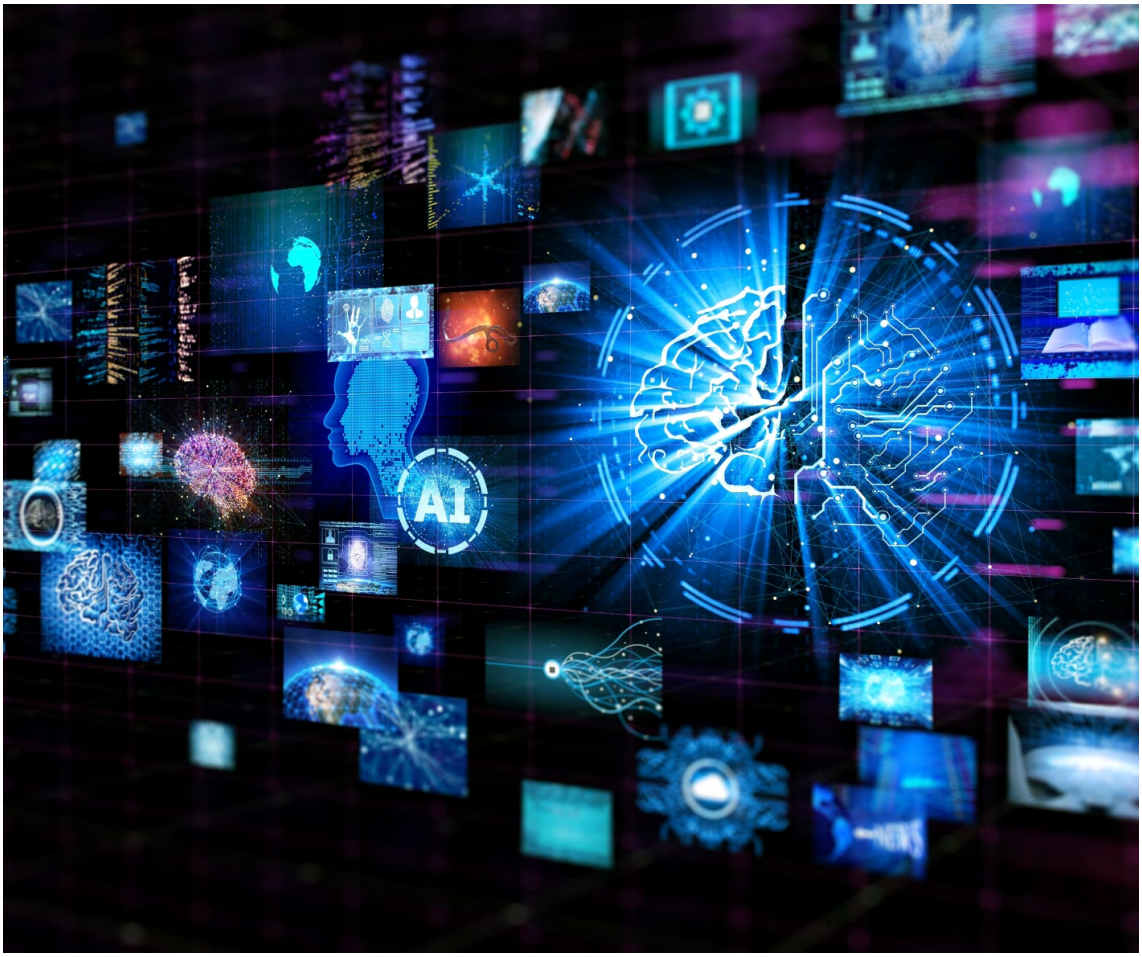
SMART CITIES, ARTIFICIAL INTELLIGENCE AND DIGITAL TRANSFORMATION LAW

A Handbook for Students and Professionals



Edited by E. E. Akin, S. Klimbacher, G. Ziccardi

ILS
INFORMATION, LAW & SOCIETY

Milano University Press

Contents

Introduction 11

PART I DIGITAL TRANSFORMATION LAW 13

_Hlk172038128 15

Chapter I. From Legal Informatics to Digital Transformation Law 15

_Hlk172038951 27

Chapter II. The Need for Digital Skills: Digital Literacy and Education to Technology 27

_Hlk172038987 49

Chapter III. The European Union Digital Strategy: GDPR, DSA, DMA and AI Act 49

Chapter IV. The Regulation of Data in the European Union: the Data Governance Act and the. 65

_Hlk172039131 93

Chapter V. The United States of America Approach to Digital Regulation 93

_Hlk172039155 107

Chapter VI. The Chinese Approach to Information Technology Law 107

PART II ARTIFICIAL INTELLIGENCE 113

_Hlk172039170 115

Chapter VII. An Introduction to Artificial Intelligence 115

_Hlk165732389 122

_Hlk172039185 123

Chapter VIII. Artificial Intelligence, Data Protection and Responsibilities 123

_Hlk153362412 139

_Hlk172039210 139

Chapter IX. Artificial Intelligence and Ethics 139

_Hlk153746347 152

_Hlk172039274 159

Chapter X. Generative Artificial Intelligence and Copyright 159

Contents (continued)

_Hlk165740415 167
Chapter XI. Artificial Intelligence and Healthcare 167
PART III SMART CITIES AND DIGITAL SOCIETY 181
_Hlk172039315 183
Chapter XII. The Idea and Creation of a Smart City 183
_Hlk172039348 193
Chapter XIII. Surveillance, Security, Resilience and Protection of Critical Infrastructures 193
Chapter XIV. Computer Crime on the Dark Web 203
_Hlk172039424 219
Chapter XV. The Regulation of Hate Speech, Antisemitism, and Terrorism Online 219
_Hlk172039457 241
Chapter XVI. Fake News, Conspiracy Theories, Misinformation and Disinformation 241
_Hlk172039483 255
Chapter XVII. Protection of Minors Online 255
_Hlk172039500 263
Chapter XVIII. Online Reputation and Right to Be Forgotten 263

SMART CITIES, ARTIFICIAL INTELLIGENCE AND DIGITAL TRANSFORMATION LAW

A Handbook for Students and Professionals

Edited By
Eylül Erva Akin, Simona Klimbacher and Giovanni Ziccardi

Milano University Press

Smart Cities, Artificial Intelligence and Digital Transformation Law: A Handbook for Students and Professionals / Edited by Eylül Erva Akin, Simona Klimbacher and Giovanni Ziccardi. Milano: Milano University Press, 2024. (Information, Law & Society; 1)

ISBN 979-12-5510-164-2 (print)
ISBN 979-12-5510-166-6 (PDF)
ISBN 979-12-5510-167-3 (EPUB)
DOI 10.54103/infolawsoc.186

Le edizioni digitali dell'opera sono rilasciate con licenza Creative Commons Attribution 4.0 - CC-BY-SA, il cui testo integrale è disponibile all'URL: https://creativecommons.org/licenses/by-sa/4.0

Le edizioni digitali online sono pubblicate in Open Access su: https://libri.unimi.it/index.php/milanoup.

© The Author(s), 2024
© Milano University Press per la presente edizione

Pubblicato da:
Milano University Press
Via Festa del Perdono 7 – 20122 Milano
Sito web: https://milanoup.unimi.it
e-mail: redazione.milanoup@unimi.it

L'edizione cartacea del volume può essere ordinata in tutte le librerie fisiche e online ed è distribuita da Ledizioni (www.ledizioni.it)

Table of contents

Introduction 9

PART I
DIGITAL TRANSFORMATION LAW

Chapter I
From Legal Informatics to Digital Transformation Law 13
by Giovanni Ziccardi

Chapter II
The Need for Digital Skills: Digital Literacy and Education to Technology 25
by Giulia Pesci

Chapter III
The European Union Digital Strategy: GDPR, DSA, DMA and AI Act 47
by Desideria Giulia Pollak

Chapter IV
The Regulation of Data in the European Union: the Data Governance Act and the Data Act 63
by Charlotte Ducuing

Chapter V
The United States of America Approach to Digital Regulation 91
by Simona Klimbacher

Chapter VI
The Chinese Approach to Information Technology Law 105
by Eylül Erva Akin

PART II
ARTIFICIAL INTELLIGENCE

Chapter VII
An Introduction to Artificial Intelligence 113
by Giovanni Ziccardi

Chapter VIII
Artificial Intelligence, Data Protection and Responsibilities 121
by Maria Grazia Peluso

Chapter IX
Artificial Intelligence and Ethics 137
by Simona Klimbacher

Chapter X
Generative Artificial Intelligence and Copyright 157
by Eylül Erva Akin

Chapter XI
Artificial Intelligence and Healthcare 165
by Malwina Anna Wójcik-Suffia

PART III
SMART CITIES AND DIGITAL SOCIETY

Chapter XII
The Idea and Creation of a Smart City 181
by Gabriele Suffia

Chapter XIII
Surveillance, Security, Resilience and Protection of Critical Infrastructures 191
by Pierluigi Perri

Chapter XIV
Computer Crime on the Dark Web 201
by Aleksandra Klimek-Lakomy

Chapter XV
The Regulation of Hate Speech, Antisemitism, and Terrorism Online 217
by Arianna Arini

Chapter XVI
Fake News, Conspiracy Theories, Misinformation and Disinformation 239
by Paulina Kowalicka

Chapter XVII
Protection of Minors Online 253
by Samanta Stanco

Chapter XVIII
Online Reputation and Right to Be Forgotten 261
by Simone Bonavita

Introduction

The "Smart Cities, Artificial Intelligence and Digital Transformation Law" course at the State University of Milan was established more than ten years ago as part of the Master of Laws in sustainable development taught in English.

It then became a Faculty of Law course of 6 credits (42 hours) open to students attending other faculties; due to the increasingly rapid technological progress and advent of artificial intelligence, the course has lately seen a significant increase in the number of students (around 100 last year).

Since its inception, the course focuses on three main parts, namely: i) regulations aimed at governing the digital transformation of our society, ii) artificial intelligence, and iii) smart cities.

The part on smart cities deals with the technical and legal problems linked to the increasing digitalization of cities and the consequent problems of regulation, cybersecurity, vulnerability of the Internet of Things, citizen profiling, creation of digital twins for governance and administration tests, privacy, and protection of sensitive data.

The lectures on artificial intelligence have rapidly attracted the interest of many students dealing with the ethical, social, and regulatory aspects of technology, focusing on what is happening in the world (especially in Europe, the United States of America and China) as well as on the opportunities and threats that technology may pose to human beings.

Finally, a more general part is devoted to the regulatory process that seeks to govern the changing technological society. This is the most complex part, as new initiatives related to innovative technologies, platforms and services are rapidly spreading everywhere.

Most course participants are international students with different backgrounds that bring to class their diverse home country experiences with technology, including the evolution of smart cities in Asian cities, the digitalization of public administration in Northern Europe, the politics of platforms and governments in the United States of America and China, but also issues connected with the technological divide in the global South, the use of technology to alter democratic balances and elections, the different perspectives of online hate and crime.

During the previous year, students prepared for their final exam by drawing from a list of open access papers published by various world-class scholars.

The teaching team of the course decided to write this handbook to make the syllabus and method of study more homogeneous, and to support students in an in-depth study of the main topics covered during our lectures. In addition, the essential bibliography will further support those students that wish to

research specific topics on any specific subject matter that is part of the law course.

It is not easy nowadays to crystallize in a "traditional" book the law aspects of new technologies. While maintaining a scientific rigor, the authors of the handbook focus on general aspects and principles, omitting detailed descriptions of laws and regulations (constantly changing, especially at a European Union level) that students will easily retrieve from institutional websites or large legal databases.

This handbook is designed for students, professionals and scholars studying the legal and technical aspects of the digital world.

We hope that publishing the handbook in open access will increase the number of students and professionals that may read it and encourage other academics to make study materials and course books available in open access format.

Professors around the world are welcome to adopt this handbook for their digital and non-digital courses.

Milan, August 2024

Eylül Erva Akin, Simona Klimbacher, Giovanni Ziccardi
(Editors)

PART I
DIGITAL TRANSFORMATION LAW

Chapter I
From Legal Informatics to Digital Transformation Law

by Giovanni Ziccardi*

Index: 1. The importance of studying these subjects. – 2. The advent of the idea of "Legal Informatics". – 3. A traditional bipartition. – 4. The connection to the philosophy and sociology of law. – 5. New research topics. – 6. A first international map of digital transformation law.

1. The importance of studying these subjects

The study of Legal Informatics, of the law of digital transformation, and of all the numerous connections between the world of technology and that of law, including the practical use of information technology in all the legal professions, has for decades been considered obligatory in the world's best universities.

It is considered an essential element, during the studies of a lawyer, a magistrate, a notary and, in general, of a law graduate, the knowledge, at least basic, of the law of new technologies and of the academic discipline "underlying the technological transformation and economics of the legal industry" (Katz, Dolin and Bommarito, 2021).

Such an awareness has not, however, always existed in the world of the legal professions: for many years, the more conservative circles considered information technology as a fad, a simple tool and practical element to be left, in its use, to the young trainees in the firm, or to secretaries.

In the meantime, however, society was rapidly changing, and all law was turning into "digital law" (Paliwala, 2010).

The advent of the smartphone, e-commerce, social networks and, ultimately, artificial intelligence put legal professionals in front of a new framework where, today, all legal issues have a strong ICT component that must be well understood by the interpreter. Think of the digital changes in family law, commercial law, criminal law, and so on (Schweighofer, 2008).

---

* Full Professor of Philosophy of Law (Legal Informatics) at the University of Milan and Director of the Information Society Law Center (ISLC).

In this first Chapter, which introduces the part devoted to the law that governs and seeks to regulate digital transformation, we will illustrate why it is essential today for a modern jurist to have a clear understanding of these issues.

We will quickly trace the evolution of Legal Informatics and the law of new technologies by identifying the research topics that have concerned it in the past, and the new perspectives.

From the very first lines, the reader will notice five clear elements:

1. A constant evolution of law in the often-unsuccessful attempt to keep pace with technological evolution. This problem of a legal regulation that, in view of the speed of technology and the slow pace of law, is in danger of "always arriving late", is a constant that scholars of the subject have had to deal with since the early years (Susskind, 2023).
2. A clear distinction, in doctrine, between "Legal Informatics" (interpreted as the use by the jurist of the informatics tools that are most useful in his/her daily activity) and "new technology rights" (understood as the attempts to regulate with norms the new technological phenomena that characterize modern society).
3. A commonality of themes, and problems, at an international level, even if linked to the technological evolution of each Country (which, in some cases, tries to solve a legal issue before others). Think of computer crimes (the regulation of which started in the United States of America and, then, arrived in Europe) or, on the contrary, the initiatives in Europe, first in the world, to regulate e-commerce, digital markets and services, large platforms, data protection and artificial intelligence.
4. Proceeding by "blocks" or by "broad categories" in regulating individual topics, often with little legislative coordination between the various parties involved and between national and supranational law. Thus, major topics are regulated: artificial intelligence and data protection, cybersecurity and e-commerce, cyber resilience of the banking and financial system, online terrorism, whistleblowing, and copyright.
5. A very broad and strong regulatory (and sanctioning) power left to supervisory authorities, competition rules, self-regulation (especially of platforms) and codes of conduct, in an attempt (often unsuccessful) to also involve the digital society actors in this regulatory process.

All this entails the need for a student of law, or for a professional eager to understand the digital society in which he or she now operates daily, to address these issues methodically and rigorously (R. Susskind and D. Susskind, 2022).

Here is, as an example, an interesting and broad definition of "Legal Informatics" by the scholar Ron A. Dolin (https://clp.law.harvard.edu/knowledge-hub/magazine/issues/legal-informatics/legal-informatics/):

> […] legal informatics can be defined as the science of information as applied to or studied in the context of law. Legal informatics contends with the development of technology in order to organize and make use of the vast amount of information that exists in the legal profession. As Bommarito writes, law is, after all, largely a knowledge business, which, until recently, relied on minds and paper. Legal informatics offers a lens through which that knowledge business might be transformed.

2. The advent of an idea of "Legal Informatics"

Regarding the notion of "Legal Informatics", scholars have always agreed on the basic content of the subject, while retaining its distinct practical sense.

Legal Informatics is considered a scientific discipline that studies how information and telematic technologies are applied to the theory and practice of law.

Legal Informatics should not be an informatics discipline, but a "legal" one; *i.e.*, a discipline to be studied and applied more by jurists than by information technicians, and whose study would demand a series of notions proper to computer science but aimed at legal application (Jori, 2006).

In fact, it would not be a matter of building computers or devising operating systems or programming techniques, but of researching the best ways to use them in the field of law.

This distinction between "Legal Informatics" and "Computer Literacy" in the broad sense (or the acquisition of computer knowledge) takes on considerable importance.

The study of Legal Informatics has, first, a purely didactic aspect: in this sense, it is understood, in the academic sphere, as an activity aimed at bringing the jurist closer to the computer, with a sort of computer-legal literacy, or at perfecting his computer and telematic knowledge with reference to certain evolved uses of the electronic machine in the legal field.

