Naturally, societal stability is favored when people have faith in the fairness of courts, judges, and legal processes. Thus, trust in governance nurtures social cohesion, and legitimacy of political actions, such as important decision-making.

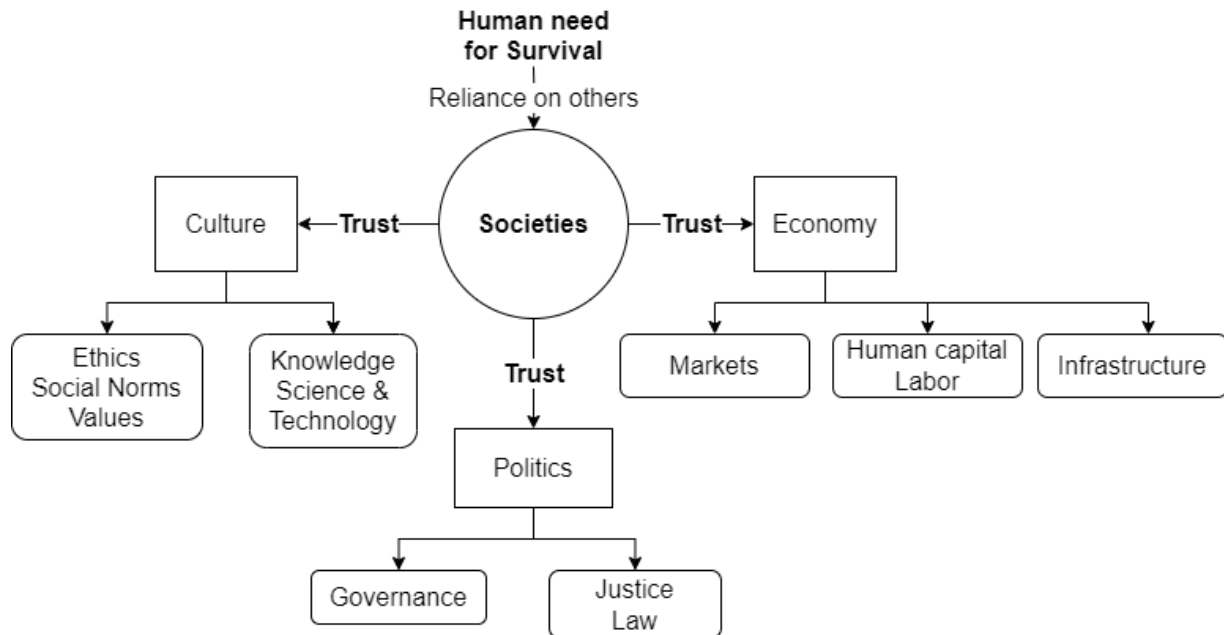


Image 21: A visual representation of Societies and their aspects.

When it comes to institutional trust, it has been found that it also increases feelings of security, and therefore promotes interpersonal trust among members of a society, even when those are strangers (Spadaro et al., 2020). Recent studies by multiple associations, such as the OECD (Organization for Economic Cooperation and Development), indicate that societies with high levels of interpersonal trust show stronger economic growth than communities marked by mistrust among their members. What is more, they are consisted of happier citizens, who also tend to participate more in public affairs (Portela et al., 2013, Algan & Cahuc, 2010).

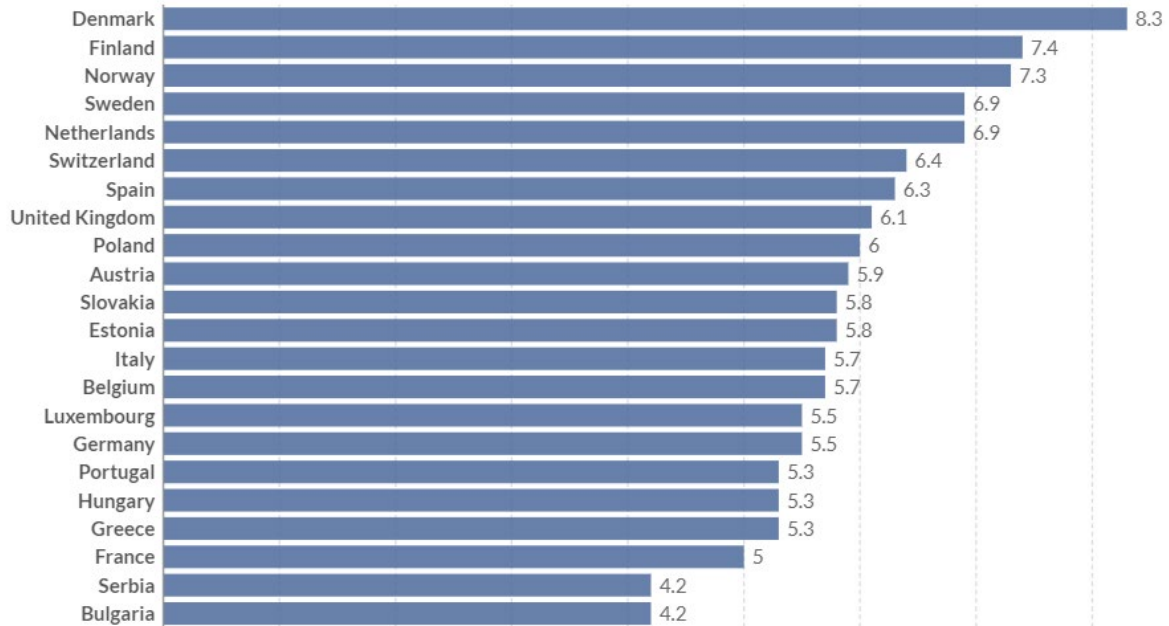
Characteristically, a 2013 survey regarding the estimates of interpersonal trust in Europe showed that countries with good quality of liberal democracy and trustworthy legal systems have a higher rate of interpersonal trust. In the following graph, the tendency of Nordic countries to trust their fellow citizens is in most cases higher than those of Eastern Europeans:

Interpersonal trust in Europe, 2013

Respondents answered the survey question "would you say that most people can be trusted?" on a scale ranging from 0 (low trust) to 10 (high trust). Shown is the average answer for each country.



[+ Add country or region](#)



Source: Trust – Eurostat (2015)

OurWorldInData.org/trust • CC BY

Image 22: Estimates of interpersonal trust in Europe, 2013
 [Source: Trust – Eurostat (2015) <https://ourworldindata.org/trust>]

This is directly linked to how each person evaluates democracy in their respective countries. In countries like Norway (NO), Sweden (SE), Finland (FI), Denmark (DK) and the Netherlands (NL), individuals are heavily inclined to hold liberal democracy, social justice, and direct democracy to high esteem. In the second graph, it is clearly shown that these are citizens from countries that score the highest when it comes to trusting the political system as a whole, police (as a part of the state), and the legal system – the laws and how they are implemented by legal instruments in their country (e.g., courts).

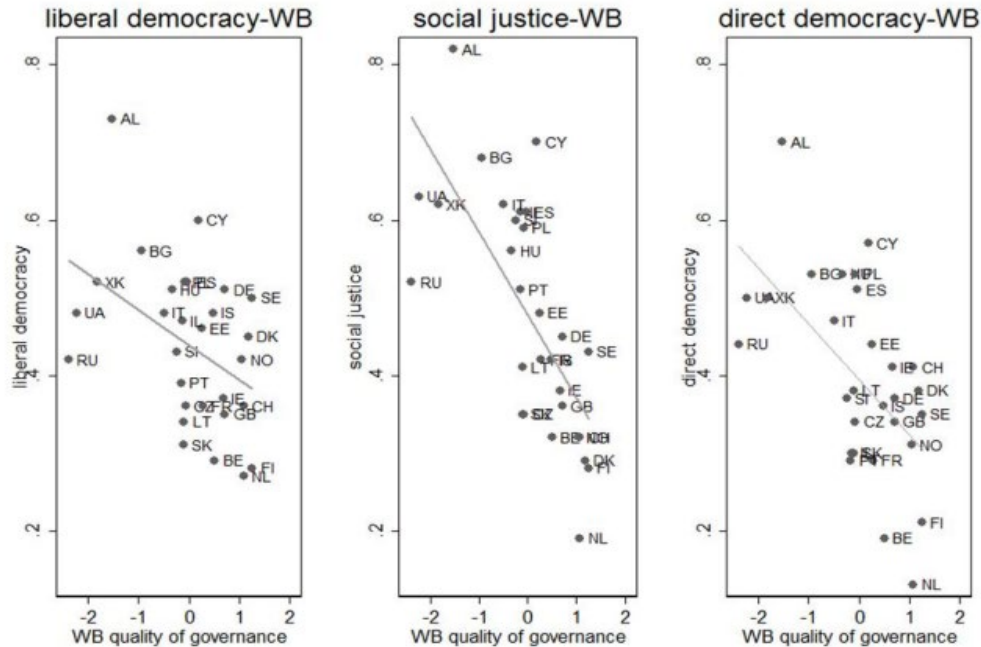
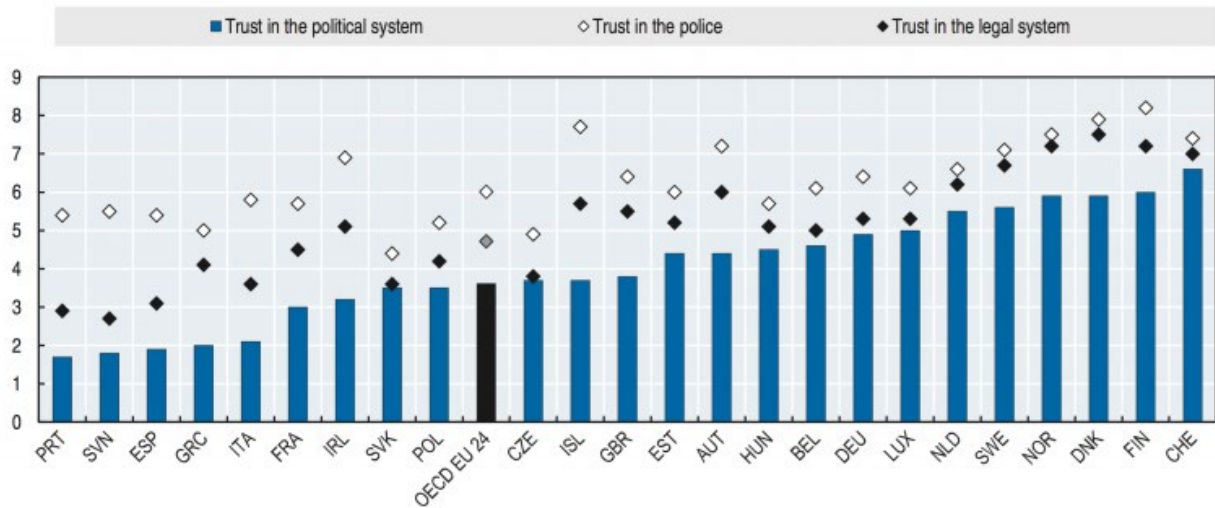


Image 23: Liberal Democracy/Social Justice/Direct Democracy by Quality of democracy (2013) – How Europeans View and Evaluate Democracy, 2014
 [Source: OECD <https://www.oecd.org/statistics/ESS%20seminar%20Hanspeter%20Kriesi.pdf>]



Note: Response options range from 0 ("No trust at all") to 10 ("Complete trust"). The OECD EU average is the population-weighted average of the values included in the chart.
 Source: Eurostat (2015), European Union Statistics on Income and Living Conditions (EU-SILC), http://appsso.eurostat.ec.europa.eu/nui/show.do?dataset=ilc_pw03&lang=en.