The second aspect of Legal Informatics is connected to pure research activity: in this case, in addition to eminently practical issues, such discipline includes topics of broad theoretical and interdisciplinary scope, incidentally concerning, for example, private law, administrative law, criminal law or trade and contract law.

In the modern era, Legal Informatics is understood as a means of illustrating information and communication theories to provide a simple and straightforward working tool for jurists.

A first noticeable change has taken place, for instance, in that area of study that practitioners traditionally refer to as Documental Legal Informatics, since it devotes special scientific attention to the concept of the "document": today, document informatics has lost the typical static connotation it had in its early

days and has taken shape, essentially, in the study of the management of legal documents on the web, on platforms, on the Internet and in large databases used in law firms, in the analysis and use of cooperative editing and content management tools (authoring), in the elaboration of tag or label systems (in many cases using artificial intelligence tools) in order to mark documents and make them recognisable to networked machines, especially within databases that dialogue with tools, engines and browsers, or to create cooperative spaces of users processing notes related to their research.

The topic of knowledge and documentality – or, rather, of the documentary heritage within large firms – is today central both in terms of its importance and its economic value. Much of the investment in legal tech is geared towards finding solutions for ever better automated document management. Some of the most technologically advanced systems have started, on an experimental basis, to implement knowledge-sharing tools and research results in these environments as well.

All of them have for some time now borrowed the technologies of the large generalist search engines, especially Google, to help the user in his search by means of suggestions, auto-completion of queries or predefined search paths, and the use of advanced predictive and artificial intelligence systems. The plan is to be able to predict even the most common queries, to speed up the consultation process and to ensure the accuracy of the result provided after the query has been entered.

A similar change has also occurred in the traditional sector of "legal databases" and databanks in general: today, Legal Informatics studies, essentially, large online databases, accessible via the Internet, institutional or private and, in some cases, fed directly, in their contents, by the administrations themselves and often related to very interesting phenomena such as *open access*. Big data, as is well known, are at the heart of today's entire social and technological system and are, therefore, also the subject of great interest for Legal Informatics.

Last but not least, the traditional legal management informatics, *i.e.*, the static use of the computer by the lawyer for the management of the profession or daily business (word processing, creation of spreadsheets, use of software for the management of the law office, client archives and fees), has been partly replaced by the dynamic use of the computer/smartphone/tablet connected to the Internet, e-mail, network technologies and office packages that can be used directly online and for functions related to telematic processes. In this case, the advent of the economic crisis first, and the pandemic later, has radically changed the habits of professionals, who are less and less tied to a place, or to the walls of the office, and are able to use cloud, laptops and sharing spaces also to reduce costs.

The traditional categories of information retrieval and management informatics have evolved to include, in the age of telematics, the analysis of the

functioning of the telematic process, the relationship between deontological rules and the online and social network presence of the legal professional, legal research on the Internet, the use of certified e-mail, digital signatures and computerised documents, the implementation of anti-money laundering and electronic invoicing systems and the safekeeping of documents, the process of optical archiving of paper material, especially in public administration, and the free availability of the Official Gazette and other types of legal information on the Internet possibly through standards that are open and understandable by all types of operating systems.

In addition to this, considerable resources, in today's legal framework, have been invested in systems for the creation of institutional databases on the web aimed at both the collection and provision of regulatory material and the telematic connection between subjects for the purpose of exchanging information, especially in the investigation phase.

3. A traditional bipartition

A traditional bipartition in the relationship between information technology and law is that between Legal Informatics in the strict sense and the rights (or law) of information technology.

The discipline concerning the rights of information technology is characterised by not being homogeneous, but by bringing together, under a single heading, a bundle of sub-disciplines that can be linked to one or more relevant subjects (Jori, 2006).

One might think, for instance, of the criminal law of information technology, the tax law of information technology, or the public law of information technology.

Usually, teaching and research activity in the field of computer law is conducted by specialists in the prevailing discipline, also because, in some cases, the branches of law are very far apart.

A further distinction could be that between "IT laws" and "laws related to information technology", the former being understood as legal regulations of IT activities and the latter as activities regulated by law with a strong IT component.

The topic of the boundaries between computer law and Legal Informatics has been debated for years.

Computer law, or information technology law, is that regulatory field of contemporary legal systems consisting of all the provisions aimed at regulating the new information and communication technologies, information, and telematics.

Similarly, the normative propositions and reasoning of legal theorists whose purpose is to analyse, interpret, systematise or criticise the normative field that regulates information and telematics would also fall under computer law.

The sources and structure of computer law are said to belong to the traditional branches of law; this same interdisciplinary character – or transversal spirit – that characterizes computer law has, however, given rise to a theoretical debate aimed at highlighting whether it is a sector of disorganised norms pertaining to different legal disciplines or whether it constitutes a unitary complex of norms (sources) aimed at regulating a well-defined object based on its own methodology, thus giving rise to an autonomous legal discipline.

4. The connection to the philosophy and sociology of law

Alongside Legal Informatics as a tool for "computer literacy" and the rights related to information technology, there is an area of research that has become increasingly important in the age of the Internet and that concerns concepts that are closely connected to the classical topics of the philosophy, sociology, and theory of law.

Alongside the logic and technique of law, traditional topics of legal philosophers, the advent of networks and artificial intelligence has introduced new reasoning, debates and questions related, for instance, to the rights of freedom – especially of manifestation of thought and with reference to online hate content –, the relationship between anarchy and regulation of cyberspace ethics and computer ethics, netiquette and deontology in the electronic environment, mass dissemination phenomena such as peer-to-peer and the emergence of online communities, problems related to digital identity, sexuality, and new forms of relationships and parasociety and paracommunity on the net. These phenomena, alongside purely legal and technical issues, constantly raise important sociological and philosophical interpretative questions.

The tumultuous spread of the Internet has given rise, first in the United States of America and, subsequently, all over the world, to a series of social phenomena and behaviours of enormous legal relevance linked to the existence of the network. These too are topics that the Legal Informatics scholar cannot avoid in their course of study and research.

These phenomena, in fact, have not only a legal aspect but also a very relevant sociological aspect, which obviously falls within the field of study of the sociology of law. On the Internet and on social networks, networks of relations and paracommunities form with both very new and very ancient aspects whose legal aspects can only be understood if their sociological aspect is also taken into account.

5. New research topics

The traditional lines of research in Legal Informatics have been enriched in recent times by topics closely related to the technological evolution that has taken place over the last ten years.

In the activities of the Coordinated Research Centre in "Information Society Law" (ISLC) at the University of Milan, at the "Cesare Beccaria" department of legal sciences, at least eight areas of research have been pursued as listed hereinafter:

1. Technological threats, computer crimes, information wars and digital investigations.

   In this area of study, often referred to computer crimes, digital forensics, digital investigations, information warfare or, again, cyberwarfare, are studied technological threats and their relationship with law and geopolitics, both from a substantive as well as a procedural and investigative point of view. Topics of study include the analysis of all threats to online users, to micro, small, medium and large enterprises, to critical infrastructures and to society in general, as well as forensics activities within the judicial process and for internal corporate investigative purposes.

   Some lines of research examine the technical and legal aspects of computer crimes, their history, their evolution, the regulatory framework, case studies, major trials and ongoing regulatory reforms; others analyse the technical and legal aspects of digital investigations (including anonymity), the use of encryption systems or hacking tools, technological frauds and the very varied world of online frauds (especially in pandemic times) via e-mail, text messages, voice messages, artificial intelligence and deepfake or "human" interlocutors. Particular attention is devoted to the banking world and the phenomenon of man-in-the-middle, ransomware and phishing, defence strategies and the psychological and criminological aspects of crime, spam, identity theft and the trade in people's and children's data on the black market, malware and its malicious potential, disinformation, and fake news, as well as the increasingly frequent attacks on IoT devices.

   The recent wars have also made the themes of online disinformation, deep fakes and war fought with electronic armies and not only with "physical" weapons of destruction highly topical: all topics that involve national security.

2. Politics, online hatred, and the renewed interest in computer ethics.

   This second area of research studies the relationships between law, politics, democracy, and technology, with a focus on hate phenomena, the use of social networks in politics, and the renewed interest in an ethical, and sustainable, use of technology and artificial intelligence. Some of the topics explored include online antisemitism, populism, polarization,

disintermediation, "mud machines", online political hatred, e-voting and the opportunities/security of the critical systems, the dangers of political discrimination using technologies, social sorting and social control of the user/citizen, the protection of the reputation of the citizen or politician online and the possibility of content removal, and conspiracy theories.

3. Minors and their online protection.
   The relationship between technologies and minors, especially in the five to sixteen age group, and with a focus on responsible use of technologies, is a very topical issue of great concern. One thinks of the constant presence of minors online, the spread (and generation) of hate and pornographic content among adolescents, dangerous challenges, mental health and social withdrawal, video games between imitation and addiction, cyberbullying, cyberstalking, child grooming, sexting, revenge porn and pro-ana groups.
4. The European Union and its technological future.
   Indispensable is the analysis of the single market of new technologies and the European Union's digital policy. The research includes the highlighting of European values behind technological reforms, the study of the regulation on artificial intelligence, the Digital Single Market and its development, data protection at the centre of EU action, copyright and content protection, public infrastructure security (NIS and NIS2) and platform law as well as liability profiles for content circulating online.
5. The hybrid jurist and new digital skills.
   In this research area, the aim is to draw the profile of the jurist of the future: hybrid, contaminated and with transversal competences. The focus is on the development of digital skills and the idea of training a new "hybrid jurist", developing and maintaining a personal, and corporate, awareness of the importance of ICT issues and digital transformation, as well as the need for quality dissemination on the most critical topics.
   Some of the lines of research are based on the idea of *interdisciplinarity* and the presence of hybrid figures/skills among jurists, on the relationship between the jurist and programming (the importance, for example, of "Coding for Lawyers" courses), on the skills needed in legal design, smart contracts, risk analysis and data governance, consulting legal databases and OSINT activities. Also central is the design, and creation, of corporate policies to make employees responsible for cyber risk, the consideration of the importance of correct disclosure on critical topics such as cybercrime, cybersecurity and digital skills, the dissemination of the idea of a safe law firm and a professional activity that is always attentive to security, as well as the need to increase skills in the Covid-19 emergency period and the acquisition of specific skills, including through data encryption and the activation of secure connections during communications in the professional sphere.
6. Smart cities, connected suburbs and the digitisation of public administration.

In this area of study, the theme of connecting cities, neighbourhoods and buildings is fundamental, with a focus on the suburbs and the move from the centre.

7. Digital products and the freedom of information, code and content.
   This area of study deals with digital content, its value, circulation, and use, especially from a freedom and inclusion perspective.
   Some of the lines of research include the legal protection of digital products and creative works, open data and linked open data, the *Creative Commons* movement, free software, the protection of information and inventions, the regulation of copyright, the right to knowledge, data of deceased persons and digital assets, legacy and inheritance.
   Central topics in this area are smart cities, their organisation and security, the problems of inclusion of "digital" citizens and the digital divide, connections in rural and suburban areas, interpersonal relations in the age of Covid-19, the security of public ICT systems and the health sector, the importance of big data in cities and related risks, and citizen profiling, the sensor society and the so-called "electronic body", video surveillance and facial recognition in their most evolved forms (including control and security), the idea of citizen data governance, data breaches, and the possible "leakage" of data, IoT and the connection of objects, the automotive between extreme connectivity and security and artificial intelligence, as well as issues of equality and discrimination.
8. The GDPR, cybersecurity, data protection and privacy.
   Central today is the importance of the general regulation on the protection of natural persons regarding the processing of personal data (GDPR), security measures, ideas of privacy, confidentiality, data governance in general and accountability in particular. Some of the lines of study include an analysis of the GDPR, its critical issues and its evolution. The past, the present and the future, adequate security measures, the idea of accountability and its practical application (also from a defense point of view), the main obligations, sanctions, the actions of the European Data Protection Authorities (also with a view to consistency in the application of sanctions in the various countries), cybersecurity and law, the concept of "privacy" and confidentiality of personal information and the concept of "security".

6. A first international map of digital transformation law

It is not easy to draw an international map that can guide the scholar in an overview of what is happening in the world regarding digital transformation law. The framework is constantly changing that the risk is to point to norms, or legislative policies, that have already changed.

There are, however, some very useful suggestions.

The first is to divide the world into "three large areas" and try to get a clear picture of how the European Union, the United States of America and China act, and how they regulate. These are the three geographical places to keep a constant eye on.

The European Union has, over the past ten years, started an incredible regulatory action on at least twenty different fronts. From the creation of a digital single market with the Digital Services Act and the Digital Markets Act to data protection with the GDPR, from the creation of the first legal framework on artificial intelligence with the AI ACT to the cyber resilience of the banking and financial sector with the DORA regulation, to the copyright directive and regulations on cybersecurity, hate content, and terrorism.

The framework, and the approach, of the European Union are interesting because it is a Continent that does not have large platforms but provides the data of Europeans to the whole world and has a strong legal tradition that puts the human being (natural person) and his dignity at the center of any regulation.

Many of the measures mentioned are also effective against large platforms that are not based in Europe, and this has caused inevitable friction in recent years.

The European framework must necessarily be compared, by the scholar, with the North American one, the land of large platforms and technology companies that has always had a more attentive approach to e-commerce and to the person seen as user or consumer.

In addition to the very interesting tradition of regulating computer crimes (with the CFAA, one of the first organic regulations in the world), there are numerous state and federal regulations on privacy and data protection, on the liability limits of providers and platforms, and on digital inheritance.

Finally, China adopts an even different approach: the idea of using technology also as a tool to control citizens for national security purposes. For example, the use of facial recognition, social sorting and the identification of all users in order to maintain strict control over every activity. It should be noted that in recent years, China has also issued specific, articulated regulations on data protection, computer crimes, cybersecurity, and deepfake and artificial intelligence.

Other parts of the world are inevitably influenced by these three main approaches. Particularly interesting for the interpreter is, for instance, the technological and regulatory situation in India (not only a land of excellent programmers but also very advanced in experimenting with artificial intelligence systems in the public sector and in the courts), Japan (with the topics of video games and robotics in the limelight) and in war zones (at the time of writing this book we are referring to Russia-Ukraine and the war in the Middle East) with the problems of using artificial intelligence for war purposes.

In addition, the coming months/years will be characterized by numerous electoral contexts where the problem of unlawful use of artificial intelligence will join the generation of fake news and disinformation with the aim of altering democratic balances and disorienting voters.

For scholars there is a need to have a clear picture that is truly international, and that goes beyond borders; this is why in this book we decided to keep an international approach and limit specific references to local regulations.

Bibliography

Jori, M. ed. (2006) *Elementi di informatica giuridica*. Torino: Giappichelli.

Katz, D. M., Dolin, R. and Bommarito, M. J. (2021) *Legal Informatics*. Cambridge: Cambridge University Press.

Paliwala, A. (2010) *A history of legal informatics*, Zaragoza: Prensas de la Universidad de Zaragoza.

Schweighofer, E. (2008) *Legal Informatics and e-Governance as Tools for the Knowledge Society*. Zaragoza: Prensas de la Universidad de Zaragoza, 2008.

Susskind, R. (2023) *Tomorrow's Lanyers: An Introduction to your Future*. 3rd edn. Oxford: Oxford University Press.

Susskind, R. and Susskind, D. (2022) *The Future of the Professions: How Technology Will Transform the Work of Human Experts*. Updated edn. Oxford: Oxford University Press.

Chapter II
The Need for Digital Skills: Digital Literacy and Education to Technology

by Giulia Pesci*

Index: 1. The concept of "digital skills". – 2. The digital divide and its consequences. – 3. Digital skills in the world. – 4. The process of digital literacy. – 5. The education to technology.

1. The concept of "digital skills"

1.1. Premises

Before analyzing the importance of digital education, it is necessary to understand the concept of "digital skills".

Digital skills are the ability to know how to use well information and communication technologies (ICTs).

In fact, the continuous, pervasive, and unstoppable spread of these technologies has resulted in the demand for more digital skills.

With the advent of the Internet and the development of information technology, which have especially developed since the 1990s, nowadays digital skills are needed in education, work, leisure, dealing with public administrations and institutions, in justice, to ensure our safety, and to inform us.

Think about how most people use technology today: we use cell phones, computers, tablets, sophisticated devices that can perform increasingly complex tasks.

Nowadays, therefore, acquiring increasingly high digital skills is proving to be extremely important.

The importance of digital skills emerged in the early 2000s and the European Union has always given special attention to the issue. In fact, in 2006 and 2018 the European Parliament and the Council of the European Union published the recommendations on key competencies for lifelong learning (2006/962/EC

* Research Fellow in Legal Informatics at the University of Milan and Research Fellow of the Information Society Law Center (ISLC).

and 2018/C 189/01) and, within the texts, defined what is meant by the term "digital competence".