Image 24: Estimates of trust in public institutions in European countries in 2013
 [Source: Figure 3.14 in the OECD report How's life? (2015). <https://ourworldindata.org/trust>]

What is particularly interesting is the timing that this survey was conducted: amidst the immediate aftermath of the 2008-2009 financial crisis. Although the global recession lasted for about two years, and was presumed to be over by June 2009, its repercussions on the overall economy lingered for a considerably

Data and AI in Social Systems: Trust Perspectives in a Digitally Developing World

longer period of time. With unemployment rates not returning to pre-recession levels until 2014 and median household incomes severely affected until 2016, systemic trust and trust in institutions also wavered, as did the feeling of security (Roth, 2022). Yet, this decline was milder in countries with a strong Rule of Law tradition, proving that there is an amphidromous relationship between trust in the public and in the private sphere.

Despite the fact that trust in individuals differs from trust in groups, building trust requires nurturing all kinds of relationships within social systems, from personal ones to professional and communal relations on the whole. Elements like consistency and credibility of decisions and actions taken by people and institutions alike, transparent communication regarding said actions, fairness and treating everyone with equity and respect, and privacy and respect of boundaries are crucial in cultivating trust and strengthening interpersonal and institutional connections throughout the social spectrum (Simon, 2020).

9. Trust in a Digitally Developing World

In the modern world, trust remains fundamental to the functioning of social systems, albeit in different contexts. Throughout history, the form of social systems has evolved, societies have grown in numbers and complexity. While threats may have changed from wild animals to global challenges, such as economic instability, climate change, or technological risks, trust is still essential for individuals to come together and work towards creating a more cohesive and harmonious environment that benefits everyone. When certain predicaments appear to (directly or not) affect humanity as a whole, the answers ought to come from an international, holistic point of view.

Regardless, establishing trust in a digital networking environment encompasses a broader range of factors compared to the physical world. This is due to the reliance of computer network communications on humans and their intentions, as well as digital components, such as Big Data and undecipherable Artificial Intelligence algorithms.

One significant challenge is interconnected with the quality of data used by the systems, and how flawed or biased data can affect public perception and reinforce prejudices and stereotypes about certain groups of people. As Artificial Intelligence systems and algorithms are increasingly used to automate processes, they can diminish objectivity of decision-making processes and perpetuate existing inequalities (Image 25). The problem intensifies upon reconsidering that AI algorithms often operate as black boxes, making it challenging to understand how decisions are being made. Thus, the sentiment of injustice deepens, due to a non-human agent making life decisions for humans (e.g., job candidate qualification assessment based on CVs). This hinders social and professional collaboration and intensifies divisions and mistrust among individuals across social systems.

An infamous case in point is the use of certain AI-based computer models by Amazon.com Inc for recruiting purposes. Essentially, based on the data it was given, Amazon's system independently taught itself to favor male candidates (Dastin, 2018). Provided with data from the male-dominated tech-industry, it prejudiced against resumes that mentioned "women's", as in "women's chess club captain", effectively sidelining qualified female candidates. After the initial public outcry, Amazon was forced to adjust the software module in order to eliminate bias towards women. However, this did not guarantee that the algorithm would not develop alternative methods of candidate evaluation that could potentially prove sexist or biased in any way, and Amazon finally decided to dissolve the group of machine-learning specialists who had created the program (Aumüller-Wagner, 2019).

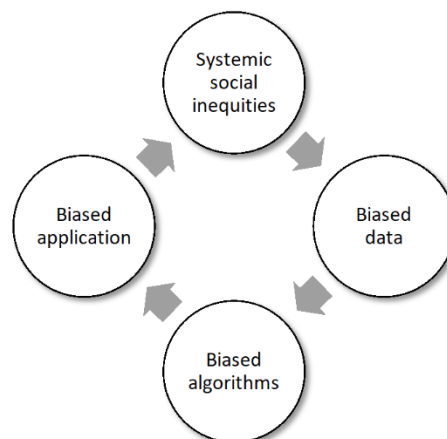


Image 25: Bias and AI Systems [Source: <https://liberalarts.utexas.edu/psychology/news/research-paper-from-adela-timmons-looks-at-ai-bias-in-mental-health-apps>]

Notwithstanding, even if it were possible to find a panacea for “bad data” and system transparency, a key part of the problem lies in the overload of data and information available, as much as the spreading of disinformation (fake news). The ascendance of Big Data seems inevitable when, as of 2023, approximately 65% (5.2 billion people) of the global population uses the internet, and almost 60% (4.2 billion people) use social media (Petrosyan, 2023). Especially in the Western world, where 9 out of 10 adults use the world wide web leaving their online mark, Big Data is being generated at an accelerating pace on a daily basis (Image 26).

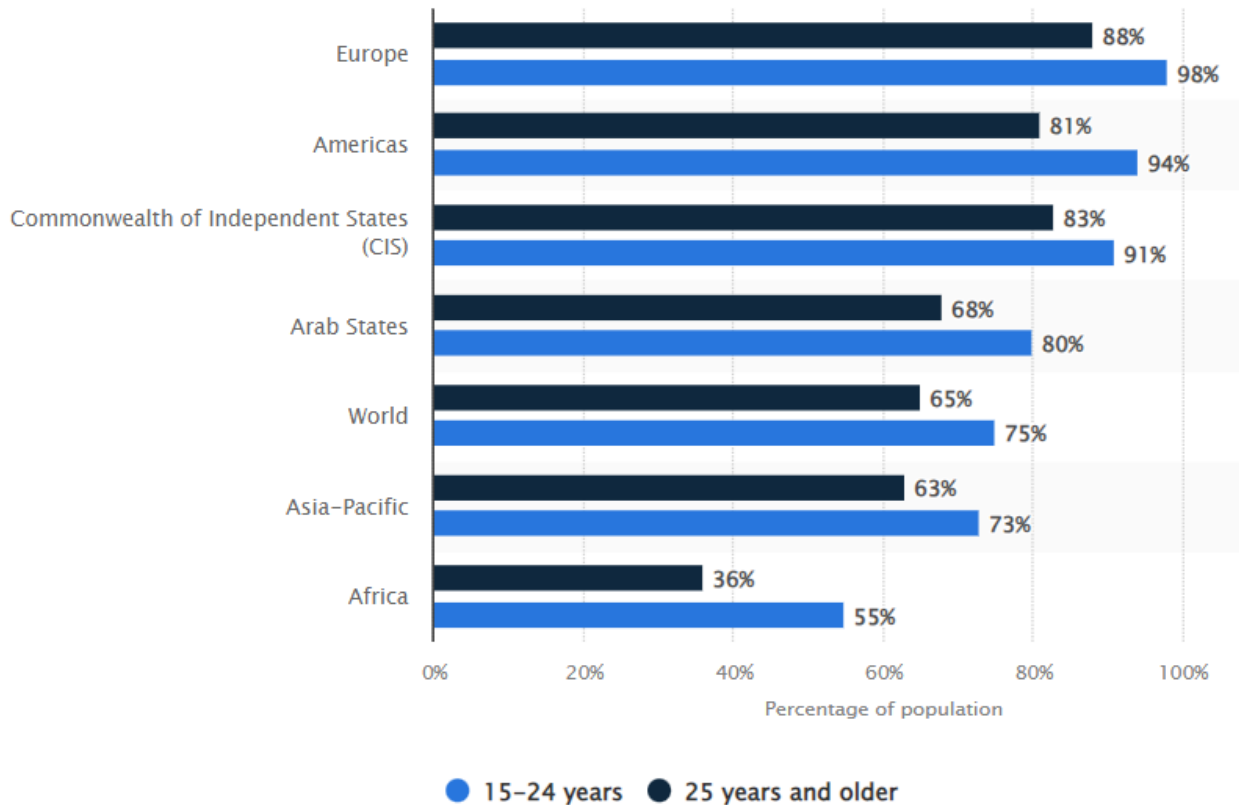


Image 26: Age distribution of internet users worldwide in 2022, by region (Statista)

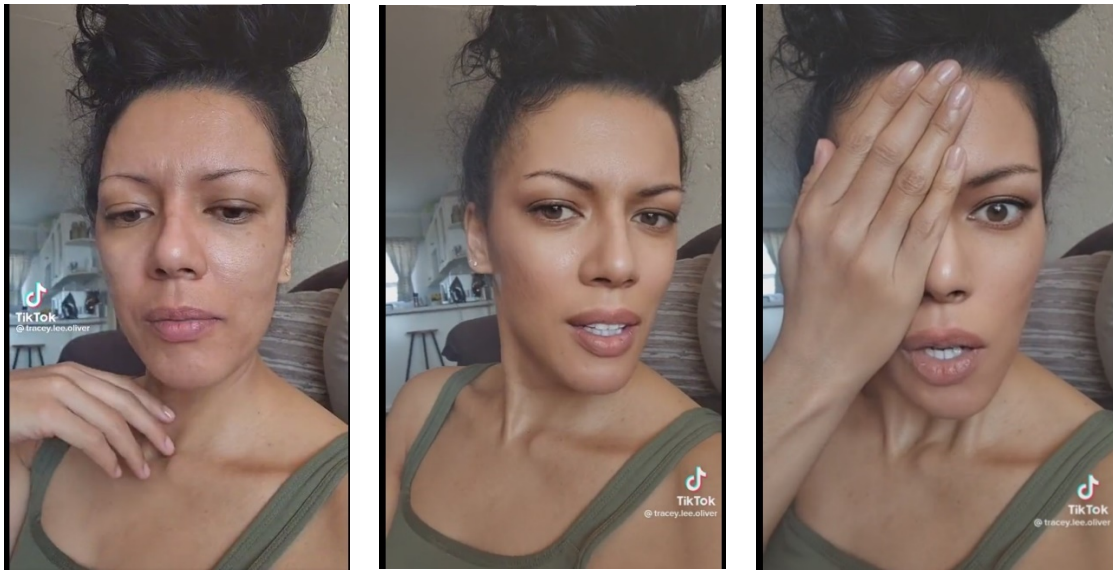
[Source: <https://www.statista.com/statistics/209101/age-distribution-internet-users-worldwide-by-region/>]

While this abundance of data and information everywhere online has its advantages (e.g., being able to instantly be informed about various issues, or being able to track on Google Maps all the places we have visited since the creation of our accounts), the sheer volume of information available can be overwhelming. With the proliferation of social media platforms, news websites, blogs, and online forums, anyone can publish content and share information, regardless of its accuracy or credibility. Thus, it becomes difficult to discern what is true and reliable.