The recommendation focuses on the skills citizens should develop and maintain to stay abreast of new contemporary challenges. Indeed, the dual role of education is highlighted: social and economic, because it is precisely education that has the role to "play in ensuring that Europe's citizens acquire the key competencies needed to enable them to adapt flexibly to such changes".

According to the European Union the key competencies are communication in the mother tongue and communication in foreign languages.

These first two skills can be further distinguished into: literacy and multilingual competences; mathematical competence and basic competences in science and technology; digital competence; personal, social and learning to learn competence - in other words, the ability to reflect upon oneself, effectively manage time and information, work with others in a constructive way, remain resilient and manage one's own learning and career; social and civic competences or "citizenship competence" (the ability to act as responsible citizens and to fully participate in civic and social life, based on understanding of social, economic, legal and political concepts and structures, as well as global developments and sustainability); sense of initiative and entrepreneurship (capacity to act upon opportunities and ideas, and to transform them into values for others); cultural awareness and expression (ability of being engaged in understanding, developing and expressing one's own ideas and sense of place or role in society in a variety of ways and contexts).

1.2. Definitions

"Digital competence" has been defined by the European Commission in this way: "digital competence involves the confident, critical and responsible use of, and engagement with, digital technologies for learning, at work, and for participation in society. It includes information and data literacy, communication and collaboration, media literacy, digital content creation (including programming), safety (including digital well-being and competences related to cybersecurity), intellectual property related questions, problem solving and critical thinking".

In addition, the last European recommendation mentioned above added "essential knowledge, skills and attitudes related to this competence", namely:

> (Knowledge) Individuals should understand how digital technologies can support communication, creativity and innovation, and be aware of their opportunities, limitations, effects and risks. They should understand the general principles, mechanisms and logic underlying evolving digital technologies and know the basic functions and uses of different devices, software, and networks. Individuals should take a critical approach to the validity, reliability and impact of information and data made available by digital means and be aware of the legal and eth-

> ical principles involved in engaging with digital technologies. (Skills) Individuals should be able to use digital technologies to support their active citizenship and social inclusion, collaboration with others, and creativity towards personal, social or commercial goals. Skills include the ability to use, access, filter, evaluate, create, program and share digital content. Individuals should be able to manage and protect information, content, data, and digital identities, as well as recognise and effectively engage with software, devices, artificial intelligence or robots. (Attitudes) Engagement with digital technologies and content requires a reflective and critical, yet curious, open-minded and forward-looking attitude to their evolution. It also requires an ethical, safe and responsible approach to the use of these tools.

The European Union continues to invest heavily in this area, and ongoing projects include digital skills for citizens; digital skills for ICT professionals; digital skills for non-ICT workforce; digital skills in education (digital skills for educators, youth and students); measurement of digital economy and society indicators.

It is the European Union itself that in its publications defines digital skills as "vital" for increasingly digitized societies and economies.

Other institutions are also extremely attentive to the issue. UNESCO, for example, writes on its website, in the area "Digital Competencies and Skills" that digital technologies have changed the way knowledge and information is accessed, shared and produced and emphasizes how the COVID-19 Pandemic has resulted in the rise of "digital transmission of knowledge" making digital skills crucial for participation in society, including lifelong learning and employment opportunities.

Currently, UNESCO projects focus on the following objectives: supporting digital competencies training for teachers and instructors; support accessible open and distance learning; supporting youth coding and digital skills training for entrepreneurship.

1.3. The Digital Economy and Society Index

On the other hand, regarding the European study area related to "Measurement of digital economy and society indicators", it is necessary to deepen the contents of the "The Digital Economy and Society Index (DESI)" developed by the EU since 2014.

This index is used to monitor progress towards meeting the digital goals of European States.

The data that emerge are used to measure and record the progress made by the member States in the digital sector, and finally, a graph is compiled summarizing the various levels achieved by the States.

All this work is, clearly, closely linked to the European Union's commitment to digitization. In fact, the EU is investing quite a lot of funds to support digital transformation.

The EU, in the context of national recovery and resilience plans, has earmarked EUR 127 billion for digital-related reforms and investments.

The DESI index is structured based on four factors: human capital; connectivity; integration of digital technology; digital public services.

"Human capital" refers to basic and advanced digital skills; regarding "connectivity", connectivity possibilities, speed and accessibility to the network are considered; with digital technology integration, the use of digital in business and commerce is assessed; and the factor "digital public services" measures the digitization of public administration (the "e-Government").

The chart with 2022 DESI indices and the ranking of European States according to the efforts they are putting in and their technological and digital developments is shown below.

Figure 2.1 European Commission, Digital Economy and Society Index (DESI) 2022.

We also read from the report published by the European Commission that "much of the EU population still lacks basic digital skills, even though most jobs require them. Even though most jobs require such skills". In this regard, and about digital skills, the EU also highlights in the context of the DESI index that the "Path to the Digital Decade" aims for at least 80 percent of citizens to have basic digital skills.

Speaking of the European Union and the approach to digital, it is inevitable to briefly anticipate a topic that will be addressed extensively within this Handbook: "artificial intelligence".

The EU, says, in fact: "the way we approach Artificial Intelligence (AI) will define the world we live in the future".

The issue that immediately arouses attention, however, is this: how can citizens be prepared to meet the challenges of artificial intelligence when, as we have seen, basic digital skills are lacking in many cases?

Artificial intelligence systems can be complex, sophisticated, and based on difficult computational processes. For this very reason, the commitment of institutions and Governments proves essential.

1.4. Sustainable development is also digital

With regard to both artificial intelligence and digitization in general, the key words used by the European Union are: human-centric, sustainability, and development.

This means that the EU is setting itself the goal of achieving high levels of social and economic development in digital but in taking these steps forward, it maintains its focus on sustainability and the possibility of keeping humans (and their rights) at the center of progress.

Speaking of sustainability, mention must be made of the 17 Sustainable Development Goals (SDGs) adopted by all United Nations Member States in 2015.

Through the "2030 Agenda", therefore, the United Nations has set a series of ambitious and important goals to be completed by 2030.

Figure 2.2 United Nations, Sustainable Development Goals (SDGs) - Agenda 2030.

Among these, in addition to fighting poverty, hunger, protecting health, promoting education, or achieving gender equality, goal number 9 makes explicit references to digital.

In fact, the goal "Build resilient infrastructure, promote sustainable industrialization and foster innovation" promotes investment in infrastructure (transportation, irrigation, energy, and information and communication technologies) to achieve sustainable development and to strengthen the capacities

of communities in many countries. This means that information and communication technologies also play an essential role in sustainable development.

Even today, basic infrastructure such as roads, information and communication technology or electricity is still scarce in many developing countries, and the lack of adequate infrastructure prevents access to jobs, information, and education, creating barriers to economic and cultural activities.

Among the goals the United Nations wants to achieve with goal number 9 there is the following: "significantly increase access to information and communications technology and strive to provide universal and affordable access to the Internet in least developed countries".

1.5. The challenges and opportunities of artificial intelligence

Google – in the "Google Cloud" section of the search engine – defines generative artificial intelligence as "the use of AI to create new content, like text, images, music, audio, and videos" and adds that "generative AI is powered by foundation models (large AI models) that can multi-task and perform out-of-the-box tasks, including summarization, Q&A, classification, and more. Plus, with minimal training required, foundation models can be adapted for targeted use cases with very little example data".

If we ask one of the most famous generative artificial intelligences, ChatGPT, whether artificial intelligence can be useful in the development of digital skills, it responds with several interesting points:

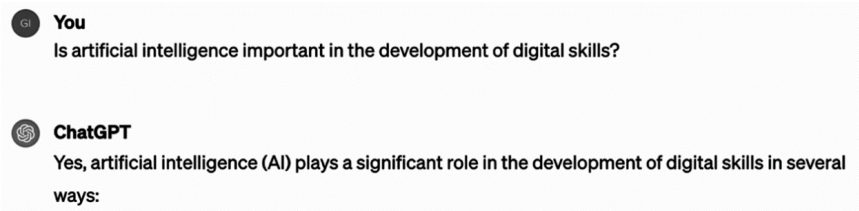


Figure 2.3 Interview with ChatGPT.

For digital skills, the benefits identified by ChatGPT are as follows: personalized learning tools; simulations and virtual learning environments; personalized tutoring and feedback; data analysis for improvement; automation of complex tasks.

But that's not all. The European Union clearly states that artificial intelligence is already part of our lives: "from using a virtual personal assistant to organise our working day, to travelling in a self-driving vehicle, to our phones suggesting songs or restaurants that we might like, AI is a reality".

The use of artificial intelligence is often easy and intuitive, but this does not mean that, even in this case, people actually know the functionality, risks, and features of AI systems.

Recent research conducted by the Organisation for Economic Co-operation and Development (OECD) about the impact of AI on the workplace also conducted surveys regarding "How is AI changing skill needs" (Lane, Williams, Broecke, 2023). The "main findings" of the research are as follows:

- while employers say that AI has increased the importance of specialized AI skills, they suggest that it has increased the importance of human skills and the need for highly educated workers even more so;
- employers appear to be addressing AI-related changes in skill needs primarily through training;
- most workers who use AI did not consider themselves to have specialized AI skills, more than 70% said that they were enthusiastic to learn more about AI. Broadly, the same groups of AI users who were more likely to already have specialized AI skills (male, university educated, aged 16-24) were also more enthusiastic to learn more;
- more than half of workers who use AI said that their company had provided or funded training so that they could work with AI. These workers were more likely to report that AI had improved working conditions and more likely to say that AI had increased wages in the sector, compared to those who had not received training. However, they were also more likely to report AI-related worries regarding job stability.

In conclusion, even because of this data, it can be said that artificial intelligence is playing an increasingly pivotal role in the world of work. This means that skills in the field must also be increased, especially through education and training.

The challenges may be many, the fears in relation to the future of the world of work, but, as noted, AI is now part of our lives.

It is therefore necessary, as far as possible, to know how AI systems work, to define legislative frameworks (as is happening, for example, with the EU AI Act), and to develop training and updating projects on the subject and its developments.

2. The digital divide and its consequences

The term "digital divide" refers to all the disparities that arise regarding digital.

The expression was first used in the 1990s in the United States precisely to define the disparity in opportunities for access to telematics services among the population of the US.

At the time, the Internet and digital services were just beginning to spread widely among citizens, and right from the start, differences in access to the Internet and in the ability to use these new tools were noticed. Today, the digital divide is still there and, in some cases, noticeable.

2.1. Characteristics of the phenomenon

The digital divide can be considered according to two aspects. The first concerns access to the net and the availability of technological tools, while the second concerns the presence of digital skills.

The phenomenon, then, is considered and analyzed according to the people it affects. One can speak of an intergenerational digital divide when age differences are considered and, in this case, it is often the older people who are part of a greater gap. There can be, then, the "linguistic-cultural" digital divide that affects migrant people, the "gender digital divide" when it mainly affects women. The digital divide, then, may affect people with low levels of education, with disabilities, or who are detainees.

The digital divide is, therefore, the "gap" in the population's access to digital and ability to use it by exploiting its potential.

The phenomenon, then, can be further analyzed according to three other aspects. It can be "global" when measuring the degree of digitization of each country compared to other countries, it can be "social" when analyzing the inequalities within each country, or "democratic", when the use of technology results in differences regarding people's participation in political and social life.

According to the latest 2023 data published by the International Telecommunication Union (ITU) – the United Nations agency specializing in information and communication technologies-ICTs –, 2.6 billion people of the global population remain offline and 67 percent of the world's population is now online.

In conclusion, then, digital disparities may include deficiencies from an infrastructural perspective (when there is a lack of access to the Internet, devices capable of connecting to the network, etc.) and a lack of IT knowledge.

2.2. Network access as a fundamental right

The Internet and technology are, by now, essential tools for every person's life. The net provides access to information, work, and education. The Internet makes it possible to exercise basic human rights.

Globally, therefore, there is a growing awareness of making digital technologies accessible to as many people as possible, precisely so that no one is left behind in this race towards the future.

Network access, however, today represents a fundamental right, precisely because it enables the exercise of different essential rights. Those who are, therefore, affected by the digital divide find themselves excluded from several essential services. Also, digital identity today represents an essential aspect of people's lives.

These gaps had a huge impact during the most difficult months of the COVID-19 pandemic, when most people found themselves in situations

where travel was restricted and digital was the only way to study, work, and communicate.

Those who do not have access to the net or those who lack the skills to use digital technology limit their right to information and expression; reduce their knowledge; may have less access to work; and may be restricted in their economic and government relations.

It is important, therefore, not to assume that everyone has the same opportunities to access the network and obtain the benefits offered by digital.

For these reasons, digital should no longer be a "luxury" or an option for the few. Major efforts are needed from the point of view of infrastructure, in the instruction and the educational sector, at the political and social level.

2.3. Inclusion and equity in artificial intelligence

Differences in access to technology are now also affecting the use of artificial intelligence.

UNESCO, in one of its recent publications "Guidance for generative AI in education and research" (2023) shows the most important aspects of the issue, relating them to the fourth UN's Sustainable Development Goal.

The critical importance of inclusion must be recognized and addressed throughout the generative AI life cycle as well. More specifically, GenAI tools must be made inclusively accessible.

Therefore, it is deemed necessary to:

- identify those without access to Internet connectivity or data, and take action to promote universal connectivity and digital skills to reduce barriers to equitable and inclusive access to AI applications;
- establish sustainable funding mechanisms for the development and provision of AI-enabled tools for students with disabilities or special needs;
- promote the use of AI to support students across the lifespan, at whatever age and wherever they are;
- develop criteria for the validation and operation of GenAI systems to ensure that algorithms do not incorporate gender bias, discrimination, or hate speech;
- protect linguistic and cultural diversity when using GenAI in education and research.

AI is getting closer and closer to the lives of citizens around the world and is characterized by increasingly pervasive systems capable of having a significant impact on each person's life. For this reason, it is important to always keep the level of attention on the issue high and to observe how governments and institutions in all parts of the world respond to all of this.

3. Digital skills in the world

Although mention has been made of the digital divide and the use of the Internet and technology around the world, it is necessary to briefly mention some characteristics of digital skills in other countries not yet analyzed within this Chapter.

As it turns out, important differences can be seen within each state in their approach to digital.

In the United States, for example, it resulted from an analysis conducted by the OECD ("Skills Matter. Additional Results from the Survey of Adult Skills", 2019) that one in three workers lack adequate digital skills.

Although the United States is known as the home of Silicon Valley, or the place where some of the most important technology breakthroughs have been made and pursued, digital skills also deserve special attention and efforts to be implemented in this region of the world.

Moving to the other side of the world, India is a country that has been racing rapidly towards technological development in recent years. India is investing heavily in the sector, improving access to the Internet and digital technologies, investing in software development and school and academic training, and developing digital services. Through the "Digital India" initiative, indeed, the country has set ambitious goals, such as promoting digital literacy, improving access to digital services, and encouraging the adoption of digital technologies in various sectors, including education, healthcare, and e-governance.

Other countries, such as Japan, China, and the Republic of Korea, are also experiencing significant and steady advances from a technological standpoint, and some of them will be examined in depth in subsequent chapters of this Handbook.

Moving further to the African continent, we briefly mention two digital skills development projects launched in 2023 by the Global Education Coalition (founded by UNESCO) and the Ivorian Ministry of Technical Education and Vocational Training (METFPA). The projects aim to strengthen digital skills and train in digital marketing and entrepreneurship. Over the course of 2024, the training in question will be delivered to help more than 300 young students find employment and to prepare an estimated more than 300 ministry leaders and teachers for the digitization of the country's education system.

We bring up this recent example precisely because, even in the statements of those involved, we read the following statements:

> digitization is an engine of transformation and development that allows us to adapt to the realities of the modern world (...) The digitization of our administration and the practical use of digital skills offers many benefits for our TVET sector - [it can] improve the efficiency and quality of our educational services, offer new teaching and training methods, promote innovation and creativity, and strengthen the employability of students.

Currently, digital skills development opportunities for TVET teachers/trainers in sub-Saharan Africa are very limited, resulting in poor digital skills of teachers compared to TVET teaching staff in other regions of the world. UNESCO, however, points out that the digitization of education is a key policy priority in Africa.

As we have seen in the previous sections, digital skills to be developed need a lot of effort and many practical aspects that need to be considered, such as, for example, having access to the net.

Returning to consider the digital divide and its consequences, it is important to consider the differences that exist in different areas of the world and how to reduce inequalities, including the digital ones.

4. The process of digital literacy

The term "literacy" identifies continuous learning in individuals. Learning concerns the development of their knowledge, potential and skills. Literacy means also the ability to read, write and, in a broad sense, the possession of education and knowledge of a given subject.

It has just been pointed out that differences in the digital divide may also relate to the presence or absence of digital skills.

UNESCO defines digital literacy in this way:

> digital literacy is the ability to access, manage, understand, integrate, communicate, evaluate and create information safely and appropriately through digital technologies for employment, decent jobs and entrepreneurship. It includes competences that are variously referred to as computer literacy, ICT literacy, information literacy and media literacy (Antoninis, Montoya, 2018).

Digital literacy, then, consists of the set of skills used to locate, understand, and use information with information and communication technologies.

These skills must be considered both from a technical point of view (thus, computer and digital literacy) and in terms of the correct, ethical, and responsible use of digital content and tools.

It has been pointed out that computer literacy and digital skills are essential within societies today.

These disciplines are, by now, the subject of attention by governments and institutions and subjects of study at the school and academic levels. Nowadays, disciplines related to digital literacy are included in the curricula at every level and are studied both as a subject and in the context of other teachings.

A key word in this regard is undoubtedly "lifelong learning", meaning that this lifelong learning must be continuous and constantly developing in each person's life. Indeed, digital, information technology and technology in general

are constantly developing and growing. This requires continuous updating and a constant study of these sophisticated and essential subjects.

Digital competencies encapsulate within them digital literacy, the ability to use data, communication, digital content creation, and all related activities exercised in a safe and conscious manner.

4.1. The Digital Decade policy program of the European Union

For the European Union, digital society and technologies bring with them "new ways to learn, entertain, work, explore, and fulfil ambitions", they also bring "new freedoms and rights, and give EU citizens the opportunity to reach out beyond physical communities, geographical locations, and social positions".

The European Union's agenda for the digitalization process in these years can be summarized in four points: (i) a digitally skilled population and highly skilled digital professionals (more graduates and ICT specialists + gender balance and 80% of adults can use tech for everyday tasks); (ii) secure and sustainable digital infrastructures (gigabit connectivity for everyone, high-speed mobile coverage, fast data access); (iii) digital transformation of businesses (75% of companies using Cloud, AI or Big Data, double the number of "unicorn startups", 90% of SMEs taking up tech); and (iv) digitalization of public services (Key Public Services 100% online, access health records online and eID for everyone).

4.1.1. The advantages of media literacy

For the European Union (2023/C 66/02), media literacy plays a critically important role:

> media-literate people are able to make informed choices, understand the nature of content and services and take advantage of the full range of opportunities offered by different communication technologies. They are better able to protect themselves and their families from harmful or illegal content. Media literacy can also serve as a valuable tool for combating the spread of disinformation by enabling users to critically assess the source of information and thus detect false or misleading content (…). Media literacy therefore empowers people to participate in a more open and informed democratic debate.

The European Union, with the Digital Education Action Plan (2021-2027) promotes a common vision of high-quality, inclusive, and accessible digital education in Europe. Indeed, the EU aims to support the adaptation of Member States' education systems to the digital age and its characteristics.

A further important reflection reported in the Communication from the EU Commission "Guidelines pursuant to Article 33a(3) of the Audiovisual Media Services Directive on the scope of Member States' reports concerning measures for the promotion and development of media literacy skills" is about the aforementioned importance of not limiting media literacy to learning about

tools and technologies, but must also be about equipping people with the critical thinking skills needed to make judgements, analyze realities complexities and recognize the difference between opinions and facts.

With these goals, and through the Audiovisual Media Services Directive (Directive (EU) 2018/1808), the European Commission requests Member States to report on legislative and other measures of a legal nature in place as well as on those possibly planned for the promotion and development of media literacy. In addition, as far as possible, Member States should also provide information on broader policy measures, such as national and/or regional strategies and action plans for the promotion and development of media literacy.

Building on the legal and policy measures of the European Union and individual Member States, they must implement the organizational measures through public funds and other funding mechanisms for media literacy and engage in engagement and awareness-raising activities, including media literacy programs.

Because of, and in the wake of, the multiple damages created by the Coronavirus health emergency, the European Commission, the European Parliament, and EU leaders, agreed on a recovery plan for the Union and its Member States. This new growth and recovery policy uses the financial instrument called NextGenerationEU, a € 750 billion facility created for a "sustainable, uniform, inclusive and equitable recovery".

Figure 2.4. European Commission, A Europe fit for the digital age. Empowering people with a new generation of technologies.

To track progress in implementing the plan, the European Commission launched an online public platform in 2021.

Digital plays an essential role in the plan's goals: "it is now time to get to work, to make Europe greener, more digital and more resilient" and Member States are working hard towards this direction. In Italy, for example, 27 percent of the total resources of the National Recovery and Resilience Plan are dedicated to digital transition.

4.2. Some reflections

Digital literacy therefore affects everyone, regardless of age or educational level.

Certainly, for people who already have a knowledge base in the field, learning may prove easier, but the focus must be on everyone.

To make effective use of digital education technologies, it is essential to build digital capacity at all levels of the education system, among educators and institution leaders, but also among students, parents, and administrators. Public authorities should support education institutions in selecting the right digital tools to meet their needs, facilitate their interactions with innovative education technology solutions and empower leadership teams to build a culture of digital education in schools and higher education institutions. Strengthening capacity among local authorities and within the wider education ecosystem can further support the successful implementation of digital education policies. Most importantly, supporting educators, promoting peer-learning, and offering continuing professional learning opportunities will be critical to ensuring that digital education technologies are used to advance quality and equity in education.

As a final thought on the topic of digital literacy, it is necessary to reflect on the skills possessed by those who work in digital and those who are involved in making decisions on the topic of ICT and rights related to new technologies.

This aspect is of paramount importance and, the key points, are the following: 1) the creation of appropriate and up-to-date courses of study; 2) constant updating on the subject; 3) a curious and not "know-it-all" approach: information technology is constantly developing; and 4) the selection, in the work and political spheres, of people who are truly experts and prepared on the issues.

The preparation of experts in the field and, above all, of those who are called upon to make political, legislative, and social decisions on the issue is really important. The issues are complex, the problems that may arise concern people's fundamental rights, and technological development requires increasingly rapid and appropriate responses.

4.3. The importance of starting with the basics: children's digital literacy

If we consider digital literacy starting from the base of education, then considering children and their formation, specific aspects need to be assessed regarding the early years of education.

The main problems noted for digital literacy are as follows: absence of sufficiently trained trainers, lack of ICT infrastructure, connectivity problems and lack thereof in remote and underdeveloped areas, and often absence of proper attention from policymakers.

Moreover, relevant curricula and training programs should adapt to the context in which they are applied, considering individual characteristics, and working with impact assessments.

Focusing on children, then, three essential aspects should be considered. The first is that digital literacy should not only be technical; children can have fun, play, and socialize through technology as well; therefore, the approach should

be friendly and curious. The second concerns digital literacy across the board: ICT is increasingly pervasive and present in everyone's life, even from the earliest years of life. This means that children should be put in a position to use digital across the board in their lives, even as they develop other skills. The third, on the other hand, relates to safety: children should be digitally literate even when they are not connected; they should be supported and assisted in their online operations and protected from the activities of profiling and indiscriminate collection of their data by increasingly sophisticated systems.

However, one should not only consider basic digital skills or those to be acquired by children.

In fact, in recent years, the academic field in ICT has been expanding as well. The world of work internationally demands increasingly advanced and sophisticated digital skills. Skills that are intertwined: from the world of computer science, computing, law, ethics, data science, economics, marketing and so on. Thus, it is an ever-expanding field that will find its greatest development with artificial intelligence and the challenges it will continue to bring.

For these reasons, it is important to continue to promote an ongoing, attentive, and up-to-date learning style and that policy makers should have a better understanding of the challenges promoted by digital and the characteristics of digital literacy.

Digital technologies are, therefore, a key resource for education and training systems today. When used properly and effectively, they also enable improved teaching processes and promote equity efficiency and inclusion.

In recent years, investment in technology in education has increased worldwide. Since the almost forced adoption of technology during months of "distance learning", there has been an unprecedented increase in its use. In fact, "hybrid learning" is still widely used in education, especially at the University and in professionalizing courses.

We need to ask ourselves, therefore, how educational systems can make the best use of digital technologies.

We have seen that there is a need for constant updating by educators and for major investments (including economic ones) in the field. The sharing, then, of best practices may prove essential.

4.3.1. "The digital revolution is shaping children's lives"

In its latest analysis "Analyzing how digital transformation affects children and advancing policy recommendations", UNICEF reiterates a clear concept: the digital revolution is shaping children's lives in profound ways. Given the pervasiveness of the web and the frequency with which children also use the Internet to communicate, play, work in school, access information and express themselves, the following question must be asked: how can we maximize

children's well-being through the positive use of digital technologies while mitigating the risks of harm?

UNICEF answers this question by developing recommendations and policy advice for governments and the private sector.

To answer the very complex question, UNICEF analyzes the main emerging issues:

- Neurotechnology and children: the potential benefits of neurotechnology are revolutionary (just think of its use in the medical field), but the risk to privacy and free will is high.
- For this reason, UNICEF experts believe that it is of paramount importance to prohibit "neuromarketing" involving the study of consumers' sensorimotor, cognitive, and affective responses to marketing stimuli towards children.
- Digital inclusion for children: digital inclusion goes beyond the concept of the digital divide and encompasses a broader set of barriers (social inequalities, limited local content, differences in the way digital platforms and frontier technologies are used).
- The metaverse and extended reality: the technologies behind virtual environments (artificial intelligence and extended reality devices), are becoming increasingly sophisticated and mainstream.
- Artificial intelligence: artificial intelligence systems are diverse and present in so many areas (filters on social, apps, IoT...) even in contact with children.
- Digital misinformation and children: children have cognitive abilities that are still developing and are particularly vulnerable to the risks of misinformation. We need to keep this information in mind and consider, at the same time, that young people may also be able to play an active role in countering its flow and mitigating its negative effects.
- Young people's digital civic participation: online, children and adolescents can access social movements, mobilize on activism issues, and actively participate in society, however, we need to ask how well children and adolescents have the digital and civic skills needed to relate to these digital environments.
- Digital literacy: every person should be able to make the most of the opportunities offered by digital and the Internet. This is especially true for children and young people, but they also face particular risks when they are connected. For this reason, digital literacy from the earliest years of education plays a key role.

Indeed, as UNICEF states, "investing in children's digital literacy means building more responsible, employable and tolerant future world citizens".

As will be discussed in more detail within this Handbook, the Internet and technology can present significant risks to children, but, again, UNICEF,

emphasizes that digital education can help children and youth experiencing certain difficulties take a leap towards a brighter future.

Problems that may affect children and young people include the following: conflict, disaster, or poverty; girls kept out of classrooms; young people excluded from school because of a disability.

At the Transforming Education Summit (TES), UNESCO and UNICEF launched Gateways to Public Digital Learning, a global initiative for digital learning and transformation.

The global initiative aims to work with governments and partners to make digital education a public good through high-quality, inclusive national digital learning platforms and content. The initiative, as noted on its website:

> will map, describe and analyze national public platforms and content; help countries create and strengthen national platforms; identify and showcase best practices; and establish international norms and standards to guide the development of platforms in ways that advance national and international goals for education.

The division of analysis into two distinct themes frequently occurs in the field of artificial intelligence as well: on the one hand, the advantages proposed by these tools, including in the world of education, and on the other hand, the risks that poorly informed or distorted use might entail.

4.3.2. AI in education

Thanks in part to the advent of generative AI, there has been considerable concern about the possible impact of AI in education.

UNESCO, on the topic reports, in the publication "Education in the age of artificial intelligence" (2023) that the education community needs important support to understand the risks generated by AI and measures to best protect children and education.

It is therefore important that teachers and educators also develop digital and AI skills to use programs capable of supporting and improving their activities (lesson planning, preparation of materials, feedback to students, and evaluation of assignments). Developments in studies of the subject show that AI can simplify teaching, personalizing learning, and saving time.

We need to be aware, however, of the tools used: generative AI applications today are still capable of creating content that is inaccurate or dangerous because it is false or misleading.

Finally, AI often has an economic cost: this, too, needs to be considered by policymakers and institutions. Ultimately, UNESCO also highlights the environmental impact of AI technologies, which require a significant expenditure of energy resources.

In conclusion, again, UNESCO reports this important observation: "international bodies have already significantly engaged in major efforts to shape

ethical and regulatory frameworks related to AI. It's crucial to ensure that education is equally protected".

4.3.3. Making the Internet safe for children

The European Union, as we have seen, is particularly focused on protecting children online and ensuring a safe online environment for them. Since 2012, the EU has been working on the issue through the "European Strategy for a Better Internet for Kids".

Realizing the frequent lack of adequate digital skills, it is necessary for the EU to improve media literacy and online safety education for children in schools. In fact, digital skills, and competencies, including media literacy and understanding how personal data are used, are crucial for today's minors. These skills enable minors to learn, connect, and contribute in active and informed ways to the world around them.

As has already been stated, these skills must also be possessed by adults responsible for minors precisely to guide and assist them.

The BIK+ strategy, therefore, proposes actions that focus on three pillars:

- safe digital experiences to protect children online from harmful and illegal content, conduct, contact, and consumer risks and to improve their online well-being through an age-appropriate and safe digital environment;
- empowerment and digital responsibility so that children acquire the skills and abilities needed to make informed choices and express themselves in the online environment safely and responsibly;
- active participation, respecting minors, who will be given a voice in the digital environment, with more activities led by minors to promote safe, innovative, and creative digital experiences.

To implement this strategy, the EU relies on cooperation and coordination at the European and international level, including with the help of associations such as the UN, UNICEF, OECD, Council of Europe, and NGOs working on children's rights.

5. The education to technology

Familiarity with digital does not mean actual competence.

This means that even if a person uses a digital device daily, he or she may not really know the tool and its risks.

For this reason, it is essential to be curious about technology. Curiosity must go hand in hand with awareness and education.

Starting with the basics, a person approaching digital should know the hardware and software parts of the device they are using. By hardware parts we mean knowledge of, at least, the basic functionality and characteristics of the

parts that make up the tool. When we focus, however, on the software part, we consider the computer programs that are run by the device.

As anticipated at the beginning of the paragraph, often the new generations, being "digital natives" demonstrate a strong familiarity with digital and technology. This does not mean that they really know their way around the digital and connected world.

In addition to the strictly technical aspects, it is important to know two other essential elements: data collection processes and digital-related rights.

This, then, represents the key tripartition for the informed use of IT and network tools:

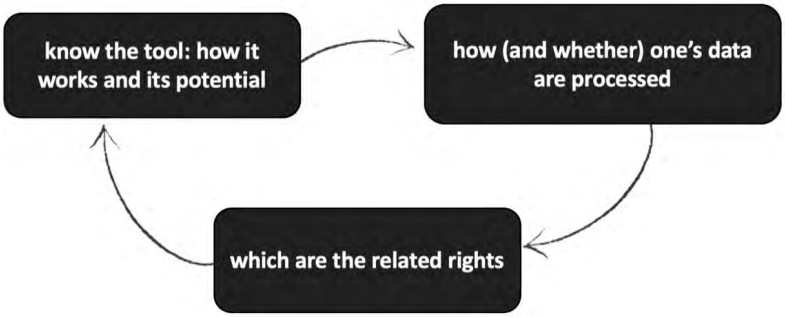


Figure 2.5 Informed use of IT and network tool.

Knowledge of these three aspects is of essential importance for the conscious, satisfactory, and efficient use of new technologies.

Knowledge of these aspects makes it possible to understand the risks and problems that can develop in connection with distorted or uninformed use of the net, like misinformation, online hate, threats to the respect of one's digital identity and the protection of children online.

A person's digital identity online represents both the totality of a person's online information and the protection of one's personality even online. In fact, it is important to consider the "online world" as an integral part of a person's life as he or she performs transactions online (from purchasing on e-commerce sites, to e-government activities, to the use of social networks and interaction with artificial intelligence systems). What happens online can have direct repercussions on people's lives, and for this reason, there is a need to be aware of and knowledgeable about the three areas mentioned above.

In technology education, efforts are, today, needed.

Efforts that are already put in place by institutions, universities, education systems, and the educational community. We must continue to work to create a network of knowledge that is increasingly advanced, accessible, and easy to share.

Starting with the basics, then, people who have no expertise in this regard (*e.g.*, children or those with low digital literacy) need to be accompanied in their approach to digital and network. The potential of digital, its features, risks and problems must be explained to them. They need to be made aware of their rights, the rights that can be violated by misuse of the digital, and the rights they can exercise to protect themselves.

More and more advanced training in these areas should continue to be encouraged, then: digital is the present and the future, and the demands in the workplace are constantly expanding.

In conclusion, we can summarize in categories the digital skills that should be learned and enhanced today:

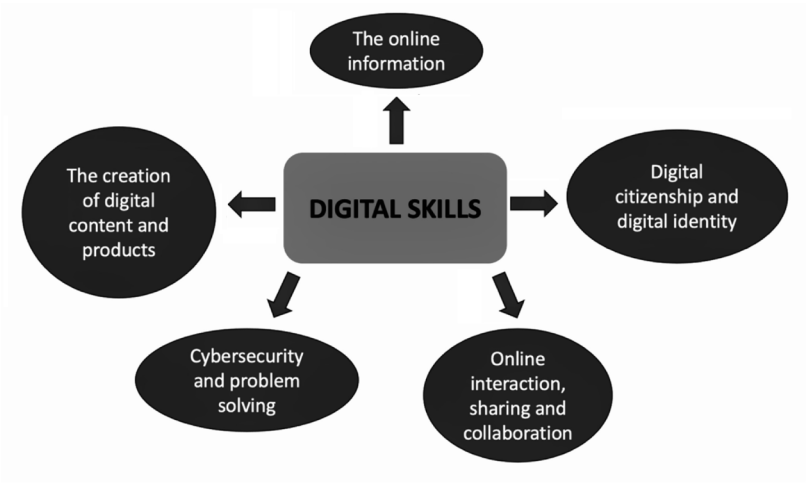


Figure 2.6 Categories of digital skills.