Additionally, the rise of Generative AI and automated bots capable of generating fake news and visual content (e.g., deepfakes) and distributing it to manipulate public sentiment further aggravates the problem (Kreps, 2022). Especially visual signals (such as images, videos, and GIFs) are more rendering compared to written text, thereby inspiring a stronger sense of trust in the content that the human user receives (Sherwin, 2006), which makes such content ideal for propaganda and spreading disinformation. Widely distributed fake news, coupled with the consequences of data and information overload, can erode trust in both traditional institutions and interpersonal relationships, since receiving conflicting or contradictory information often

invokes feelings of skepticism and uncertainty. This immediately results in people questioning the credibility of multiple news sources, government and political institutions, scientific studies, and even the intentions of friends, colleagues, or acquaintances who share information with them (Mangold, 2022).

The concern grows yet more when it becomes easier to alter content with simple actions, like using a filter to take a photo on social media with a simple tap. For example, TikTok offers its users access to live make-up filters able to adapt to virtually any face movement of the user, allowing them to appear flawless to both their own selves and others.



27.A

27.B

27.C

Image 27.A: the user without the filter

Image 27.B: the user upon activating the “Bold Glamour” filter

Image 27.C: the user is surprised to see how the filter does not falter when she moves and touches her face

[Source: TikTok video @tracey.lee.oliver]

It is worth noting that the vast majority of such beauty filters target (young) female users. It has also been proven that such social media filters affect women’s perceptions of self and beauty, leading to lower self-esteem and depression vulnerability (Solomon, 2016). Recent neuroscience research suggests that people with depressive symptoms tend to be more isolated and show less interpersonal trust, and vice versa (Fermin et al., 2022), indicating that actions must be taken to support mental health – especially this of the youth, in order to encourage stronger bonds of trust within society.

Supplementally, with the increasing reliance on technology and the collection of personal data through illegitimate means, individuals have legitimate concerns about their privacy and the security of their data and personal information stored online. High-profile data breaches, unauthorized access to personal data, and the misuse of data by corporations (similar to the Cambridge Analytica scandal) or governments (like the cases in Asia or even Europe regarding citizen surveillance) can lead to a loss of trust in the institutions responsible for safeguarding this data.

Consequently, concerns about the imperfections of AI systems have intensified the sense for urgent action against AI misuse among civilians and authorities alike. In order to address the situation and attempt to make cyberspace a safer and less mistrustful place, there has been a series of measures taken by governments, companies, and individuals alike on an international level.

Besides the official legislation implemented on an international level, as explored in *The Legal Aspect of Data*, there have been more suggestions about introducing regulations that explicitly govern the use of

Artificial Intelligence. The European Union has been a pioneer in this aspect, as it continues to consider new legislative acts, to enhance the level of efficiency of the current framework. Specifically, the Digital Services Act (Regulation (EU) 2022/2065), aiming to tackle illegal content and disinformation online, while promoting transparent advertising, received several amendment proposals, the main of which was obliging big tech companies (e.g., Google, Facebook, TikTok, etc.) to mandatorily label AI-generated content as such (Goujard, 2023). Moreover, the European Centre for Algorithmic Transparency was founded, with the goal of enforcing newer rules on platforms regarding the way their Platform Recommender Systems (PRSs) work: from which type of content the systems promote, to removing material that does not comply with the legal framework (Reviglio & Santoni, 2023).

At the same time, the European Commission updated its technology Ethics Guidelines (AI HLEG, 2022) to shift focus towards *Trustworthy AI*, Artificial Intelligence systems that are reliable, fair, transparent, and accountable, in a manner that respects ethical values, human rights, and societal well-being. The Guidelines set out a set of seven essential criteria that AI systems should meet to be considered trustworthy:

- i. Human agency and oversight: AI should be empowering individuals, allowing them to make informed decisions and safeguarding their fundamental rights.
- ii. Technical Robustness and Safety: AI systems must be resilient and secure. They should also exhibit accuracy, reliability, and reproducibility to minimize unintentional harm.
- iii. Privacy and Data Governance: AI systems must respect privacy and data protection principles.
- iv. Transparency: The data, functioning, and business models of AI systems should be transparent. Traceability mechanisms can contribute to this, ensuring awareness of interaction with an AI system.
- v. Diversity, Non-Discrimination, and Fairness: AI systems should avoid unfair biases that can lead to negative consequences such as marginalization, prejudice, and discrimination.
- vi. Societal and Environmental Well-being: AI systems should benefit all humans, including future generations. They should be designed with sustainability and environmental considerations in mind, taking into account the well-being of other living beings. Social and societal impacts should be carefully assessed.
- vii. Accountability: Mechanisms should be implemented to establish responsibility and accountability for AI systems and their outcomes.

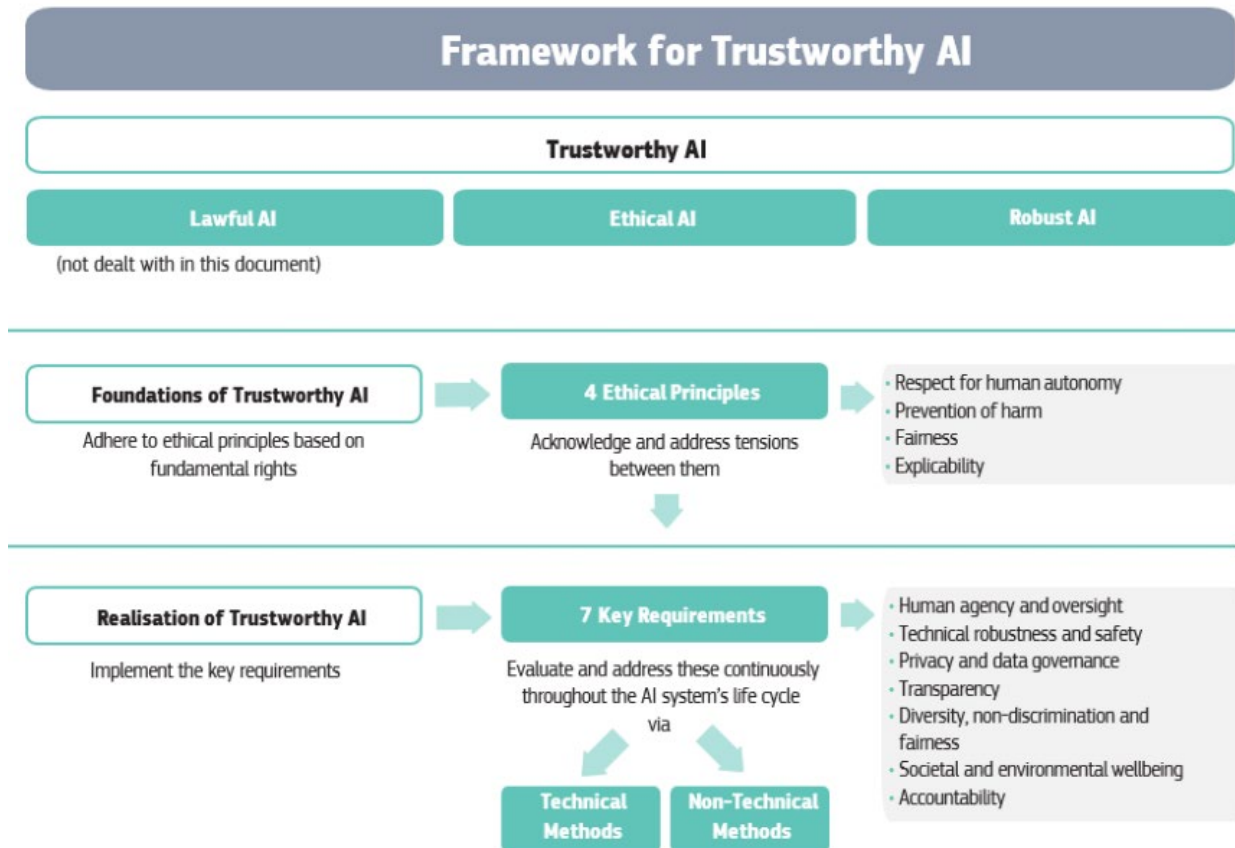


Image 28: Framework for Trustworthy AI (European Commission)

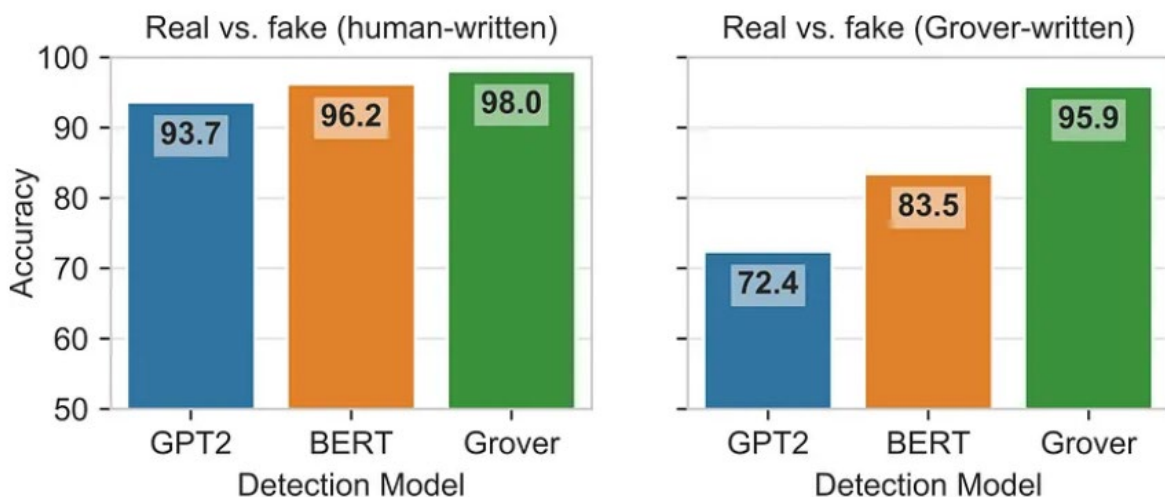
[Source: <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>]

In the meantime, besides government organization actions, some companies such as Meta (parent-company of Facebook, Instagram, and WhatsApp) claim to have launched their own campaign against misinformation and fake news across the internet. Meta has launched Sphere, a new tool that utilizes Artificial Intelligence to identify and tackle "fake news" on Wikipedia. The corporation has claimed that Sphere "*the first model capable of automatically verifying hundreds of thousands of citations at once*" to verify their alignment with the associated claims (Meta/Facebook website, 2022). Undoubtedly, there is a lot yet to be achieved, particularly since Meta (and other popular social media platforms like Twitter) currently does not actively moderate non-violent content on its own websites, in an attempt to keep a balance between user traffic (company revenue), freedom of speech that it claims to support, and putting a barrier on misinformation (Liu et al., 2021, Meta website, 2022). Google has also attempted to combat the spread of fake news through its search engine, by utilizing algorithms and AI technologies to assess and rank search results based on relevance and quality (Gorwa et al., 2020).