The area of online information includes: the ability to inform oneself online and the ability to obtain information online; recognizing and defending against misinformation, disinformation, and fake news; knowing how to recognize fake content, such as deepfake content; knowing how to evaluate the quality of information obtained online; archiving, processing, and deleting information online.

The creation of digital content and products is constantly increasing. The creation of digital products through increasingly sophisticated tools, such as artificial intelligence tools, must also be considered. This category includes production activities, related rights and licensing of use, and the ability to process and reprocess digital content and share it.

Cybersecurity and problem solving are about the security of people, infrastructure, and data.

Problem solving enables people to know how to act consciously and effectively online, including at work.

Cybersecurity, on the other hand, is a very broad area that needs to be considered at the international, state and government levels, in the workplace and on a personal level.

One must consider the security of networks, technological and network-related infrastructure, data transfer, and devices. Security is about protecting one's data, protecting against cyber-attacks and malware.

Online interaction, moreover, concerns the activities of interacting with devices and with other people or groups of people using devices. Interaction can start from devices such as smartphones, PCs, tablets, the Internet of things and augmented reality systems. One needs to know the devices one is using, pay attention to the rules, policies, language to be used, and how to interact online.

Interaction, then, can also include sharing and collaboration activities. Personal contexts, social networks, work and study situations, and relationships with government and authorities are involved.

The last category analyzed, digital citizenship and digital identity, includes all participatory and active citizenship activities. Citizenship, therefore, is also exercised online.

Digital citizenship includes digital rights, the exercise of rights including through the web and technology, the protection of fundamental rights online and through the web, access to public administration information and open data, interaction with authorities and public administrations, and the promotion of initiatives and tools available to citizens.

Digital identity management includes creating digital identities, protecting one's identity and reputation online, protecting one's data, and all activities to protect people online.

Chapter III
The European Union Digital Strategy: GDPR, DSA, DMA and AI Act

by Desideria Giulia Pollak*

Index: 1. Introduction: The EU Digital Strategy. – 2. The General Data Protection Regulation (GDPR). – 3. The Digital Services Act package. – 4. The Artificial Intelligence Act (AIA). – 5. Other regulations.

1. Introduction: the EU Digital Strategy

In an era defined by rapid technological advancement and digital transformation, the European Union has emerged as a frontrunner in shaping policies and regulations to govern the digital landscape. With the proliferation of digital technologies permeating every aspect of society and economy, the EU has recognized the need for comprehensive frameworks to safeguard individuals' rights, promote fair competition and foster innovation while addressing the challenges and risks associated with digital transformation.

Central to this effort is the EU's aim to create a single, integrated digital market (Digital Single Market) across member states where goods, services, and data can flow freely, and where high-speed broadband and 5G connectivity are available to all individuals, bridging the digital divide and promoting digital inclusion and accessibility.

Actions are being taken to strengthen the EU's digital infrastructure (including cloud computing, data centers, and cybersecurity capabilities), and significant attention is being paid to education initiatives directed at improving digital literacy and upskilling. Recognizing the value of data as a driver for innovation and economic growth, the EU seeks to promote a responsible and secure use of data and is committed to protecting individuals' digital rights while ensuring transparency, accountability, and ethical standards in the use of digital technologies, including artificial intelligence.

The regulatory interventions illustrated below aim to put an end to the existing legislation fragmentation in the EU, giving rise to a greater law harmonization of online activities (and not only). For this purpose, the format of legislation is regulation instead of directive, establishing the same level of rights

* Criminal Lawyer. Research Fellow in Legal Informatics at the Faculty of Law at the University of Milan and of the Information Society Law Center (ISLC).

and obligations across member states and enabling the coherent and effective application of rules within the Union.

2. The General Data Protection Regulation (GDPR)

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 ("GDPR") regulates the protection of the personal data of natural persons, setting uniform data privacy standards within the EU.

The right to privacy was first recognized in Europe by the 1950 European Convention for the protection of human rights and fundamental freedoms. With the advancement of technology and the invention of the Internet, the EU acknowledged the need for more effective and modern protection, enacting Directive 95/46/EC (European Data Protection Directive) – which established minimum data protection standards – whose transposition in each member state caused a significant fragmentation of the discipline across the EU. Hence, the need to ensure a coherent and harmonized level of protection within the EU finally led to the entry into force of the GDPR in 2016 and its application in May 2018 (with the repeal of the EU Data Protection Directive).

Representing a key act of the EU's secondary law, it signals the EU's resolute stance on data privacy and security at a time when the evolution of information technologies has changed the data processing environment, and an increasing number of individuals are relying on cloud services for the storage of their personal data, with data breaches occurring with alarming frequency. It is believed to be the toughest privacy and security law in the world (Krzysztofek, 2021).

The GDPR applies to organizations that are based in the EU, even if the data is being stored or used outside of it, on the basis of the extra-territorial effect. Its scope also covers the activities of organizations that are not based in the EU, if two conditions are met: such organizations offer goods or services to people in the EU, or they monitor their online behavior. For example, if an Indian company uses web tools that allow the tracking of cookies or IP addresses of people visiting the website from EU countries, then it should strive to be GDPR-compliant. It does not apply to purely personal or household activities. Unless they process sensitive data, also small and medium-sized enterprises (SMEs, with fewer than 250 employees) are partially exempt from the regulation, freeing them from record-keeping obligations in most cases (Art. 30.5).

The GDPR defines personal data as any information relating to an identified or directly/indirectly identifiable person, a so-called data subject (pseudonymous can also fall under the definition when it is easy to identify someone from it). Data processing consists of any operation or set of operations performed on personal data, whether automated or manual (such as collection, recording, organization, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or

combination, restriction, erasure or destruction). Other relevant definitions are those of: (i) data controller, the natural or legal person, public authority, agency, or other body which, alone or with others, determines the purposes and means of the processing of personal data; and (ii) data processor, the natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller (*e.g.*, cloud servers or email service providers).

The law is governed by the personal data processing principles, which constitute its framework and serve as guidelines for the interpretation of detailed legal provisions. The overriding nature of such principles implies that, in specific circumstances, they remain unalterable even with the consent of the data subject. For instance, in accordance with the data minimization principle, the maximum allowable extent of data which may be processed in a given case may not be expanded, despite the data subject's consent.

The fundamental key principles are seven and include the following:

1. **Lawfulness, fairness and transparency:** require the personal data to be processed in accordance with statutory laws and regulations issued on the basis of data protection laws (lawfulness), and that such processing be fair, respectful of the interests and reasonable expectations of data subjects, balanced between the data controller's commercial interests and the data subject's privacy (fairness), as well as transparent (transparency). The transparency principle includes the data subjects' right to have control over the processing of their data both through privacy policies (under Art. 13 and Art. 14, controllers have a specific obligation to provide data subjects with relevant information on the processing of their data) and through the exercise of the right of access (and other requests, such as the right of the data subject to obtain a copy of the personal data undergoing processing). Data subjects shall be informed on their data processing in a concise, transparent, intelligible and easily accessible form, using clear and plain language (Art. 12).
2. **Purpose Limitation:** requires data to be collected only for specific, explicit and legitimate purposes and must not be further processed in a way that is no longer compatible with the purposes for which it was collected (initial purposes). The data controller determines such purposes and must do so prior to starting the data processing.
3. **Data minimization:** together with the accuracy principle and the storage limitation principle, it is part of the so-called data quality principles. Under this principle, data processing shall be adequate, relevant and limited to what is necessary in relation to the declared purposes. If the data processing exceeds the scope necessary for the processing purposes, then such processing shall be prohibited.
4. **Accuracy:** requires the personal data being processed to be accurate and, where necessary, kept up to date. The controller shall ensure that inaccurate

data, having regard to the purposes for which it is processed, is erased or rectified without delay.

5. **Storage limitation:** under this principle, personal data shall be kept in a form which allows the identification of data subjects for no longer than what is necessary for its purposes. Intended to prevent data storage for an indefinite amount of time, only in certain cases may data be kept for longer periods (*e.g.*, archiving purposes in the public interest, scientific or historical research, statistical purposes). Following the achievement of the set purposes, personal data must be erased.
6. **Integrity and confidentiality:** require data processing to ensure appropriate security of the personal data, including protection against unauthorized or unlawful processing, and against accidental loss, destruction or damage through technical or organizational measures (*e.g.*, encryption or two-factor authentication).
7. **Accountability:** under this principle, the controller shall be responsible for and be able to demonstrate compliance with all fundamental key principles. The GDPR creates a chain of accountability that leaves very little leeway for adjustment. Observance of the principle entails appropriate procedures being introduced and accurate documentation being kept, indicating the implemented data processing method and related guarantees. The measures to be adopted by controllers depend on a prior risk-based assessment of the data processing's impact on data subjects' rights and aim specifically at avoiding the assessed risks (risk-based approach). Such measures include pseudonymization and encryption, the ability to ensure the ongoing confidentiality, integrity, availability and resilience of processing systems, the capacity to promptly restore the availability and access to personal data in case of a physical or technical incident, and processes for regularly testing the security's effectiveness. As part of the accountability principle, data protection shall be applied by design and by default (Art. 25). While the privacy by design approach demands the implementation of privacy safeguards even at the stage of creating, designing and developing services, systems or applications, the privacy by default requires a systemic approach to constructing services, systems or applications, ensuring that data protection is built into the processing by default.

Data processing shall be deemed lawful if the data subject has freely given specific, informed and unambiguous consent to process the data for one or more purposes (previously given consent may be withdrawn at any time, while children under the age of 13 must give consent with parental permission).

It shall also be deemed lawful if necessary to: (i) execute, prepare or enter into a contract to which the data subject is a party; (ii) comply with a legal obligation to which the controller is subject; (iii) protect the vital interests of data subjects; (iv) perform a task in the public interest or to carry out some official function; and (v) pursue the purpose of legitimate interest of the controller.

A personal data breach implicates a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to personal data transmitted, stored or otherwise processed. The controller must notify any personal data breach to the independent national supervisory authority without undue delay, within 72 hours of becoming aware of it (art. 33). The processor must, similarly, notify any such breach to the controller without undue delay. In case the breach triggers relevant risks for data subjects' rights, they must be informed by the controller.

The GDPR formulates a set of rights of data subjects towards the data controller, enabling them to exercise the principles of data processing. Data subjects' rights include information rights (right to be informed, right of access, right to data portability), correction rights (right to data rectification), restrictive rights (right to erasure, also known as the right to be forgotten, right to restrict processing and right to object).

The GDPR also introduces a relevant novelty in disciplining the new formal role of the data protection officer (DPO), responsible for helping the organization maintain data protection compliance. The DPO has the duty to inform, advise and monitor compliance with the GDPR but is not responsible for the entity's compliance, which rests upon the management. The officer shall act as an intermediary between the interested parties (*e.g.*, the personal data protection authority and data subjects) and the data controller (Kuner C. *et al.*, 2020). The major tasks assigned to the DPO are: to receive questions from data subjects related to their personal data processing; inform the organization and its employees of their obligations under the GDPR and any other applicable EU member state data protection provisions; monitor the entity's compliance with the GDPR (and any other applicable EU member state data protection provisions), train staff on compliance and perform audits; perform data protection impact assessments; cooperate with the data protection supervisory authority; act as the focal point for the data protection supervisory authority on matters relating to the processing of personal data.

Entities shall designate a DPO when: (i) they are a public authority or another public body; (ii) their core activities consist of processing operations requiring regular and systematic monitoring of data subjects on a large scale (large scale, regular data monitoring); and (iii) their core activities consist of processing on a large scale of special categories of data or personal data relating to criminal convictions and offenses (large scale, special data monitoring).

Penalties for violating the GDPR are severe and reach the maximum of €20 million or 4% of global annual revenue (whichever is higher, according to Art. 83). They are applied by the data protection authority in each EU country, which will determine both if the infringement has occurred and the severity of the fine. In addition, data subjects have the right to seek compensation for damages for GDPR infringements.

3. The Digital Services Act package

As part of the digital single market strategy embraced by the EU Commission and as a response to the increasing impact of digital services in people's everyday lives, in 2022 the EU adopted the Digital Services Act package, consisting of the Digital Market Act (DMA) and the Digital Services Act (DSA).

It sets out a comprehensive rulebook that uniformly applies throughout the EU and aims at creating a safer digital space where the fundamental rights of users are protected, and establishing a level playing field to foster innovation, growth, and competitiveness in the European digital ecosystem.

Two years before its implementation, through the "Shaping Europe's Digital Future" communication, the EU Commission stated:

> For Europe to truly influence the way in which digital solutions are developed and used on a global scale, it needs to be a strong, independent, and purposeful digital player in its own right.

To achieve this, a clear framework that promotes trustworthy, digitally enabled interactions across society, for people as well as for businesses, is needed" and pointed out the three key objectives for European future and growth, which certainly inspired the DSA package: (i) technology that works for people; (ii) a fair and competitive economy; and (iii) an open, democratic and sustainable society.

3.1. The Digital Market Act (DMA)

Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022, the DMA, regulates contestable and fair markets in the digital sector across Europe, where gatekeepers are present, to the benefit of business users and end users. It aims to ensure a level playing field in the digital environment and sets out rights and rules for large online platforms (gatekeepers), making sure that none of them abuse their prominent role in today's digital economy.

The DMA applies to core platform services provided or offered by gatekeepers to business users established in the Union or end users established or located in the Union. Core platform services are the most widely used services by business users and end users (under Art. 3.2 *b)* of the DMA, a designated core platform service must have more than 45 million monthly active end users established or located in the EU).

These services are only those where there is strong evidence of: (i) high concentration, where usually one or very few large online platforms set the commercial conditions with considerable autonomy from their potential challengers, customers or consumers; (ii) dependence on a few large online platforms acting as gateways for business users to reach and have interactions with their customers; and (iii) the power of core platform service providers often being misused by means of unfair behavior *vis-à-vis* economically dependent business users and customers.

They include online intermediation services (e.g. Google Maps, Amazon Marketplace, App Store), online search engines (*e.g.*, Google Search), online social networking services (*e.g.,* TikTok, Facebook, Instagram, LinkedIn), video-sharing platform services (*e.g.*, YouTube), number-independent interpersonal communications services (*e.g.*, WhatsApp, Messenger), operating systems (*e.g.*, Android, iOS), web browsers (*e.g.*, Google Chrome, Safari), virtual assistants (as of today, Siri, Alexa and Google Assistant seem not to satisfy the DMA quantitative thresholds), cloud computing services, online advertising services (including any advertising networks, advertising exchanges and any other advertising intermediation services, provided by an undertaking that provides any of the core platform services listed before).

Gatekeeper status is determined either on the basis of quantitative metrics or on a case-by-case qualitative assessment by means of a market investigation.

In particular, a gatekeeper:

- has a significant impact on the internal market (achieves an annual Union turnover equal to or above euro 7,5 billion in each of the last three financial years, or its average market capitalization or its equivalent fair market value amounted to at least euro 75 billion in the last financial year, and provides the same core platform service in at least three member states);
- provides a core platform service which is an important gateway for business users to reach end users (provides a core platform service that in the last financial year has at least 45 million monthly active end users established or located in the Union and at least 10 thousand yearly active business users established in the Union); and
- enjoys an entrenched and durable position, in its operations, or it is foreseeable that it will enjoy such a position in the near future (where the thresholds at the preceding paragraph are met in each of the last three financial years).

Companies that meet the above criteria are presumed to be gatekeepers and shall notify the EU Commission, which has the power to supervise the fulfillment of each step of the regulation (centralized enforcement model), with the assistance of the Digital Markets Advisory Committee composed of representatives of member states. Gatekeepers have the right to rebut the presumption and submit substantiated arguments to demonstrate why they should not be designated as such.

Following their designation, gatekeepers have six months to comply with the DMA and provide a compliance report detailing all implemented solutions. Non-compliance with the DMA entails severe fines, up to 10% of the gatekeeper's total worldwide turnover that can go up to 20% for repeated infringement. In cases of systematic infringements, the Commission is also empowered to adopt additional remedies, such as obliging a gatekeeper to sell a business or parts of it or banning the gatekeeper from acquiring additional services related to the systemic non-compliance.