Significant efforts have been made by smaller, less centralized entities as well. In 2019, a team of computer scientists from the University of Washington and Allen Institute for AI (AI2) launched a new system named "Grover". Alarmed by the success that fake news seemed to have online, they decided to create an application, which could detect fake news (Zellers et al., 2019). The plan was to first teach Grover how to write fake news upon a prompt given by a human agent, and then allowing it to write a completely fabricated text. After careful data processing from its part, the program was able to recognize texts written by similar AI tools, with a success rate of 92%, as opposed to other generators, which could only identify fake news with approximately 73% accuracy (Image 29). Although this would not stop the spreading of human or bot-written

Data and AI in Social Systems: Trust Perspectives in a Digitally Developing World

false information online, it would at least make spotting it effortless and time-effective, contributing to the fight against fake news.



Grover succeeds at telling apart not just human-written real news from neural fake news, it can also tell apart real news from human-written disinformation. A Grover model that is trained to do both gets over 95% accuracy in both settings.

Image 29: Grover can detect fake news written by both humans and machines (Rowan Zellers, AI2)

[Source: <https://blog.allenai.org/counteracting-neural-disinformation-with-grover-6cf6690d463b>]

As Manohar Paluri, director on Facebook's AI team noted during an interview, when it comes to systems creating disinformation online "*if you have the [false news] generative model, you have the ability to fight it*" (Metz & Blumenthal, 2019). Therefore, the conversation should involve – but not exclusively revolve around – the technical means to hinder fake news and unlock the black box of Artificial Intelligence.

As humans, we are inherently prone to biases, and given that we are the creators of AI systems, these biases can automatically be reflected onto technology. Rectifying human biases is often more difficult than addressing biases in AI systems themselves, especially when only 400 AI safety specialists contribute to this effort (Hilton, 2022). Notwithstanding, AI and data safety is gaining awareness, and the ethical problems of Artificial Intelligence and data use for algorithm development are being brought to the vanguard of technological policy agenda.

As a matter of fact, a proposed solution to amend prejudice and produce unbiased models could come from practicing Open-source Data Science (OSDS), and Open-source Intelligence (OSINT) in general. Open-sourcing technology would mean sharing the programs with the public and allowing more independent data scientists to work with the systems while offering their own perspective and skills. Open-source Intelligence would allow for the collection, analysis, and correlation of publicly available data from various open data sources, such as mass media (e.g., online newspapers), social network platforms, blogs, public government data, webpage metadata and commercial data (Pastor-Galindo et al., 2020).

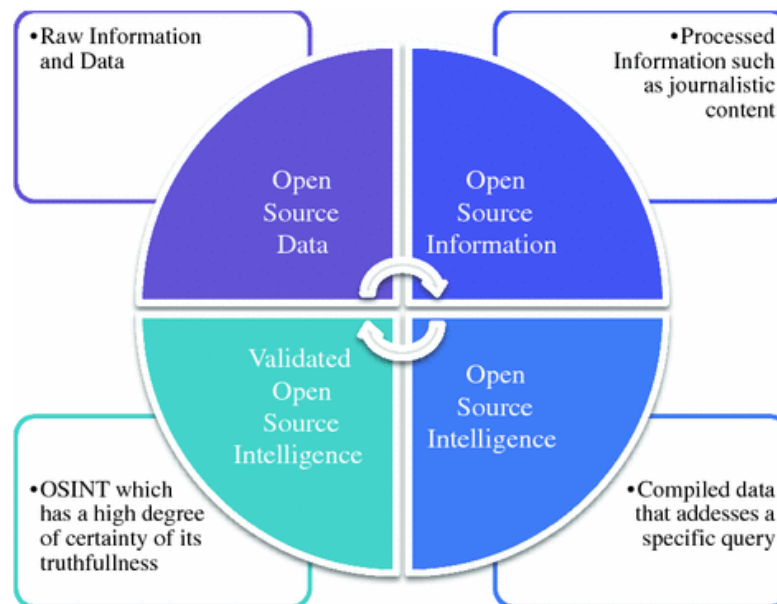


Image 30: From open source data to validated OSINT (Gibson, 2016)
 [Source: https://link.springer.com/chapter/10.1007/978-3-319-47671-1_6/figures/1]

As discussed in *the Business Aspect of Data*, cybercriminals can evoke data from such open sources and exploit it for their own benefit. This has led to a debate on whether using open data for other than personal purposes is ethical to begin with. For example, finding a plumber's phone number on their website (open data) and saving it on our personal device to be used in case of need would generally be considered an "innocent" act. If, however, we were to use that data to teach a location-detection algorithm to decipher which area the phone number corresponds to (based on the country/area code in the beginning), the type of data utilization differs from the first case. In the latter, it stands to reason that the plumber might feel like their privacy is being threatened or even violated, as the system would collect and exploit personal identifiable information for non-personal use.

Despite this argumentation, it is understandable that OSINT's wider implementation with goodwill and benevolence would signal an era of openness and broad collaboration in technology sciences (Böhm & Lolagar, 2021), which would strive for a more humanistic, ethical-centric Artificial Intelligence. The latter would, in turn, contribute to the battle against cybercrime, and aid in creating a safer and more trustworthy cyberspace for all users.

Government entities also generate and acquire large volumes of data and information. As a result, several countries have already adopted OSINT in accordance with a Directive (legislation proposal) of the European Parliament regarding open data and the re-use of public sector information, and in alliance with the approach method of their respective national intelligence agencies (Nouh et al., 2019, EU Directive 2019/1024). By implementing the concept of Open Government Data (OGD), they make these datasets openly available, and enhance transparency and accountability to citizens. This, in return, empowers public trust in governing institutions. Moreover, the encouragement of data utilization, reuse, and unrestricted distribution by governments promotes the creation of new businesses and the development of innovative services centered around citizen needs.

The map below, created by the Open Data Inventory (ODIN), assesses the extent to which a country's statistical data is comprehensive and corresponds to international standards of transparency. Openness scores are determined by factors such as the availability of data in machine-readable and non-proprietary formats, the inclusion of metadata, the presence of download options, and the existence of an open terms of use or data license.

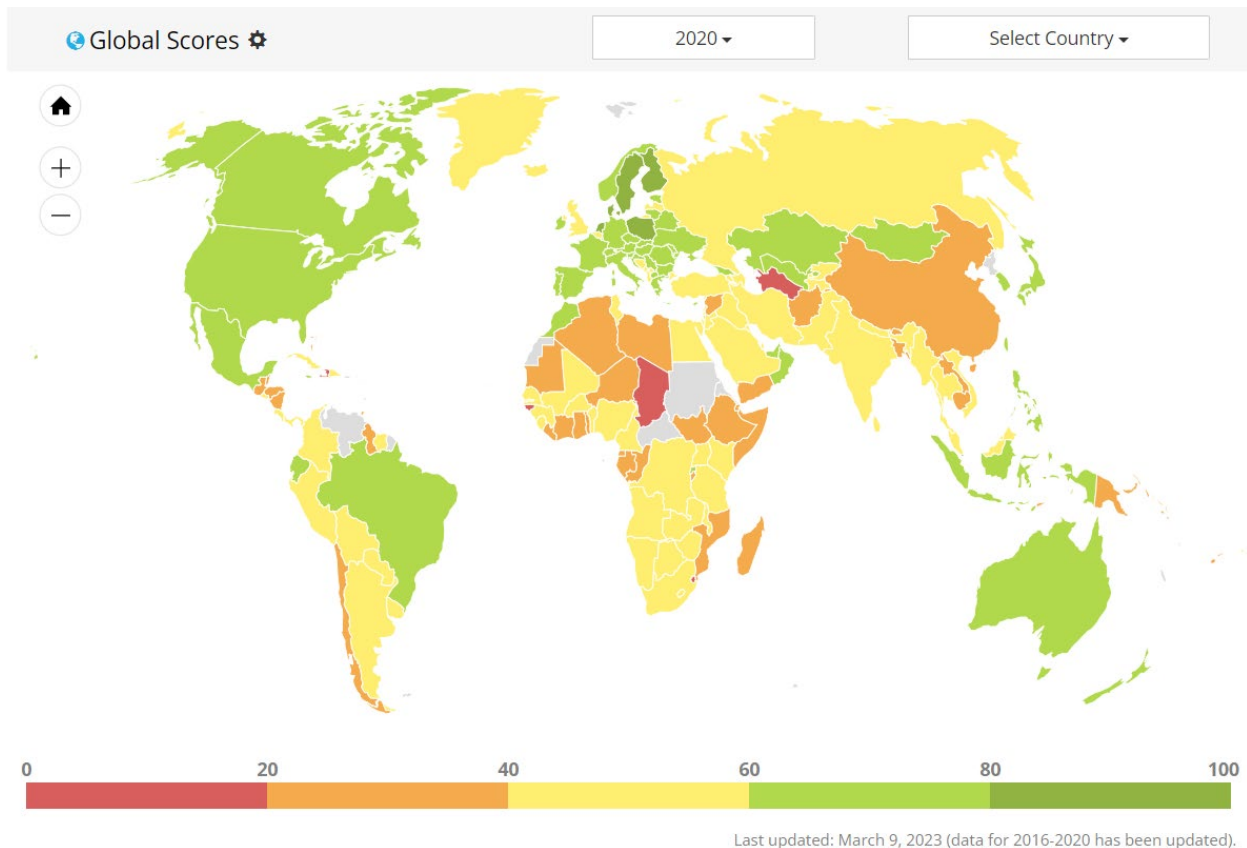


Image 31: Open Data Inventory (ODIN) 2020: Open Data Quality Scores submitted to the United Nations Statistics Division (ODIN, Open Data Watch) [Source: <https://odin.opendatawatch.com/>]

Addressing all these challenges requires a multifaceted approach. It requires actions from those who create and program systems and algorithms, but also a general shift in the way we, as users, deal with data in the Digitally Developing World.

For these reasons, it is also important to provide cybersecurity education to users of all ages, to enhance understanding of potential threats and encourage the implementation of proactive security measures. Promoting digital literacy and critical thinking skills is crucial, so that users are able to assess sources and verify information, without participating in the vicious cycle of reproducing fake news.

Simultaneously, governing authorities and technical experts on their part should demonstrate a sense of responsibility when interacting with personal data, especially when it comes to collecting and processing it. Finally, it would be wise for users not to share unnecessary data online or at least be extremely cautious when they do so. Allowing only people we are familiar with to see our full social media profiles, reducing the sharing of personal photos and personal information, using a strong password, and enabling two-factor authentication (signing in via an extra security step) are only a few of the measures we can actively take in order to feel more secure, and ensure that our digital endeavors remain a pleasant yet responsible experience.

Establishing trust in digital social systems while guarding our privacy is the principal challenge faced by humans in the Digitally Developing World, because it requires both digital literacy and social maturity. Promoting responsible information sharing and creating a culture of trust and integrity online are imperative especially in the case of social media platform data sharing, as nowadays our online profiles complement our “real lives”. This would not suppose that, until the internet becomes a completely safe place, we as users should be mistrustful by default of those whom we come across online. Rather, it would demand that we learn

Data and AI in Social Systems: Trust Perspectives in a Digitally Developing World

to *value* our *privacy* and display caution as if we had to cross a street: you always think that the car will stop before it hits you, but still double-check before you cross.

Conclusion

The erosion of trust can lead to social division, conflict, and the destruction of the social fabric. It is crucial for societies to actively cultivate and maintain trust both in the physical and the digital sphere of the contemporary state of affairs, the Digitally Developing World. This can be achieved through transparent governance, accountability, and open communication. The collaboration between (technology) companies, policymakers, and civil society plays a vital role in developing effective strategies to mitigate social and data bias, tackle disinformation, improve transparency on online platforms, and facilitate the dissemination of trustworthy information.

Furthermore, promoting responsible information sharing and creating a culture of trust and integrity online are vital, and so is encouraging individuals to value and protect their privacy. It requires the active participation and engagement of individuals, institutions, and society as a whole to reduce the negative impact of data and information overload and lessen the dangers that Artificial Intelligence brings with it. While there might always be financial, political and legal puzzles to be solved as emerging technologies continue to be ubiquitous in our lives, it is important to consider that we can use these innovations to work towards a more secure and prosperous future.

In the final chapter we saw how tools provided by technology can either be viewed as a threat or become threat deterrence tools themselves. As such, the final point of this thesis is that it would be of no use to panic over AI technologies or robots and machines taking over our consciences or lives. Therefore, questions for future research could revolve around (a) whether it would it be better to have a state-monitored AI, in the same way that technology legislation is created by the State, (b) if that would not protect citizens' rights to an efficient degree. Moreover, it would be interesting to examine (c) whether such interference would "kill" the innate freedom of technology and its innovation and (d) how would that enhance or hinder the establishment of trust in the Digitally Developing World. Finally, the question at hand remains regarding the growing presence of Generative AI in cyberspace: will we be able to tell with absolute accuracy what is real and what is counterfeit online, when Generative AI tools are becoming exceptional at imitating human ways of expression? How will that change the nature of the internet as a means of democratizing information sharing? The answers to the above questions remain to be explored and, eventually, seen.

Bibliography

- Luppi, A.I., Mediano, P.A.M., Rosas, F.E. et al. (2022). A synergistic core for human brain evolution and cognition. *Nature Neuroscience* 25, 771–782. <https://doi.org/10.1038/s41593-022-01070-0>
- Data. Merriam-Webster.com Dictionary, Merriam-Webster, <https://www.merriam-webster.com/dictionary/data>. Accessed 27 Sept. 2022.
- Aboelimged, M. & Ali, I. & Hashem, G. (2022). Mobile apps use for wellness and fitness and university students' subjective wellbeing. *Information Development*. 38. 672-687. 10.1177/02666669211020498.
- African Union Convention on Cyber Security and Personal Data Protection. (2015, upd. 2022). <https://au.int/en/treaties/african-union-convention-cyber-security-and-personal-data-protection>. Accessed on 28 Mar. 2023.
- Albrecht, B. & Naithani, G. (2022). Digital authoritarianism: A global phenomenon. <https://akademie.dw.com/en/digital-authoritarianism-a-global-phenomenon/a-61136660>. Accessed 25 Oct. 2022
- Algan Y, Cahuc P. (2010). Inherited trust and growth. *Am Econ Rev*. 100: 2060–2092.
- Allen, G. C. (2019). Understanding China's AI Strategy: Clues to Chinese Strategic Thinking on Artificial Intelligence and National Security. Center for a New American Security (CNAS). <https://www.cnas.org/publications/reports/understanding-chinas-ai-strategy>. Accessed 28 Feb. 2023.
- Andersen, A. (2020). The Panopticon is already here. *The Atlantic*. September 2020 Issue. <https://www.theatlantic.com/magazine/archive/2020/09/china-ai-surveillance/614197/>. Accessed on 28 Feb. 2023.
- Angel, W. R. III (2018). The Pulse of a Nation: A Study on the Cybersecurity Issues Plaguing the Healthcare Industry. College of Education, Criminal Justice, and Human Services. University of Cincinnati, Ohio. USA.
- Attaran, M. & Deb, P. (2018). Machine Learning: The New 'Big Thing' for Competitive Advantage. *Int. J. Knowledge Engineering and Data Mining*, Vol. 5, No. 4, 2018. 5. 277-305. Doi: 10.1504/IJKEDM.2018.10015621.
- Aumüller-Wagner, S. (2019). Encoded Bias in Recruitment Algorithms. *Excellent Student Paper Series (ESPS); STS En-counters journal*. Pp 4-6.
- Banks, M. (2018). Cambridge Analytica/Facebook Scandal: A 'Stab in the Heart of Democracy', Warn MEPs, *The Parliament Magazine*, <https://www.theparliamentmagazine.eu/news/article/cambridge-analyticafacebook-scandal-a-stab-in-the-heart-of-democracy-warn-meps>. Accessed on 27 Jan. 2023.
- Banks, M. (2021). New EU regulations on AI seek to ban mass and indiscriminate surveillance. *The Parliament Magazine*. <https://www.theparliamentmagazine.eu/news/article/can-the-eus-new-ai-regulation-become-another-european-success-story>. Accessed on 27 Mar. 2023.
- Bartz, D. (2020). Senators urge U.S. Justice Department to probe TikTok, Zoom. Reuters. <https://www.reuters.com/article/us-usa-china-tiktok-zoom-idUSKCN24V36O>. Accessed 20 Mar. 2023.
- Bendovschi, A. (2015) Cyber-Attacks – Trends, Patterns and Security Countermeasures. *Procedia Economics and Finance*, 28. 24 – 31 [https://doi.org/10.1016/s2212-5671\(15\)01077-1](https://doi.org/10.1016/s2212-5671(15)01077-1)
- Bischoff, P. (2022). Surveillance camera statistics: which cities have the most CCTV cameras? Comparitech. <https://www.comparitech.com/vpn-privacy/the-worlds-most-surveilled-cities/>. Accessed on 28 Feb. 2023.
- Boerboom, C. (2020). Cambridge Analytica: The Scandal on Data Privacy. Augustana Center for the Study of Ethics Essay Contest. <https://digitalcommons.augustana.edu/ethicscontest/18>. Accessed 27 Nov. 2022.
- Böhm, I., Lolagar, S. (2021). Open source intelligence. *Int. Cybersecurity. Law Rev.* 2, 317–337. <https://doi.org/10.1365/s43439-021-00042-7>
- Boucher, P. (2020). STUDY Panel for the Future of Science and Technology EPRS | European Parliamentary Research Service. Brussels, Belgium. © European Union. ISBN: 978-92-846-6770-3. doi: 10.2861/44572

- Brazilian General Data Protection Law (LGPD, English translation). Lei Geral de Proteção de Dados Pessoais, or LGPD; *Lei 13709/2018*. <https://iapp.org/resources/article/brazilian-data-protection-law-lgpd-english-translation/>. Accessed on 28 Mar. 2023.
- Bumblauskas, D. & Nold, H. & Bumblauskas, P. (2015). Data Collection, Analysis and Tracking in Industry. *Journal of Applied Business and Economics*. 17. 92-100.
- California Consumer Privacy Act (CCPA) Fact Sheet (2018). State of California, Department of Justice, Office of the Attorney General. https://www.oag.ca.gov/system/files/attachments/press_releases/CCPA%20Fact%20Sheet%20%28000000%29.pdf. Accessed 25 Mar. 2023.
- Caprotti, F., & Liu, D. (2022). Platform urbanism and the Chinese smart city: the co-production and territorialisation of Hangzhou City Brain. *GeoJournal*, 87(3), 1559-1573.
- CareerBuilder. (2018). More Than Half of Employers Have Found Content on Social Media That Caused Them NOT to Hire a Candidate. CareerBuilder Press Releases. <https://press.careerbuilder.com/2018-08-09-More-Than-Half-of-Employers-Have-Found-Content-on-Social-Media-That-Caused-Them-NOT-to-Hire-a-Candidate-According-to-Recent-CareerBuilder-Survey>. Accessed 17 Feb. 2023.
- Chan, R. (2020). The Cambridge Analytica whistleblower explains how the firm used Facebook data to sway elections. Business Insider. <https://www.businessinsider.com/cambridge-analytica-whistleblower-christopher-wylie-facebook-data-2019-10>. Accessed 27 Nov. 2022.
- Chavez, T. & O'Hara, C. & Vaidya, V. (2018). Data Driven: Harnessing Data and AI to Reinvent Customer Engagement. *McGraw Hill*. ISBN-13978-1260441536.
- China 2022. (2022). Amnesty International Report 2022/23. <https://www.amnesty.org/en/location/asia-and-the-pacific/east-asia/china/report-china/>. Accessed on 25 Mar. 2023.
- Chrisafis, A. (2023). France under fire over fast-track plan for AI video surveillance at Paris Olympics. The Guardian. <https://www.theguardian.com/world/2023/jan/31/france-paris-olympics-ai-video-surveillance-law>. Accessed 19 Feb. 2023.
- Ciancaglini, V., Gibson, C., Sancho, D., McCarthy, O., Eira, M., Amann, P. & Klayn, A. (2020) Malicious Uses and Abuses of Artificial Intelligence. Trend Micro Research. https://www.europol.europa.eu/cms/sites/default/files/documents/malicious_uses_and_abuses_of_artificial_intelligence_europol.pdf. Accessed 12 Nov. 2022
- Cisco (2018). What are cookies? What are the differences between them (session vs. persistent)?. Cisco website. Accessed on 6 Feb. 2023.
- Colizza, V., E. Grill, and R. Mikolajczyk, et al. (2021). Time to evaluate COVID-19 contact-tracing apps. *Nature Medicine* 27(3): 361–362.
- Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe, Brussels, 25.4.2018 COM (2018) 237 final.
- Cowan, C. W., & Watson, P. J. (Eds.). (2006). The Origins of Agriculture: An International Perspective. *The University of Alabama Press*. ISBN 9780817353490.
- Curran, V. G. (2001). Romantic Common Law, Enlightened Civil Law: Legal Uniformity and the Homogenization of the European Union. *Colum. J. Eur. L.*, 7, 63.
- Dastin, J. (2018). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G>. Accessed on 5 Jun. 2023.
- Delgado, J., de Manuel, A., Parra, I. et al. (2022). Bias in algorithms of AI systems developed for COVID-19: A scoping review. *Bioethical Inquiry* 19, 407–419. <https://doi.org/10.1007/s11673-022-10200-z>
- Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast). PE/28/2019/REV/1. OJ L 172, 26.6.2019, p. 56–83.