The DMA lays down a set of obligations (practices of gatekeepers that limit contestability or are unfair), which include:

- data processing and targeted advertising: gatekeepers must obtain user consent before collecting and processing personal data for targeted advertising purposes; they are prohibited from combining personal data from different core platform services or cross-using personal data in other services provided by the gatekeeper;
- nondiscrimination: gatekeepers must not discriminate against business users by preventing them from offering the same products or services through third-party online intermediation services or their own direct online sales channels at different prices or with different conditions;
- freedom of communication and promotion: gatekeepers must allow business users to communicate and promote their offers to end users acquired through the gatekeeper's core platform services, regardless of whether they use the gatekeeper's services; this promotes fair competition and ensures that business users have the freedom to reach their target audience;
- access to content and services: gatekeepers must allow end users to access and use content, subscriptions, features, or other items through their core platform services by using the software application of a third-party business user; this promotes user choice and prevents gatekeepers from limiting access to certain services;
- complaints and dispute resolution: gatekeepers must not prevent or restrict business users or end users from raising issues of noncompliance with relevant laws to public authorities, including national courts; gatekeepers are encouraged to establish lawful complaint-handling mechanisms to address any concerns raised by users.

On the 6th of September 2023, the EU Commission issued a press release (the non-confidential version of the decisions will be made available once confidentiality issues have been solved) detailing the designation of six gatekeepers: Alphabet, Amazon, Apple, ByteDance, Meta, and Microsoft, which have until the 6th of March 2024 to comply with the DMA. For the moment, Samsung has not fallen under the scope of this designation. In parallel, the Commission has opened four investigations to further assess Microsoft's and Apple's rebuttal that, despite satisfying the thresholds, Bing, Edge and Microsoft Advertising (for Microsoft) and iMessage (for Apple) do not qualify as gateways.

3.2. The Digital Services Act (DSA)

Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022, the DSA, regulates a single market for digital services and amends Directive 2000/31/EC, also known as the e-commerce Directive, the application of which shall not be affected. On the assumption that what is illegal offline should also be considered illegal online, the DSA establishes:

- a framework for the conditional exemption from liability of providers of intermediary services;
- rules on specific due diligence obligations tailored to certain specific categories of providers of intermediary services;
- rules on implementation and enforcement, including those concerning the cooperation of and coordination between the competent authorities.

It applies to intermediary services, consisting of mere conduit, caching and hosting services, provided to recipients of the service that have their place of establishment or are located in the Union, irrespective of where the providers of those intermediary services have their place of establishment.

Given the exponential growth and spread of such services, the DSA aims at limiting the proliferation of online illegal content and at ensuring the safety of users. To that purpose, it includes a set of conditions under which intermediary service providers (ISPs) meet exemption from liability for the third-party information they transmit and store without affecting the possibility for a court or administrative authority, in accordance with member states' legal systems, of requiring the service provider to terminate or prevent an infringement.

For example, a hosting service provider shall not be held liable if: (i) it does not have actual knowledge of illegal activity or content and, as regards claims for damages, is not aware of facts or circumstances from which the illegal activity or illegal content is apparent; or (ii) upon obtaining such knowledge or awareness, acts expeditiously to remove or disable access to the illegal content.

Upon receipt of an order to act against one or more specific items of illegal content issued by the relevant national judicial or administrative authorities, providers must inform such authorities of any effect given to the order without undue delay, specifying if and when the order was applied.

Within their due diligence obligations, all ISPs are required to:

- designate single points of contact for electronic communications with member state authorities, the Commission, the European Board for Digital Services and with users (providers without an EU establishment must appoint an EU-based legal representative to receive, comply with, and enforce DSA-related decisions);
- explain in their terms and conditions, subject to periodical revision, how they restrict and moderate (*e.g.*, remove) content that is illegal or incompatible with their T&Cs (*e.g.*, by using algorithms). Any relevant modification to the T&C must be notified to users;
- outline their content moderation practices in annual transparency reports (except for small ISPs), easily accessible, in machine-readable format and written in a plain, intelligible and unambiguous language.

The hosting ISP must, additionally, implement user-friendly electronic mechanisms (*e.g.*, online forms) that enable users and third parties to report illegal content by submitting sufficiently precise and substantiated notices. Decisions

on notified content (*e.g.*, content removal or demotion) must be taken in a timely, diligent, non-arbitrary and objective manner. The hosting ISP must also provide the users affected by content moderation with a statement of reason, giving clear and specific explanations for their decision (where automated means are used for decision-making, this must be disclosed to the user).

The DSA provides for additional obligations also for online platforms, defined as hosting ISPs that store third-party content and, at their recipients' request, disseminate that content to the public (*e.g.*, social networks).

They are required to:

- implement an internal complaint handling system, enabling users to file complaints against their content moderation decisions;
- inform complainants of their reasoned decision along with available redress options, including mandatory out-of-court dispute settlements;
- decide with priority upon notices submitted by "trusted flaggers", appointed by member states' Digital Services Coordinator;
- act against misuse by suspending their service for users who frequently provide manifestly illegal content or by suspending the processing of complaints by users who frequently submit unfounded notices;
- forbid targeted advertisements based on profiling using sensitive data of users;
- label all advertisements and inform users on who is promoting and financing them;
- refrain from designing their platform in a way that seeks to shape user behavior in a particular way (prohibition of dark patterns) or in any other deceitful or manipulative manner that would impair users' ability to make free and informed decisions;
- notify suspicions of serious criminal offenses to the competent authorities.

Providers of online platforms accessible to minors must implement appropriate and proportionate measures to maintain a high level of security, privacy, and safety for minors. For instance, they are prohibited from presenting targeted advertisements based on profiling using the personal data of minors.

To fight the online sale of illegal products and services, providers of B2C online marketplaces are subject to special obligations, such as designing their online interfaces in a way that allows traders to comply with their obligations (pre-contractual compliance and product safety information) and ensuring mechanisms of random checks for illegal products and services.

The DSA introduces a significant novelty for Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs), on the assumption that larger intermediary services with significant societal impact shall be subject to stricter rules. VLOPs and VLOSEs are defined as platforms (VLOPs) and engines (VLOSEs) which have at least 45 million average monthly active users in the EU and have been designated as such by the Digital Services Coordinator. They face additional compliance obligations and must:

- conduct risk assessments on the systemic risks related to the use or misuse of their service and implement risk mitigation measures;
- implement a crisis response mechanism (for extraordinary circumstances affecting public security and public health);
- commission annual independent compliance audits;
- implement at least one non-profiling-based recommender system option;
- observe heightened advertising transparency and file additional transparency reports;
- ensure data access and scrutiny by supervisory authorities;
- appoint one or more compliance officers (DSA officers);
- pay annual supervisory fees.

In cases of non-compliance, the Commission can impose fines up to 6% of global annual turnover.

On the 25th of September 2023, the EU Commission designated 17 VLOPs, namely Alibaba AliExpress, Amazon Store, Apple AppStore, Booking.com, Facebook, Google Play, Google Maps, Google Shopping, Instagram, LinkedIn, Pinterest, Snapchat, TikTok, Twitter, Wikipedia, YouTube, Zalando, and 2 VLOSEs, Bing and Google Search.

4. The Artificial Intelligence Act (AIA)

The use of artificial intelligence (AI) in the EU is governed by Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024, known as the AIA, which is considered the first comprehensive legal framework on AI worldwide. The EU Commission originally published a proposal in April 2021, upon which the EU Parliament adopted a negotiating position in mid-June 2023 (499 votes in favour, 28 against and 93 abstentions). Following a protracted trialogue (a three-way negotiation process), at the beginning of December 2023 the Parliament and the Council reached a provisional agreement. On such occasion, President Ursula Von Der Leyen stated:

> Artificial intelligence is already changing our everyday lives. And this is just the beginning. Used wisely and widely, AI promises huge benefits to our economy and society. Therefore, I very much welcome today's political agreement by the European Parliament and the Council on the Artificial Intelligence Act. The EU's AI Act is the first-ever comprehensive legal framework on Artificial Intelligence worldwide. So, this is a historic moment. The AI Act transposes European values to a new era. By focusing regulation on identifiable risks, today's agreement will foster responsible innovation in Europe. By guaranteeing the safety and fundamental rights of people and businesses, it will support the development, deployment, and take-up of trustworthy AI in the EU. Our AI Act will make a substantial contribution to the development of global rules and principles for human-centric AI.

The proposal became law on 12 July 2024, with the publication of Regulation (EU) 2024/1689 in the Official Journal and has been in force since 1 August 2024. The AIA establishes rigorous, harmonized rules for the development, placement on the market, putting into service and use of AI within the EU. These rules ensure that AI systems operating in the EU are secure and uphold fundamental human rights, in accordance with Union values, while also providing the flexibility needed to accommodate investment and support innovation. Recital 6 states that, given the significant impact that AI can have on society, it should be a human-centric and trustworthy technology, serving as a tool for people with the ultimate aim of enhancing human well-being (anthropocentric approach). The AIA also sets out prohibitions of certain AI practices, specific requirements for high-risk AI systems and obligations for operators of such systems. It introduces harmonized transparency rules for certain AI systems, rules for the placement on the market of general-purpose AI models, provisions on market monitoring, surveillance, governance and enforcement, and measures to support innovation, with a particular focus on small and medium-sized enterprises (SMEs), including start-ups.

The AIA grants a technology-neutral, uniform and future-proof definition of AI, focusing on its outputs and objectives rather than its underlying technology or algorithms. An AI system is defined as "a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the output it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments". The Regulation clarifies that a key characteristic of AI systems is their capability to infer – a process of generating outputs (such as predictions, content, recommendations, or decisions) which can influence physical or virtual environments. This also includes the capability of AI systems to derive models or algorithms, or both, from inputs or data through machine learning, logic-based and knowledge-based approaches. It also specifies that AI systems are designed to operate with varying levels of autonomy, meaning they possess some degree of independence in their actions and the ability to operate without human intervention.

It applies to various actors: (i) providers (or their authorised representatives) who place AI systems or general-purpose AI models on the market or put them into service in the Union, regardless of whether they are established or located within the EU or in a third country; (ii) deployers of AI systems (*i.e.* users, except when used for personal non-professional activities) that have their place of establishment or are located within the EU; (iii) providers and deployers of AI systems that have their place of establishment or are located in a third country, where the output produced by the AI system is used in the EU; (iv) importers and distributors of AI systems; (v) product manufacturers placing on the market or putting into service an AI system together with their product and under their own name or trademark; (vi) affected persons located in the EU.

For regulating AI, as had been previously done with the GDPR, the EU opted for a defined risk-based approach. This approach is intended to avoid unnecessary restrictions on trade and to ensure that legal intervention is tailored to cases of justified cause for concern or where such concerns may be reasonably anticipated in the near future. The main principle is that the higher the risk, the stricter the rules.

The AIA identifies four levels of risk, based on the intended use of AI systems:

1. **Unacceptable** – AI systems that are strictly prohibited include: (a) systems deploying subliminal, manipulative or deceptive techniques to distort behaviour and impair informed decision-making (e.g., cognitive behavioural manipulation); (b) systems exploiting vulnerabilities related to age, disability, or socio-economic status; (c) social scoring systems that classify individuals based on social behaviour or personal traits; (d) systems assessing or predicting the risk of a person committing a criminal offence, solely based on profiling or personality traits (except when used to support human assessment of involvement in a criminal activity, which is already based on objective and verifiable facts); e) systems that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or closed-circuit cameras footage; f) systems inferring emotions in the workplace and educational institutions (except for medical or safety reasons); g) biometric systems categorizing people based on race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation (except for any lawful labelling or filtering of biometric datasets for law enforcement purposes); (h) real-time remote biometric identification systems (e.g., facial recognition) in publicly accessible spaces for law enforcement (except when strictly necessary for the targeted search for victims of abduction, trafficking or sexual exploitation, or for searching for missing persons, preventing imminent threat to life or foreseeable terrorist attack, or detecting suspects in serious criminal offenses).
2. **High** – these systems may cause adverse effects on fundamental rights and are subject to a detailed certification regime before being placed on the market and throughout their lifecycle, requiring them to meet stringent requirements in addition to passing a conformity assessment. An AI system is considered high-risk if it is used as a safety component of a product, or if it is a product itself that is covered by EU legislation. Certain AI systems are always considered high-risk, unless they do not pose a significant risk to people's health, safety or rights. These include (i) systems that fall into eight specific areas (listed in Annex III) that will have to be registered in an EU database: non-banned biometric identification and categorization systems, and non-banned systems intended to be used for emotion recognition; management and operation of critical infrastructure (such as water supply or gas); education and vocational training; employment, workers management and access to self-employment; access to and enjoyment of essential private and public services (such as evaluating creditworthiness);

law enforcement; migration, asylum and border control management; administration of justice and democratic processes; (ii) systems that are used in products categorized under the EU's product safety legislation, such as toys, aviation, cars, medical devices. The obligations of high-risk AI providers encompass the implementation of risk management systems, the conduct of data governance and record-keeping of technical documentation and the development of systems with the capabilities of automatic recording, allowing deployers to provide human oversight. Crucially, the EU emphasizes that AI systems should be overseen by people, rather than by automation, to prevent harmful outcomes.

3. **Limited** – these AI systems are required to comply with light transparency obligations that enable users to make informed decisions. They include systems that generate audio or video content and chatbots. Individuals should be made aware that they are interacting with AI and given the option to either continue or opt out, unless it is obvious, or the AI is used for legal purposes (such as crime detection). AI-generated content shall be made identifiable by deployers and when such content (AI-generated text, audio or video) is published to inform the public on matters of public interest, it must be clearly labelled as artificially generated, including content constituting deep fakes. Deployers must also inform users when AI is used for emotion recognition or biometric categorisation, except when used for legal purposes. The AI Office will encourage and facilitate the development of guidelines for detecting and labelling artificially generated content.
4. **Minimal** – these systems are unregulated as they present minimal or no risk to individuals' safety and rights. They include applications such as spam filters or AI-enabled video games and represent the vast majority of AI systems currently used in the EU (although this may change with the spread of generative AI). Providers of minimal-risk systems may voluntarily commit to codes of conduct.

Additionally, the AIA addresses general-purpose AI models. These models are trained with large amounts of data using self-supervision at scale and are designed to perform a wide range of tasks and functions, similar to the versatility of human intelligence. Since they may be used as high-risk AI systems or integrated into them, providers of general-purpose AI models must comply with specific obligations, including: (i) drafting technical documentation (covering training, testing processing and evaluating results) as well as documentation to supply to downstream providers that intend to integrate the model into their AI system; (ii) establishing specific policies to respect copyright; and (iii) publishing a sufficiently detailed summary of the content used for training the model. Providers of general-purpose AI models with systemic risk – those having a significant impact on the EU market due to their reach or potential negative effects on fundamental rights or society, which can propagate across the value chain – must evaluate their models using standard

protocols and report any serious incidents to the AI Office and national authorities. They must also ensure their AI models and infrastructure are secure.

Non-compliance with the AIA attracts significant penalties ranging from 35 million euros or 7% of global turnover to 7,5 million or 1,5% of turnover, depending on the severity of the infringement and size of the company.

The AIA not only aligns perfectly with EU values, but also sets a global standard for responsible AI development. As AI continues to evolve, the AIA will serve as a critical tool for safeguarding public interest, encouraging sustainable innovation, and will significantly strengthen the EU Digital Strategy, while setting a robust precedent for AI governance worldwide.

5. Other regulations

With the aim of creating a Europe fit for the digital age, the EU is undertaking an ambitious transition towards climate neutrality and digital leadership, pursuing a human-centric, sustainable vision for digital society throughout the digital decade.

Among the many recent legislative initiatives by the EU, it is worth mentioning the Data Governance Act (Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022, so-called DGA) on European data governance, and the Data Act (Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023, so-called DA) on harmonized rules on fair access to and use of data (see Chapter IV); as well as the Digital Operational Resilience Act (Regulation (EU) 2022/2554 of the European Parliament and of the Council of 14 December 2022, so-called DORA) on digital operational resilience for the financial sector, the Chips Act (Regulation (EU) 2023/1781 of the European Parliament and of the Council of 13 September 2023) establishing a framework of measures for strengthening Europe's semiconductor ecosystem, and the EDIRPA (Regulation (EU) 2023/2418 of the European Parliament and of the Council of 18 October 2023) establishing an instrument for the reinforcement of the European defense industry through common procurement.

Bibliography

Allen, R. and Masters, D. (2020) *Regulating for an equal AI: a new role for equality bodies – meeting the new challenges to equality and non-discrimination from increased digitalization and the use of artificial intelligence*. Brussels: Equinet.

Bernitz, U., et. al. (2020) *General Principles of EU Law and the EU Digital Order*. Alphen aan den Rijn: Wolters Kluwer.

Broussard, M. (2019) *Artificial unintelligence. How computers misunderstand the world*. Cambridge: MIT Press.

Cate, F. H. and Dockery, R. (2018) *Data Privacy and Security Law*. Oxford: Oxford University Press.