- Donahue, P. (2010). The Right to Privacy and Tibetans in China: China's Use of Cyber-Surveillance. *Harvard Human Rights Journal*, 23(1), 161–210.
- Dunn, S. (2021). Women, Not Politicians, Are Targeted Most Often by Deepfake Videos. <https://www.cigionline.org/articles/women-not-politicians-are-targeted-most-often-deepfake-videos/>. Accessed 21 Nov. 2022.
- Dworkin, R. M. (1967). Law's Empire. Belknap Press. ISBN: 9780674518353.
- EC High-Level Expert Group on Artificial Intelligence. (2018). A definition of AI: Main capabilities and scientific disciplines. https://ec.europa.eu/futurium/en/system/files/ged/ai_hleg_definition_of_ai_18_december_1.pdf. Brussels, Belgium. © European Union Accessed 10 Nov. 2022.
- EU AI Convention: stronger protection of fundamental rights is necessary. (2022). https://edps.europa.eu/system/files/2022-10/EDPS-2022-24-AI-CONVENTION_EN.pdf. Accessed 25 Oct. 2022.
- European Commission, High-Level Expert Group on AI (AI HLEG). (2022). Ethics Guidelines for Trustworthy AI. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>. Accessed on 11 Jun. 2023.
- European Commission. (upd. 2023). A European approach to artificial intelligence. <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>. Accessed on 28 Mar. 2023.
- European Data Protection Board - Press Release Statement. (2021). EDPB & EDPS call for ban on use of AI for automated recognition of human features in publicly accessible spaces, and some other uses of AI that can lead to unfair discrimination. https://edpb.europa.eu/news/news/2021/edpb-edps-call-ban-use-ai-automated-recognition-human-features-publicly-accessible_en. Accessed on 27 Mar. 2023.
- European Parliament and Council of the European Union. (2016). Article 17. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, L 119/1. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679&from=EN>. Accessed 27 Mar. 2023.
- European Union. (2012). Consolidated version of the Treaty on the Functioning of the European Union Reference: 2012/C 326/01. *Official Journal of the European Union*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:12012P/TXT&from=EN>. Accessed on 27 Mar. 2023.
- European Union. (2016). REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>. Accessed on 25 Mar. 2023.
- Fact Sheet: Health Insurance Portability and Accountability Act (HIPAA). U.S. Department of Labor, Employee Benefits Security Administration, Washington, DC 20210. <https://www.dol.gov/sites/dolgov/files/EBSA/about-ebsa/our-activities/resource-center/fact-sheets/hipaa.pdf>. Accessed 25 Mar. 2023.
- Feng, Y. (2010). Legal Culture in China: A Comparison to Western Law. *16 Revue Juridique Polynésienne*. pp 115-121.
- Fermin, A.S.R., Kiyonari, T., Matsumoto, Y. et al. (2022). The neuroanatomy of social trust predicts depression vulnerability. *Sci Rep* 12, 16724. <https://doi.org/10.1038/s41598-022-20443-w>
- Gerodimos, R. et al. (2023). Φάκελος "Κίνα": Ψηφιακοί Δούρειοι Ίπποι - Μια Παγκόσμια Υποδομή Παρακολούθησης, Παραπληροφόρησης και Εθισμού | (Greek: Fakelos "Kina": Psifiakoi Doureioi Ippoi - Mia Pangosmia Ypodomi Parakolouthisis, Parapliroforisis kai Ethismou) | "China" Files: Digital Trojan Horses; A Global Infrastructure of Surveillance, Disinformation and Addiction. *Athens Voice Magazine*. <https://www.athensvoice.gr/epikairota/diethni/796053/fakelos-kina-psifiakoi-doureioi-ippioi/>. Accessed on 4 Apr. 2023.

- Getz, A. (2022). Number of jailed journalists spikes to new global record. Committee to Protect Journalists. <https://cpj.org/reports/2022/12/number-of-jailed-journalists-spikes-to-new-global-record/>. Accessed on 21 Jan. 2023.
- Gibson, H. (2016). Acquisition and Preparation of Data for OSINT Investigations. In: Akhgar, B., Bayerl, P., Sampson, F. (eds) Open Source Intelligence Investigation. Advanced Sciences and Technologies for Security Applications. Springer, Cham. https://doi.org/10.1007/978-3-319-47671-1_6
- Gorwa R., Binns R., Katzenbach C. (2020) Algorithmic content moderation: Technical and political challenges in the automation of platform governance. *Big Data & Society* 7(1): 2053951719897945.
- Goujard, C. (2023). EU wants Google, Facebook to start labeling AI-generated content. Político. <https://www.politico.eu/article/chatgpt-dalle-google-facebook-microsoft-eu-wants-to-start-labeling-ai-generated-content>. Accessed on 10 Jun. 2023.
- Government of the Netherlands. Counterterrorism and national security. <https://www.government.nl/topics/counterterrorism-and-national-security/counterterrorism>. Accessed on 17 Feb. 2023.
- Gruessner, V. (2015). The History of Mobile Health: From Cell Phones to Wearables. <https://mhealthintelligence.com/news/the-history-of-mobile-health-from-cell-phones-to-wearables>. Accessed 8 Nov. 2022.
- Harvard Business Review. (2022). Empowering Decision Makers with Self-Service Analytics Webinar. <https://hbr.org/sponsored/2022/03/empowering-decision-makers-with-self-service-analytics-webinar>. Accessed 22 Jan. 2023.
- Henley, J. & Booth, R. (2020). Welfare surveillance system violates human rights, Dutch court rules. The Guardian. <https://www.theguardian.com/technology/2020/feb/05/welfare-surveillance-system-violates-human-rights-dutch-court-rules>. Accessed on 15 Feb. 2023.
- Hernandez, M. & Pinzón, C. & Díaz- López, D. & Garcia, J. & Pinto, R. (2018). Open source intelligence (OSINT) in a Colombian context and sentiment analysis. *Revista vínculos*, vol. 15, no. 2, pp. 195–214, 2018.
- Hilton, B. (2022). Preventing an AI-related catastrophe. <https://80000hours.org/problem-profiles/artificial-intelligence/#neglectedness>. Accessed on 15 May 2023.
- Hobbes, T. (2010). Leviathan: Or the Matter, Form, and Power of a Common-Wealth Ecclesiastical and Civil, ed. by Ian Shapiro. New Haven & London: Yale University Press.
- Holyoake, M. (2021). How LinkedIn Has Changed the Recruiting Industry - and How it Hasn't. LinkedIn. <https://www.linkedin.com/pulse/how-linkedin-has-changed-recruiting-industryand-hasnt-mark-holyoake/>. Accessed on 23 Feb. 2023.
- Hurst, L. (2022). Chatgpt: Why the Human-Like AI Chatbot Suddenly Has Everyone Talking. EuroNews. <https://www.euronews.com/next/2022/12/14/chatgpt-why-the-human-like-ai-chatbot-suddenly-got-everyone-talking>. Accessed 21 Nov. 2022.
- Indian Government Gazette. (2021). Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021. <http://egazette.nic.in/WriteReadData/2021/224361.pdf>. Accessed on 25 Jan. 2023.
- Indian Joint Parliamentary Committee (JPC) on Personal Data Protection: The Personal Data Protection Bill, 2019, As Introduced in Lok Sabha. Bill No. 373 of 2019.
- Japanese Personal Information Protection Commission. Act on the Protection of Personal Information (Act No. 57 of 2003 as amended in 2023). <https://www.ppc.go.jp/en/legal/>. Accessed on 2 Apr. 2023.
- Jeffery, M. (2010). Data-driven marketing: the 15 metrics everyone in marketing should know. John Wiley & Sons, Inc., Hoboken, New Jersey, USA. ISBN 978-0-470-50454-3.
- Jiang, E. (2020). Chinese public toilet forces people to scan their faces before being allowed loo paper 'to reduce waste'. The Daily Mail UK. <https://www.dailymail.co.uk/news/article-9017797/Chinese-public-toilet-facial-recognition-paper-dispenser-sparks-privacy-concerns.html>. Accessed 20 Mar. 2023.

- Kaffenberger, L., & Kopp, E. (2019). Cyber Risk Scenarios, the Financial System, and Systemic Risk Assessment Cyber Policy Initiative Working Paper Series | "Cybersecurity and the Financial System" #4.
- Kayali, L. (2023). France plots surveillance power grab for Paris 2024 Olympics. Politico. <https://www.politico.eu/article/paris-2024-olympic-games-raise-surveillance-fears-camera-facial-recognition>. Accessed on 19 Feb. 2023.
- KBV Research via Reportlinker.com. (2022). Global Machine learning as a Service Market Size, Share & Industry Trends Analysis Report By End User, By Offering, By Organization Size, By Application, By Regional Outlook and Forecast, 2022 – 2028. Report ID: 6289268. https://www.reportlinker.com/p06289268/Global-Machine-learning-as-a-Service-Market-Size-Share-Industry-Trends-Analysis-Report-By-End-User-By-Offering-By-Organization-Size-By-Application-By-Regional-Outlook-and-Forecast-.html?utm_source=GNW. Accessed 3 Nov. 2022.
- Khatsenkova, S. (2023). Smile, AI is watching you: Paris slammed for new video surveillance ahead of 2024 Olympics. Euronews. <https://www.euronews.com/next/2023/02/06/smile-ai-is-watching-you-paris-slammed-for-new-video-surveillance-ahead-of-2024-olympics>. Accessed 19 Feb. 2023.
- Kitchin, R. (2014). The real-time city? Big data and smart urbanism. *GeoJournal* 79: 1–14. Doi: 10.1007/s10708-013-9516-8
- Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences. *The Data Revolution*, 1-240.
- Kopp, M., Nikl, M., & Holeňa, M. (2017). Breaking CAPTCHAs with Convolutional Neural Networks. J. Hlaváčová (Ed.): ITAT 2017 Proceedings, pp. 93–99. *CEUR Workshop Proceedings Vol. 1885*. ISSN 1613-0073.
- Kreps, S., McCain, R., & Brundage, M. (2022). All the News That's Fit to Fabricate: AI-Generated Text as a Tool of Media Misinformation. *Journal of Experimental Political Science*, 9(1), 104-117. doi:10.1017/XPS.2020.37
- Kuner, C. & Svantesson, D. & Cate, F. & Lynskey, O. & Millard, C. (2017). The rise of cybersecurity and its impact on data protection. *International Data Privacy Law*. 7. 73-75.
- Kurichenko., V. (2020) The Surprising Truth Behind Massive Spam on Celebrity Instagram. <https://bettermarketing.pub/the-surprising-truth-behind-massive-spam-on-celebrity-instagram-accounts-5c03913d36ea>. Accessed 21 Nov. 2022.
- Lambert, L. & Woodford, C. & Poole, H. & Moschovitis, C. J. P. (2005). The Internet: A Historical Encyclopedia. *ABC-CLIO*. ISBN-13978-1851096596.
- LexisNexis® Risk Solutions. (2014). Survey of Law Enforcement Personnel and Their Use of Social Media. www.lexisnexis.com/investigations; <https://centerforimprovinginvestigations.org/wp-content/uploads/2018/11/2014-social-media-use-in-law-enforcement-pdf.pdf>. Accessed on 15 Feb. 2023.
- Li, Z. (2020). In: Poff, D., Michalos, A. (Eds.). Guanxi. Encyclopedia of Business and Professional Ethics. Springer, Cham. https://doi.org/10.1007/978-3-319-23514-1_187-1
- Liu, Y., Yildirim, P., Zhang, Z. J., (2021) Implications of Revenue Models and Technology for Content Moderation Strategies Available at SSRN: <http://dx.doi.org/10.2139/ssrn.3969938>
- Lyon, B., Tora, M. (2023). Exploring Deepfakes: Deploy powerful AI techniques. *Packt Publishing*. ISBN 978-1801810692.
- Mangold, F., Bachl, M., & Prochazka, F. (2022). How News Audiences Allocate Trust in the Digital Age: A Figuration Perspective. *Journalism & Mass Communication Quarterly*, 0(0). <https://doi.org/10.1177/10776990221100515>
- Manyika, J., Lund, S., Bughin, J. et al. (2016). Digital Globalization: The New Era of Global Flows. McKinsey Global Institute.
- Maranto, L. (2020). Who Benefits from China's Cybersecurity Laws? CSIS - Center for Strategic & International Studies. <https://www.csis.org/blogs/new-perspectives-asia/who-benefits-chinas-cybersecurity-laws>. Accessed on 25 Mar. 2023.

- Maras, M. H., & Alexandrou, A. (2019). Determining authenticity of video evidence in the age of artificial intelligence and in the wake of Deepfake videos. *International Journal of Evidence & Proof*, 23(3):255–262. <https://doi.org/10.1177/1365712718807226>
- Marchant, G. & Abbott, K. & Allenby, B. (2013). *Innovative governance models for emerging technologies*. Edward Elgar Publishing USA. ISBN: 9781782545637. pp.130-136. <https://doi.org/10.4337/9781782545644.00020>.
- Marr, B. (2015). How Big Data Is Changing Insurance Forever. *Forbes Magazine*. <https://www.forbes.com/sites/bernardmarr/2015/12/16/how-big-data-is-changing-the-insurance-industry-forever/?sh=15fa0907289b>. Accessed on 17 Feb. 2023.
- Marr, B. (2022). The Problem With Biased AIs (and How To Make AI Better). <https://www.forbes.com/sites/bernardmarr/2022/09/30/the-problem-with-biased-ais-and-how-to-make-ai-better/>. Accessed 4 Nov. 2022.
- Marvin, S., While, A., Chen, B. (2022). Urban AI in China: social control or hyper-capitalist development in the post-smart city? *Frontiers in Sustainable Cities*, 4. ISSN 2624-9634 <https://doi.org/10.3389/frsc.2022.1030318>.
- Matthews, E. (2022). China suspected in hack of journalists at News Corp. Reuters. <https://www.reuters.com/business/media-telecom/news-corp-says-one-its-network-systems-targeted-by-cyberattack-2022-02-04/>. Accessed on 21 Jan. 2023.
- Mayer-Schönberger, V. & Cukier, K. (2013). Big Data: A Revolution That Will Transform How We Live, Work, and Think. *American Journal of Epidemiology*, Volume 179, Issue 9, Pages 1143–1144, <https://doi.org/10.1093/aje/kwu085>
- Meijer, A. J. & Torenvlied, R. (2016). Social media and the new organization of government communications: An empirical analysis of Twitter usage by the Dutch police. *American Review of Public Administration*, vol. 46, n2, pp. 143–161. doi: <https://doi.org/10.1177/0275074014551381>. Accessed 15 Feb 2023.
- Meta (2022). How AI could help make Wikipedia entries more accurate. <https://tech.facebook.com/artificial-intelligence/2022/07/how-ai-could-help-make-wikipedia-entries-more-accurate/>. Accessed on 11 Jun. 2023.
- Metz, C., Blumenthal, S. (2019). How A.I. Could Be Weaponized to Spread Disinformation. *The New York Times*. <https://www.nytimes.com/interactive/2019/06/07/technology/ai-text-disinformation.html>. Accessed on 16 Jun. 2023.
- Metz, R. (2022). Facebook and YouTube say they removed Zelensky deepfake. *CNN Business*. <https://edition.cnn.com/2022/03/16/tech/deepfake-zelensky-facebook-meta/index.html>. Accessed 21 Nov. 2022.
- Mitchell, T. (1997). *Machine Learning*. New York, McGraw Hill. USA. pp 5-12 ISBN 0-07-042807-7.
- Mitrou, L. & Kandias, M. & Stavrou, V. & Gritzalis, D. (2014). Social Media Profiling: A Panopticon Or Omnipticon Tool? *Proc. of the 6th Conference of the Surveillance Studies Network*. Athens University of Economics and Business. Pp 1-15.
- Mormina, M. (2019). Science, Technology, and Innovation as Social Goods for Development: Rethinking Research Capacity Building from Sen's Capabilities Approach. *Science and Engineering Ethics*, 25(2), 671-692. <https://doi.org/10.1007/s11948-018-0037-1>
- Mourtzis, D. & Vlachou, K. & Siganakis, E. & Zogopoulos, V. & Kaya, M. & Tekin Bayrak, I. (2017). Mobile Feedback Gathering App for Frugal Product Design. *Procedia CIRP*. 60. 151-156. 10.1016/j.procir.2017.01.042.
- Mousadakos, D. for Big Blue Data Academy. (2022). What is Machine Learning: Sentiment Analysis <https://bigblue.academy/en/what-is-machine-learning>. Accessed 3 Nov. 2022.
- Müller, B., Reinhardt, J., & Strickland, M. T. (1995). Neural networks: an introduction. *Springer Science & Business Media*.
- National Institute of Standards and Technology. (1993). Computer Security Basics. NIST Special Publication 800-12. <https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800-12.pdf>.

- National People's Congress of the People's Republic of China. (2016). Cybersecurity Law of the People's Republic of China. <http://www.npc.gov.cn/englishnpc/c23934/202112/1abd8829788946ecab270e469b13c39c.shtml>. Accessed on 25 Mar. 2023.
- Nouh, M., Nurse, JR., Webb, H., Goldsmith, M. (2019). Cybercrime investigators are users too! Understanding the socio-technical challenges faced by law enforcement. *Proc. 2019 Workshop Usable Security*.
- O'Carroll, L. (2023) French Court's Approval of Olympics AI Surveillance Plan Fuels Privacy Concerns. The Guardian. <https://www.theguardian.com/world/2023/may/18/french-courts-approval-of-olympics-ai-surveillance-plan-fuels-privacy-concerns>. Accessed on 19 Jun. 2023.
- OECD (2015). Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development, The Measurement of Scientific, Technological and Innovation Activities. *OECD Publishing*. Paris, <http://oe.cd/frascati>.
- Palka, S., McCoy, D. (2015). Usenix. Fuzzin E-mail Filters with Generative Grammars and N-Gram Analysis. <https://www.usenix.org/system/files/conference/woot15/woot15-paper-palka.pdf>. Accessed on 12 Nov. 2022
- Parsons, T. (1951). The Social System. England: *Routledge*. ISBN 978-0-203-99295-1.
- Pastor-Galindo, J. & Zago, M. & Nespoli, P. et al. (2020). Spotting political social bots in twitter: a use case of the 2019 Spanish general election. *IEEE Transactions on Network and Service Management*, vol. 17, no. 4, pp. 2156–2170.
- Pastor-Galindo, J., Mármol, G., & Martínez Pérez, G. (2021). Nothing to hide? On the Security and Privacy Threats beyond Open Data. *IEEE Internet Computing*, vol. 25, no. 4, pp. 58-66, doi: 10.1109/MIC.2021.3088335.
- Pastor-Galindo, J., Nespoli, P., Gómez Mármol, F. & Martínez Pérez, G. (2020). The Not Yet Exploited Goldmine of OSINT: Opportunities, Open Challenges and Future Trends. *IEEE Access*, vol. 8, pp. 10282-10304. doi: 10.1109/ACCESS.2020.2965257.
- Peerenboom, R. (2010). Judicial Independence in China: Lessons for Global Rule of Law Promotion, *Cambridge University Press*. SSRN: <https://ssrn.com/abstract=1543003>.
- Pendergrass, W. S. by Skovira, R. J., Rota, D. R., & Draus, P. J. (2013). What is Anonymous: A Case Study Analysis of an Information Systems Hacker Activist Collective Movement. Doctoral Dissertation. ISBN: 978-1-3031-4741-8
- Petrosyan, A. for Statista. (2023). Number of internet and social media users worldwide as of April 2023. <https://www.statista.com/statistics/617136/digital-population-worldwide/>. Accessed on 10 Jun. 2023.
- Pfeifer, Rolf & Iida, Fumiya. (2003). Embodied Artificial Intelligence: Trends and Challenges. *Embodied Artif. Intell.* 3139. 1-26. 10.1007/978-3-540-27833-7_1.
- Portela M, Neira I, Salinas-Jiménez M del M. (2013). Social capital and subjective wellbeing in Europe: A new approach on social capital. *Soc Indic Res.* 114: 493–511.
- Protection of Personal Information Act (POPI Act) – Official Website. <https://popia.co.za/>. Accessed on 28 Mar. 2023.
- Qiang, X. (2019). The Road to Digital Unfreedom: President Xi's Surveillance State. *Journal of Democracy* 30(1), 53-67. doi:10.1353/jod.2019.0004.
- Ramírez Sánchez, J. & Campo-Archbold, A. & Zapata Roza, A. & Díaz-López, D. & Pastor-Galindo, J. & Gómez Mármol, F. & Aponte Díaz, J. (2021). Uncovering Cybercrimes in Social Media through Natural Language Processing. *Complexity*, vol. 2021, Article ID 7955637. <https://doi.org/10.1155/2021/7955637>
- Ramzan, S. (2020). Leveraging Big Data and IoT technology into Smart Homes. *International Journal of Scientific and Research Publications (IJSRP)*. 10. 372. 10.29322/IJSRP.10.09.2020.p10545.
- Rawls, J. ([1971] 1999). A Theory of Justice. *Oxford University Press*. ISBN: 9780198250548.