Chamberlain, J. (2023). 'The risk-based approach of the European Union's proposed artificial intelligence regulation: some comments from a tort law perspective', *European Journal of Risk Regulation*, 14(1), pp. 1–13. Doi: 10.1017/err.2022.38.Cole, M. D., Etteldorf, C., and Ullrich, C. (2021) *Updating the rules for online content dissemination: legislative options of the European Union and the Digital Services Act proposal*. Baden-Baden: Nomos.

Edwards, L. (2022) *The EU AI Act: A Summary of its Significance and Scope*. London: Ada Lovelace Institute.

Floridi, L. (2021) *Ethics, governance, and policies in artificial intelligence*. Berlin: Springer.

Hacker P., Engel A. and Mauer M. (2023) 'Regulating ChatGPT and other Large Generative AI Models'. Available at: https://doi.org/10.48550/arXiv.2302.02337.

Krzysztofek, M. (2019) *GDPR: General Data Protection Regulation (EU) 2016/679: post-reform personal data protection in the European Union*. Alphen aan den Rijn: Wolters Kluwer.

Kuner, C., et. al. (2020) *The EU General Data Protection Regulation (GDPR): A Commentary*. Oxford: Oxford University Press.

Laux, J., Wachter, S., Mittelstadt, B. (2021) 'Taming the few. Platform regulation, independent audits, and the risks of capture created by the DMA and DSA', *Computer Law & Security Review*, 43. Available at: https://doi.org/10.1016/j.clsr.2021.105613

Leistner, M. (2021) 'The Commission's vision for Europe's digital future: proposals for the Data Governance Act, the Digital Markets Act and the Digital Services Act – a critical primer', J*ournal of intellectual property law & practice*, 16(8). Available at: https://doi.org/10.1093/jiplp/jpab054.

Lynch, S. (2023) *Analysing the European Union AI Act: What Works, What Needs Improvement* [Online], Human-Centred Artificial Intelligence (HAI) Stanford University. Available at: https://hai.stanford.edu/news/analyzing-european-union-ai-act-what-works-what-needs-improvement.

Lynskey, O. (2019) 'Grappling with "Data Power": Normative Nudges from Data Protection and Privacy', *Theoretical Inquiries and Law*, 20(189) Available at: https://www7.tau.ac.il/ojs/index.php/til/article/view/1613/1714.

Sharma, S. (2019) *Data Privacy and GDPR Handbook*. Newark: Wiley.

Solove, D. J., Schwartz, P. M. (2020) *EU Data Protection and the GDPR*. Boston: Aspen Opco Llc.

Vrabec, H. U. (2021) *Data subject rights under the GDPR: with a commentary through the lens of the data-driven economy*. Oxford: Oxford University Press.

Chapter IV
The Regulation of Data in the European Union: the Data Governance Act and the Data Act

by Charlotte Ducuing*

Index: 1. From data ownership to EU data legislation. – 2. The DGA: Institutions to facilitate data sharing. – 3. The Data Act: the allocation of rights on or in relation to data. – 4. Commonalities and elements of consistency within EU data legislation.

1. From data ownership to EU data legislation

Data are increasingly used as valuable resources in the data economy, which is illustrated by the now commonplace references in the media to data as a "new oil" or "even more!". Data lie at the heart of the business model of online platforms and this model – often referred to as Big Tech – is now expanding to an increasing number of sectors.

Data become central to the operation of connected devices – also referred to as the Internet of Things (IoT) – found with both individuals and businesses. Think for example of smart machinery in precision medicine or farming, or of smart wearables such as smart watches or smart fridges which generate data through their use and run on data.

The IoT is also closely connected to Artificial Intelligence (AI), visible for example with smart cars moving towards higher degrees of autonomy based on sensors that generate all sorts of data and on increasingly elaborate modes of computation.

Data are often expected to serve a whole range of purposes – both promising and challenging ones. For example, the smartization of cities refers to the growing use of data to optimize city management in various ways, such as through surveillance of the public domain or data-driven optimization of traffic.

In this context, the question has arisen how the law should take into account the function of data as a resource begging, in turn, the question what type of resource data are from an economic perspective.

* Doctoral Researcher at the Centre for IT and IP Law (CiTiP) of the KU Leuven, Belgium.

Dissensions do exist as to which kind of resource data are and whether such qualification shall constitute the starting point for regulating them (Purtova and van Maanen, 2023). This being, data are often characterized by their non-rivalry in use, entailing that consumption by one does not detrimentally affect their quality when used by another actor.

This shall be associated with the other feature of data as non-depletable resources. Data considered to have infrastructural features (OECD, 2016) or potential. In addition to their above-mentioned features, they can indeed be put to use for different purposes, especially in the Big Data era where the semantic content (information) loses significance to the benefit of the quantity and constant flow of data. Also, data do not have an inherent economic value. Their value depends on their (re)use, which explains why they are rarely exchanged as a commodity as such. Rather, their economic valorization is often indirect, through the commercialization of data-driven knowledge, products or services (Martens, 2021).

Importantly, several dimensions or interests are nested within data (Bietti, 2020). While the recognition of data as a resource relates to economic interests, data also embed non-economic interests and especially fundamental rights such as privacy and data protection. This shall notably be associated with the specific power dimensions of data, in the sense that power can be manifested or exercised "through" data (Ruppert, Isin and Bigo, 2017). Such is the case with data-driven surveillance which, when deployed in the public domain, consists in the exercise of power onto citizens with a risk of interfering with, for example, their freedom of movement.

In recent years, the European Union (EU) has attempted to design its own political project concerning the role of data in society. Initially, the European Commission (Commission) was leaning towards a form of data ownership, as visible with the idea to create a data producer's right for non-personal data (Commission, Communication of 2017 "Building a European Data Economy"). With ownership rights on data, the Commission ambitioned to turn data into a tradable commodity, thus allowing for data markets to flourish. The idea was to include only non-personal data, with data protection – and thus personal data - deemed incompatible with data ownership. This option was abandoned, including because the legal literature demonstrated the impossibility to fully distinguish "personal" from "non-personal" data (Drexl, 2018; Graef, Gellert and Husovec, 2018).

The EU has now sided for a more complex and *sui generis* approach to data regulation, following the Communication from the Commission "European Data Strategy" of 2020. The European Data Strategy sticks to the same goal to establish European data markets, with the steady ambition to steer economic growth and data use for a wide range of purposes while preserving European values and especially competition and data protection. To do that, the Strategy

envisions a horizontal layer of data regulation, further complemented by sector- or domain-specific regulations where deemed necessary. The horizontal layer comprises two main regulations. Adopted in 2022, the Data Governance Act (DGA) aims to facilitate and support voluntary data sharing in different scenarios. Adopted in 2023, the Data Act lays down rights on data – and especially rights to access and use data – in various scenarios. This Chapter discusses these two Regulations – referred to, together, as EU data legislation – in further details. Besides, sector- or domain-specific regulations make part of a broader attempt from the Commission to structure data markets within "Common European data spaces", defined loosely as the legal and technical means to enable data sharing in scenarios deemed of general interest (see Data Act, Art. 33(1), Commission SWD, 2022). Common European data spaces may include data-related legislations in the future, such as the Health Data Space Regulation proposed by the Commission in 2022 to steer the digitalization of health data records and to facilitate their (re)use.

Several authors have analyzed the European Data Strategy as the initiation of an "EU data law", in the sense of a new branch of law characterized by data as a regulatory subject-matter, by a set of specific legal principles and a common conceptual framing (Streinz, 2021; Riis, 2023). An EU data law undoubtedly reflects the ambition of the Commission to achieve a comprehensive regulation of data, with now a new economic law layer being added to the already existing fundamental rights one – namely the GDPR. Yet, at first glance, EU data legislation appears to be piecemeal and disparate. While EU data legislation does obviously focus on data, a remaining question is whether EU data legislation has some sufficient level of (conceptual) consistency and what are, if at all, its specific legal principles.

In that light, this chapter provides an outline of EU data legislation – the horizontal layer of EU regulation of data – in order to identify, beyond the diversity of the provisions, commonalities and/or elements of consistency. This constitutes a prerequisite exercise to analyze the possible emergence and features of an EU data law, based on dedicated legal principles and a common conceptual framing. To do so, this chapter analyzes the policy vision of the EU concerning data and its manifestations with the DGA and the Data Act. It provides a descriptive and explanatory perspective on the main substantive provisions. It does not provide a comprehensive overview of the specific enforcement mechanisms that EU data legislation requires member States to establish for the full spectrum of its substantive obligations. The complexity of these provisions, the importance of national implementation and the thorny question of how they relate to already-existing enforcement authorities, beg for a developed analysis which lies beyond the ambit of this chapter. This chapter does not evaluate EU data legislation, in particular against its own ambition to regulate data comprehensively, with both economic and fundamental rights

interests being kept in check. The chapter proceeds in three movements, in addition to the present introductory section 1. Sections 2 and 3 respectively outline, in turn, the main provisions of the DGA and of the Data Act. Both Regulations include several legal regimes, which are introduced in turn. Based thereon, section 4 identifies the commonalities and elements of consistency of EU data legislation, that suggest the ambition – if not squarely the emergence – of an EU data law.

2. The DGA: institutions to facilitate data sharing

The DGA consists in three main chapters, respectively dealing with three different types of data sharing scenarios, which are hereby outlined in turn. Chapter II DGA entitled "re-use of certain categories of protected data held by public sector bodies" aims to facilitate the making available of data by public sector bodies (PSBs) beyond the obligations incumbent on them under the Open Data Directive (section 2.1). Chapter III DGA entitled "data intermediation services" aims to incentivize the commercial sharing of data, through the recourse to trustworthy data intermediaries (section 2.3). Chapter IV DGA entitled "data altruism" aims to incentivize the altruistic sharing of data for purposes of general interest through the recourse to trustworthy data altruism organizations (section 2.4). Before outlining these Chapters, it is, first, necessary, to clarify what the EU means with and expects from data governance under the DGA (section 2.1).

2.1. The policy vision: towards an EU way of data governance

The ambition of the EU is to foster data sharing, thus allowing for data to be (re)used more broadly. The fostering of data sharing is expected to unleash the potential of data against the background of their economic qualification as non-rival resources. The DGA pursues the general objective to improve the conditions for data sharing in the internal market (DGA, Rec. 3). With the DGA, the EU focuses explicitly on "data governance" *albeit* without a definition. The notion of "data governance" appears also in policy documents from the Commission, such as the European Data Strategy and the Commission SWD on Common European data spaces, also without a definition. This being, it appears from these documents that the EU refers to data governance as institutions or structures that enable and/or determine the conditions for data sharing, in close connection to the operation of data spaces. That way, data governance within the meaning of EU data policies makes part of the infrastructure for data sharing (Commission SWD, 2022).

Thus understood, data governance is critical to the ambition of the EU to increase trust in data sharing. The DGA is based on the premise that actors

– both individuals and businesses – do not engage into data sharing as much as they could or would like to, because of their fear of losing control over data, thus undermining their trust and eagerness. What the DGA refers to as data control or agency is not defined, but appears to refer to the ability of actors – especially those providing data in order to share them – to have their individual preferences concerning the conditions and purposes of data sharing and further use respected by data users (or acquirers). In order to deliver on this objective, the DGA sets out to regulate the institutions or structures supporting different types of data sharing scenarios, to whom data providers and data users (or acquirers) resort for the purpose of supporting their data sharing endeavors.

In doing so, the DGA portrays itself as a counterpoint to Big Tech. Big Tech are perceived by the EU legislature as vertically integrated data monopolists who unfairly appropriate data of weaker actors. In contrast, the European way of data governance as promoted by the DGA is characterized by a vertical separation between the various phases in data lifecycle - data collection or provision, data sharing, and data (re)use (DGA, Rec. 32). Such a decentralized approach is deemed to enable weaker actors such as individuals and SMEs to take part in the data economy. In other words, with the DGA, the EU aims to establish fair and undistorted data markets. That way, data markets are both an end and a means to empower weaker actors.

The DGA also follows geo-political and industrial goals, described as a form of neo-mercantilism (König, 2022). While Big Tech have already (mis-)appropriated individuals' data via online platforms, the same phenomenon could also happen with brick-and-mortar industries, however key an economic asset of the EU. A fair and undistorted data market is thus viewed as an alternative model for the data economy within the EU. This is complemented with an external dimension, in the form of a protection against mis-use of data outside of the EU that data governance institutions and structures should contribute to as detailed in the paragraphs that follow.

2.2. The making available of data by public sector bodies beyond the Open Data Directive

PSBs, such as municipalities or administrative agencies, are at the forefront of EU data policies. Regulated mainly at national level, the right for citizens to access documents held by PSBs – such as opinions delivered by expert bodies ahead of policy decisions – is traditionally viewed as a means, in a democratic society, to hold PSBs accountable towards the general public for the decisions they make. The Public Sector Information Directive of 2003 brought about a novel approach to documents held by PSBs, actually foreshadowing the principle of future EU data legislation. This Directive was based on the realization

that these documents also have an untapped economic value and innovation potential, which could be unleashed by putting them in the hands of private economic actors. For example, start-ups could use various types of geographical data produced by municipalities to run and optimize GPSs and intermodal journey planners. The Directive has been revised on several occasions with the steady objective to facilitate the reuse of documents held by PSBs by a broad range of (private) actors and for a broad range of purposes, virtually by anyone and for any purpose. The notion of document has become gradually outmoded by this of data, as illustrated by the recast Open Data Directive of 2019.

Under the Open Data Directive, PSBs are under a by-default obligation to make the documents (or data) that they hold available to third parties. This principle has exceptions, especially taken from the presence of rights of third parties in relation to the documents (or data). In particular, the Open Data Directive does not apply in case of intellectual property rights (IPRs) nor does it apply when national law excludes that access be granted to documents (or data) because of national security, statistical confidentiality or commercial confidentiality (such as trade secrets) (Open Data Directive, Art. 1(2)). Besides, the Open Data Directive is without prejudice to data protection (Open Data Directive, Art. 1(4)).

In this context, Chapter II DGA lays down a legal regime for such data, held by PSBs, which lie outside of the scope of the Open Data Directive because of commercial confidentiality, statistical confidentiality, IPRs or data protection (DGA, Art. 3) (protected data).

Chapter II DGA is based on principles that differ from these of the Open Data Directive in the following respects.

1. While the Open Data Directive is based on the principle of an obligation for PSBs to make documents (or data) available, Chapter II DGA lays down an "optional" regime for PSBs. Should they want to make protected data available, then they should comply with the set of rules laid down by the DGA.
2. Used to play a limited role in making data available (*i.e.*, "as they are"), PSBs have increasingly been requested to play an "active" role in facilitating data reuse by reusers. This culminates with Chapter II DGA, which requires PSBs to take both legal and technical measures to that aim in view of the protected nature of data. As part of such measures, they may notably modify data to preserve this protected nature – for example through anonymization and/or aggregation, they may make data available in a secure processing environment that they control or they may require that data be reused within secured physical premises (DGA, Art. 5(3)). The Open Data Directive is based on the general principle that documents (or data) shall be made available to anyone for any purpose. In contrast, Chapter II DGA instantiates a granular and context-based approach to both who reuses data and for what purpose(s). The PSB shall take an active role in ensuring a balance between the data needs of the data reuser and the preservation of the

protected nature of data. On the one hand, PSBs are essentially requested to reserve (contractually) a right to audit data reusers (DGA, Art. 5(4), (5), (7) and (8)) to ensure that they preserve the protected nature of data. The audit may notably result in that the PSB prohibits the data reuser to use the results of certain data reuse. Besides, and in line with the neo-mercantilist approach of the DGA, the PSB shall take specific measures to preserve the protected nature of data when transferred outside of the EU (DGA, Art. 5(9) and (10)). On the other hand, where appropriate, the PSB shall assist data reusers in obtaining the consent or authorization ("permission") from actors benefiting from protection on data (such as data subjects or IPRs holders) - thus acting as a go-between (DGA, Art. 5(6)) - and in complying with their obligations (DGA, Art. 5(11)).

3. Under the Open Data Directive, and in line with their rather passive role, the making available of documents (or data) for reuse by PSBs was in principle free of charge with the exceptional possibility to charge the marginal costs of making them available (Open Data Directive, Art. 6(1)). In contrast and in line with the more active role that PSBs shall then play, Chapter II DGA is based on the opposite principle that PSBs may charge a fee. This principle has limitations; in particular the fee shall be derived from the costs incurred for making data available which are likely to be higher.

It follows that PSBs should act as data professionals which they may not be capable of, nor do they possibly dispose of the necessary resources. Anticipating the difficulty for PSBs to take on this new role, Chapter II DGA requires member States to designate competent bodies to assist PSBs (DGA, Art. 7(1)). Such competent bodies may even be requested to substitute PSBs in making protected data available. This could result in a centralized competent body established to that aim for data stemming from many or all PSBs in a given jurisdiction (DGA, Art. 7(2)), possibly as a centralized data public service or authority.