- Republic of Canada: Personal Information Protection and Electronic Documents Act (S.C. 2000, c. 5). <https://laws-lois.justice.gc.ca/eng/acts/p-8.6/>. Accessed on 28 Mar. 2023.
- Reuters. (2023). Meta rolls back measures to tackle COVID misinformation. <https://www.reuters.com/technology/meta-rolls-back-measures-tackle-covid-misinformation-2023-06-16/>. Accessed on 16 Jun. 2023.
- Reviglio, Urbano and Santoni, Giulio. Governing Platform Recommender Systems in Europe: Insights from China. *Global Jurist*, 2023. <https://doi.org/10.1515/gj-2023-0013>
- Rosenberg, S. (2018). World Cup 2018: Russia promises 'unprecedented' security. BBC News. <https://www.bbc.com/news/world-europe-44466387>. Accessed on 19 Feb. 2023.
- Rosenzweig, P. (2013). *Cyber Warfare: How Conflicts in Cyberspace Are Challenging America and Changing the World*. Praeger. Santa Barbara, CA, USA. ISBN-10031339895X
- Roth, F. (2022). The Effect of the Financial Crisis on Systemic Trust. *Contributions to Economics*. Springer, Cham. https://doi.org/10.1007/978-3-030-86024-0_11
- Rowley, Jennifer (2007). The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information and Communication Science*. 33 (2): 163–180. doi:10.1177/0165551506070706.
- Russell, S. J. & Norvig, P. (2009) *Artificial Intelligence: A Modern Approach*, 3rd edition. Prentice Hall. Upper Saddle River, New Jersey, USA. pp 1-2. ISBN-13: 978-0-13-604259-4
- Sabato, L. & Kondik, K. & Shelley, G., eds. (2017). *Trumped: The 2016 Election That Broke All the Rules*. Lanham, MD: Rowman & Littlefield. ISBN 978-1-4422-7940-7.
- Samek, W., Wiegand, T., & Müller, K.-R. (2018). Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. *ITU Journal: ICT Discoveries, Special Issue No. 1*. International Telecommunication Union, 2018. pp 39.
- Schneier, B. (2015). *Data and Goliath: The Hidden Battles to Collect Your Data and Control Your World*. W.W. Norton & Company. ISBN-13: 978-0393352177
- Schönberger, V. M., & Cukier, K. (2013). *Big Data: A revolution that will transform how we live, work, and think*. New York, NY: Houghton Mifflin Harcourt.
- Sherwin, R. K., Feigenson, N., & Spiesel, C. (2006). Law in the digital age: how visual communication technologies are transforming the practice, theory, and teaching of law. *Boston University journal of science & technology law*, 12(2), 227. <https://doi.org/10.2139/ssrn.804424>
- Simon, J. (Ed.). (2020). *The Routledge handbook of trust and philosophy*. Routledge. ISBN: 978-1-138-68746-2
- Sky News. (2020). France: A timeline of deadly attacks after latest atrocity. Sky News. <https://news.sky.com/story/france-a-timeline-of-terror-10787264>. Accessed on 19 Feb. 2023.
- Smith, I. (2021). Low-income households 'priced out' of insurance market. Financial Times. <https://www.ft.com/content/2be0ee61-30cc-494d-9e6f-b8ba2749560d>. Accessed on 23 Feb 2023.
- Solomon, M. (2016). Social media and self-evaluation: The examination of social media use on identity, social comparison, and self-esteem in young female adults. William James College.
- South Korean Personal Information Protection Commission. Personal Information Protection Act; Act No. 16930. Korea Legislation Research Institute.
- Spadaro G., Gangl K., Van Prooijen J-W., Van Lange PAM, Mosso CO (2020). Enhancing feelings of security: How institutional trust promotes interpersonal trust. *PLoS ONE* 15(9): e0237934. <https://doi.org/10.1371/journal.pone.0237934>
- Staff Writer at Lmg Security (2022). What Hackers Do with Stolen Data & How to Reduce Your Risk After Data is Taken. <https://www.lmgsecurity.com/what-hackers-do-with-stolen-data-how-to-reduce-risk-after-data-is-taken/>. Accessed 20 Oct. 2022.

- Standaert, M. (2021). Smile for the camera: the dark side of China's emotion-recognition tech. *The Guardian*. <https://www.theguardian.com/global-development/2021/mar/03/china-positive-energy-emotion-surveillance-recognition-tech>. Accessed 20 Mar. 2023.
- Statewatch (2022). Public hearing at Dutch court over police surveillance of activist. <https://www.statewatch.org/news/2022/september/europol-told-to-hand-over-personal-data-to-dutch-activist-labelled-terrorist-by-dutch-police/>. Accessed on 15 Feb. 2023.
- Su, A. (2020). A doctor was arrested for warning China about the coronavirus. Then he died of it. *The Los Angeles Times*. <https://www.latimes.com/world-nation/story/2020-02-06/coronavirus-china-xi-li-wenliang>. Accessed on 20 Mar. 2023.
- Sullivan, J. & Brands, H. (2020). China Has Two Paths To Global Domination. *Foreign Policy*. <https://foreignpolicy.com/2020/05/22/china-superpower-two-paths-global-domination-cold-war/>. Accessed 20 Mar. 2023.
- Taylor, T. (2021). Hackers, Breaches, and the Value of Healthcare Data. Secure Link official website. 2021. <https://www.securelink.com/blog/healthcare-data-new-prize-hackers/>. Accessed 20 Oct. 2022.
- Tepedino, C. (2022). How Social Media Could Impact Your Insurance. *US Insurance Agents*. <https://www.usinsuranceagents.com/how-social-media-could-impact-your-insurance/>. Accessed on 26 Feb. 2023.
- Thomson, B. (2019). China ranks top of the world's 'Big Brother' states for its 'extensive' and 'invasive' use of biometric data belonging to citizens and tourists. *Daily Mail UK*. <https://www.dailymail.co.uk/news/article-7760657/China-No-1-Big-Brother-state-invasive-use-biometric-data.html>. Accessed 20 Mar. 2023.
- Thormundsson, B. for Statista. (2021). Artificial intelligence (AI) market size worldwide in 2021 with a forecast until 2030. <https://www.statista.com/statistics/1365145/artificial-intelligence-market-size/>. Accessed 30 Jun. 2023.
- Tyner, K. (2014). *Literacy in a digital world: Teaching and learning in the age of information*. Routledge.
- United Nations Conference on Trade and Development. (2023). Data Protection and Privacy Legislation Worldwide. <https://unctad.org/page/data-protection-and-privacy-legislation-worldwide>
- United Nations, Academy of ICT Essentials for Government Leaders (2020). Social Media, Development and Governance. Asian and Pacific Training Centre for Information and Communication Technology for Development.
- Van der Sloot, B., Wagenveld, Y. (2022). Deepfakes: regulatory challenges for the synthetic society. *Computer Law & Security Review, Volume 46, 2022*, 105716, ISSN 0267-3649, <https://doi.org/10.1016/j.clsr.2022.105716>.
- Velmurugan, A. & Mayan, J. & Niranjana, P. & Francis, R. (2020). Expense Manager Applications. *Journal of Physics: Conference Series*. 1712. 012039. 10.1088/1742-6596/1712/1/012039.
- Vincent, J. (2019). This app uses neural networks to put a smile on anybody's face. *The Verge*. <https://www.theverge.com/tldr/2017/1/27/14412814/faceapp-neural-networks-ai-smile-image-manipulation>. Accessed 21 Nov. 2022.
- Wall, D. (2007). *Cybercrime: The Transformation of Crime in the Information Age*, Polity Press 2007, ISBN: 0-74562735-8, 8-58
- Wallace, Danny P. (2007). Knowledge Management: Historical and Cross-Disciplinary Themes. *Libraries Unlimited*. pp. 1–14. ISBN 978-1-59158-502-2.
- Westerlund, M. (2019). The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review*. Vol. 9, Issue 11. pp 39-52. Doi: 10.22215/timreview/1282.
- Winner, L. (1993). Upon Opening the Black Box and Finding It Empty: Social Constructivism and the Philosophy of Technology. *Science, Technology & Human Values*, 18, 362 - 378.
- Yang, Yu. & Yang, Yi. & Fei Ju, Sh. (2017). China seeks glimpse of citizens' future with crime-predicting AI. *Financial Times*. <https://www.ft.com/content/5ec7093c-6e06-11e7-b9c7-15af748b60d0>. Accessed on 20 Mar. 2023.

- Yu, R. & Ali, G. (2019). What's Inside the Black Box? AI Challenges for Lawyers and Researchers. *Legal Information Management*. 19. 2-13. 10.1017/S1472669619000021.
- Zellers, R., Holtzman, A., Rashkin, H., Bisk, Y., Farhadi, A., Roesner, F., & Choi, Y. (2019). Defending against neural fake news. *Advances in neural information processing systems*, 32.
- Zhang, P. (2021). Biometric data collection: China is the most invasive user in the world, according to a 96-country study. South China Morning Post. <https://www.scmp.com/news/people-culture/article/3122187/biometric-data-collection-china-most-invasive-user-world>. Accessed 20 Mar. 2023.
- Zhu, J. (2022). The Personal Information Protection Law: China's Version of the GDPR?. *Columbia Journal of Transnational Law*. <https://www.jtl.columbia.edu/bulletin-blog/the-personal-information-protection-law-chinas-version-of-the-gdpr>. Accessed on 25 Mar. 2023.
- Zuboff, Sh. (2019). *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. PublicAffairs. ISBN-13978-1610395694.
- Zulkarnain, N. & Anshari, M. (2016) Big Data: Concept, Applications, & Challenges. *2016 International Conference on Information Management and Technology (ICIMTech)*. 306-307.