Chapter II DGA shall thus be viewed as both a continuation and a complement of the Open Data Directive. It fulfills the same objective to enable data (or documents) held by PSBs to unleash their - i.a. economic - potential by favoring their reuse by other actors. The governing principles for doing so differ from these of the Open Data Directive to the extent rendered necessary by the protected nature of the data at stake.

Chapter II DGA deals with the situation where data: (i) held by PBSs and (ii) covered by rights of third parties – either individuals or businesses – (iii) are made available to other parties, *i.e.*, data reusers. PSBs – and competent bodies – are expected to find the balance between the facilitation of data reuse and the preservation of the different types of legal protection of data. Member States and PSBs will have to juggle with the two legal frameworks, which may raise questions as to the respective scope of these frameworks. The difficulty is reinforced by their different legal natures – the Open Data Directive is a directive,

requiring transposing national legislation while the DGA is a regulation, directly applicable without the need for transposition - (Pétel, 2020).

2.3. Commercial data sharing supported by data intermediaries

Both Chapters III and IV DGA deal with data held by any actor – with businesses and individuals as the main targets – on which they have certain rights deemed to serve as a legal instrument for data sharing (DGA, Art. 2(8), (9), (10) and (11)). Identified as "data holders"' and "data subjects"' within the meaning of data protection, these actors are supported in their endeavor to share data with other actors – data users - respectively in a commercial (Chapter III DGA) and in an "altruistic" (Chapter IV DGA) context. The DGA does not require but aims to incentivize data sharing, by arranging conditions in which these actors can trust that they have control over data and over the data sharing process. The present section outlines how the DGA aims to incentivize commercial data sharing while the following section focuses on data altruism.

Chapter III DGA aims to incentivize commercial data sharing by laying down a stringent legal regime for providers of data intermediation services (DIS), namely services supporting the matchmaking and facilitation of commercial data sharing either in the context of bilateral or multilateral data sharing, notably through the operation of data marketplaces (DGA, Art. 2(11)). Large industry players - such as Airbus, Siemens or GE - have created their own data sharing platforms, but with the risk that SMEs have little bargaining power (Impact Assessment, p. 10). In contrast, the DGA lays down stringent requirements to ensure the independence and neutrality of data intermediaries, with the expectation that they become the backbone or orchestrators of data sharing ecosystems open to all interested parties in the context of common European data spaces (DGA, Rec. 28).

Three categories of DISs are identified as falling in the scope:

- DIS specifically addressed at individuals willing to share data relating to themselves - *i.e.*, data subjects with respect to personal data - with data users, such as through personal information management systems (PIMSs);
- DIS consisting in data cooperatives constituted by either individuals or small businesses members which aim to support them in negotiating commercial data sharing with data users, thus acting as a sort of data union;
- Finally, any other DIS whereby commercial data sharing is initiated by actors who are not data subjects with respect to the data, which typically targets the sharing and possibly pooling of any sort of industrial data by businesses (DGA, Art. 10).

Any DIS belonging to any of these three categories shall comply with two sets of requirements, which read – stringently - as legal conditions for the provision of DIS. The nature of these requirements as conditions implies that the

provision of services is subject to postponement, suspension or even cessation in case of infringement. Such and other sanctions – including financial penalties – can be imposed by dedicated enforcement competent authorities to be established by member States (DGA, Art. 14). These sanctions are particularly stringent. The explicit reference to the imposition of the suspension or cessation of services outpaces for example the sanctions that enforcement authorities can impose, under the Digital Services Act, even when "very large online platforms" and "very large search engines" such as Google, X or Facebook violate this Regulation (DSA, Art. 51). It also outpaces the sanctions that DPAs can impose in case data controllers violate the GDPR (GDPR, Art. 58).

The first set of requirements consists of a prior notification of DIS provision to the competent authority. As part of the notification requirement, putative DIS providers shall include information on the nature and category of the services they intend to provide, their name, address, their public website, etc. (DGA, Art. 11). The notification shall notably feed the public registry of DIS providers operated by the Commission. In the context where the definition of DIS remains unclear (Bobev *et al.*, 2023), the notification-related requirements shift the burden of proof of the qualification of services as DISs to providers.

The second set consists of a motley list of structural and – mainly – behavioral requirements that DIS providers shall comply with when providing such services. The requirements can be clustered based on their underlying principles and rationale, with the steady objective that DIS providers come across as trustworthy in the eye of their customers. Out of independence, DISs shall be structurally unbundled from other services, namely provided through a dedicated legal entity. Also, DIS providers shall not use the data entrusted to them by their customers for purposes other than providing DIS. This prohibition is inspired by practices found in the online sector, such as the use by Facebook of user-generated content for the purpose of commercializing advertisement, that the DGA explicitly sets out to prohibit in the context of data sharing intermediation. The DGA also subjects DIS providers to neutrality requirements, such as through the imposing of fair, transparent and non-discriminatory conditions of services, including the price. Several of the requirements pertain to the quality of the services and to their adaptability to the needs of customers, namely to the ambition that customers remain in control of data (sharing). In that respect, customers may notably revendicate their data upon the termination of the service and access the data generated by their activity on the DIS provider technical environment. They shall be informed in case of unauthorized international transfer of their data. DIS providers shall also comply with security requirements and a requirement of service continuity in case of insolvency.

Some of the service quality requirements come down to delegated enforcement, such as the requirement for DIS providers to set up procedures and measures to prevent frauds by their customers – for example data users attempting

to nudge data holders and data subjects into giving in too much data. Finally, while most requirements apply equally to all DIS providers irrespective of the category, the DGA lays down an additional requirement for DIS targeted at data subjects. They shall comply with a fiduciary duty to "act in the best interest of data subjects" who engage into such data sharing (DGA, Art. 12). The DGA does not define this notion but it does provide examples, pertaining mainly to heightened information provision so that data subjects make well-informed choices and do not share data beyond their interest. It remains to be seen how ambitious the interpretation of this notion will be. Especially, a question remains whether it could extent to ex post obligations – *i.e.*, taking place after the data sharing phase – such as the conduct of audits on data users to ensure that they use data in compliance with the agreed conditions, similar to the auditing role entrusted to PSBs vis-à-vis data reusers (section 2.2).

To summarize, DIS providers are expected, on the one hand, to foster commercial data sharing, by providing the technical and legal means (such as template contract provisions) for doing so. On the other hand, this shall be done in balance with other vested interests and especially these of data subjects under data protection, perceived under the DGA as the individual preferences of data subjects concerning data relating to them – a certain account of data control. With Chapter III DGA, the EU legislature bets on a reshaped future data economy consisting of fair and undistorted (data) markets. In light of the stringency of the legal regime and of the balance that they shall find between the fluidification of data transactions and other values - such as data protection, the question has been raised whether DIS providers can find a sustainable business model (von Ditfurth and Lienemann, 2022; Bobev *et al.*, 2023).

2.4. Altruistic data sharing supported by data altruism organizations

Chapters III and IV DGA are explicitly conceived as being in principle exclusive from one another (DGA, Art. 15), although with remaining unclarities, see (Bobev *et al.*, 2023). While Chapter III deals with data markets contexts, Chapter IV deals with "data altruism", namely data sharing without a market price and for objectives of general interest (DGA, Art. 2(16)). For example, individuals could altruistically share fitness trackers or smart wearable data (access to which they will get based on the Data Act, see section 3.2 below) to support research in health.

In the Commission proposal, Chapter IV DGA had some similarities with Chapter III. The Commission proposed to regulate "data altruism organizations" as data altruism intermediaries that, similar to DIS providers, would match data providers with data acquirers without having the possibility to engage into processing data themselves. This model was thus following the same principle of a separation between data collection / provision, data intermediation and data use. Similar to DIS providers, the Commission considered this system based on independence and neutrality of intermediaries as conducive

to trust and ultimately to fostering data altruism. Similar to Chapter III, the requirements were thus focused on the intermediary or match-maker, with a lesser concern for the actual data user. While this logic has remained untouched in Chapter III DGA (see section 2.3), Chapter IV has developed into a legal regime that applies to both to data altruism intermediation and to the processing of altruistically shared data for objectives of general interest. This development is however little explicit – in particular, Chapter IV sticks to the regulation of the same, undefined and unclear notion of data altruism activities.

Similar to Chapter III, Chapter IV DGA consists of two sets of requirements accompanied by strict enforcement, expected to bring trust and thus to incentivize data altruism. One may apply to be registered in a "public national register of recognized data altruism organizations", also feeding a European register, subject to a set of conditions. The then recognized data altruism organization, having obtained the right to display the label, shall then comply with another set of requirements. A major difference with Chapter III DGA lies in that the provision of "data altruism activities" is not conditional upon compliance with these requirements. Compliance with Chapter IV DGA is indeed optional; the registration of one as a data altruism organization within the meaning of Chapter IV grants them the right to use the label "data altruism organization recognized in the Union" (DGA, Art. 17(2)). Chapter IV DGA thus consists of a certification mechanism, leading to a specific sanction regime. Failing compliance, a dedicated enforcement "competent authority" shall have the power to require from the recognized data altruism organization the cessation of the infringement. Failing that, the recognized data altruism organization shall cease to use the label and be removed from the register, thus no longer qualifying as a recognised data altruism organization. This sanction regime is thus different from that of Chapter III DGA, which may imply the cessation of the services (see section 2.3).

In order to be registered, a data altruism organization shall comply with both formalist and substantive requirements. It shall provide information – including on its name, legal status, sources of income, address, public website, the intended objectives of general interest and the nature of data that it intends to control or process (DGA, Art. 19(4)). The organization shall carry out "data altruism activities", namely either data altruism intermediation or the processing of altruistically shared data for objectives of general interest - such as research, combating climate change, official statistics, etc.. It shall be a legal person established under relevant national law to meet objectives of general interest, which is left for member States to decide upon. It shall operate on a non-for-profit basis and be legally independent from any entity operating on a for-profit basis. Data altruism activities shall also be functionally unbundled from other services. Finally, the organization shall comply with the rulebook that the Commission shall adopt (DGA, Art. 22) to increase the preservation the rights and interests of individuals and businesses who altruistically share data.

The then "registered data altruism organization" shall comply with another – although substantively connected to the first one and with the same sanction regime being applicable - set of requirements. It pertains to both transparency of their activities (DGA, Art. 20) and to the protection of the rights and interests of individuals and businesses who altruistically share data (DGA, Art. 21). Prior to data processing, data altruism organizations shall notably inform such individuals and businesses of the objectives of general interest and purposes for which they intend to process data themselves or permit the processing by a data user when acting as an intermediary. Data shall not be processed beyond the objectives or purposes for which individuals and businesses have consented or given permission. Some of the obligations share obvious similarities with the legal regime applicable to DIS providers under Chapter III DGA, such as security obligations, measures in case of international transfer of (non-personal) data and the requirement to provide tools for giving and withdrawing consent or permission. The Commission proposal included essentially an auditing requirement for data altruism organizations intermediating data altruism to ensure that data is not used by data users beyond the objectives or purposes for which individuals and businesses have agreed or given permission. This requirement was not maintained in the adopted DGA – and does not make part of the mandate given to the Commission for the adoption of the rulebook through the future adoption of delegated acts. This runs against the ambition of the EU legislature to grant individuals and businesses control data in the context of data altruism – especially when intermediated.

3. The Data Act: the allocation of rights on or in relation to data

Adopted on 13th December 2023, the Data Act has the ambitious goal to lay down "harmonized rules on fair access to and use of data" in the internal market. It consists of a set of chapters, each of them dealing with data access and use rights and obligations in a different context and with different rationales. After having introduced the general policy vision for the Data Act (section 3.1), this section outlines the different chapters of the Data Act one after the other, including their relation one to the other where appropriate.

3.1. The policy vision

Similar to the policy vision underpinning the DGA, the Data Act is based on the perception of data as having an immense potential for the economy and society as a whole, namely for a range of different purposes whether economic – data-driven services and products – or not – fighting climate change, informing policy-making, etc. Data are considered as non-rival resources that

should thus be shared and (re)used in order for their potential to be unleashed. From that perspective, data monopolization by Big Tech is considered a bad thing insofar as an obstacle to the optimal allocation of data.

In addition to the concern for an economic efficient allocation of data, the Data Act is also driven by a certain idea of "fairness". Both individuals and businesses – such as customers of connected products – should be able to use and generate value from data that they have generated, which is considered as the characterization of data control. In that context, data monopolization by Big Tech – such as car manufacturers with respect to car data, manufacturers of smart farming machinery, etc. – is considered a bad thing insofar as it prevents individuals and businesses from obtaining and making use of data generated by them. It can also prevent third parties having justified reasons to access and use data, such as maintenance or refurbishment providers of connected products.

In this context, the Data Act grants data access and use rights in different contexts and to different types of actors. The policy vision is a two pillars-one, where both economic efficiency and a certain account of fairness shall converge. Fairness under the Data Act shall be understood, similar to the DGA, in the sense of fair and undistorted (data) markets. The Data Act is based on the expectation that the prospect of deriving value from data will incentivize actors – especially individuals and businesses – to act upon the data access and use rights, thus resulting in economic efficiency. For example, based on the right to access data generated by their connected products (Chapter II Data Act, see section 3.2), individuals could choose the repairers of such products and tailor the repair or maintenance to their needs. Another example is that farmers could entrust the data generated by their smart tractors and other farming machinery to independent consultancies to get insights on their soils.

To reinforce the alignment with the economic efficiency goal, the Data Act coins a new concept of "data literacy" that dedicated enforcement authorities, in their enforcement practices, shall promote towards the beneficiaries of rights (Data Act, Art. 37(1)(a)). Not to be confused with digital literacy, data literacy is defined by the Data Act as the skills, knowledge and understanding that allows individuals and businesses to "gain awareness of the potential value of the data they generate" to incentivize them to take an active role in the data markets, viewed as empowerment (Data Act, Rec. 19). Similar to the DGA, the Data Act thus aims to provide the tools and to incentivize individuals and businesses to act as data market participants.

While the DGA focuses on the institutions and structures that support and facilitate data sharing, the Data Act focuses on substantive rights on – or in relation to – data. As further discussed in this section, both legislations – and especially the Data Act - can thus be analyzed as property institutions, defined broadly and functionally as "an institution for organizing the use of [data] as resources in society" (Merrill, 2012).

3.2. Allocation of rights on connected product data

The regulation of connected product data under Chapter II Data Act with reference to Chapters III and IV, constitutes probably the most innovative data-related legal regime deployed by the EU. It is based on the observation that data generated by connected products – smart watches, fridges, industrial machinery, cars, etc. – are often (mis)appropriated by the manufacturer of such products and/or by providers of services associated with the product. This is to the detriment of users of such products who have no data control despite the fact that, according to the Data Act, they participate in data generation either by actively using products or by having their behavior passively "datafied" by connected products. This can also be to the detriment of third parties who could make good use of such data, such as aftermarket service providers, research organizations, etc.

In this context, Chapter II Data Act lays down data rights for the user enforceable against the "data holder" (often the manufacturer), namely the actor having control over data. First, in principle, connected products shall be designed in such a way that data are made accessible to the user (Data Act, Art. 3), *i.e.,* by design compliance. Should that not be the case, then users shall dispose of a data access right enforceable against the data holder (*i.e.*, manufacturer) (Data Act, Art. 4). The data access right is explicitly aimed at enabling data use: Data shall be provided with the relevant metadata, they shall be of the same quality as the ones available to the data holder and especially "machine-readable". If technically feasible, they shall be provided continuously and in real-time. Second, users dispose of the right to require data holders to share data with a third party that they select (Data Act, Art. 5). The selected third party and the data holder shall conclude a contract pertaining to the conditions for such data sharing, which can include a fair compensation – or in other words a fair price (Data Act, Art. 8 and 9). Importantly, the user can exercise this right several times to the benefit of several third parties that they have selected and possibly with respect to the same data. The Data Act also lays down provisions instrumental to the data access and portability rights, such as transparency requirements and the prohibition of manipulative or otherwise deceiving behaviors (known as "dark patterns") impairing the autonomous exercise of their rights by users.

However ambitious, these rights – data access and data portability – are as such not new. The novelty lies in the regulation of the use of data by the three categories of actors – data holder, user and selected third party – who find themselves in a triangular relationship with one another. In line with its property ambition, the Data Act seeks to apportion the benefits of data use and value while mitigating the resulting harmful consequences on the other actors. In walking this ridgeline, the Data Act is not without some contradictions and unclarities, such as concerning the conditions under which the data holder may use data. In that respect, the Data Act distinguishes personal data – deemed to

be exhaustively regulated by the GDPR – from non-personal data which the data holder shall use only "based on a contract with the user" (Data Act, Art. 4(13)). This provision has been interpreted as an indirect exclusive allocation of connected product data to the user, literally contradicting the spirit of the Data Act that data use and value should be shared more broadly and especially between these actors that generate data – users and data holders for that sake. In contrast, the user shall in principle be free to use data for any lawful purpose. As to the selected third party, they shall process data for the purpose and conditions agreed with the user (Data Act, Art. 6), which can be assimilated to a specific (*i.e.*, triangular) form of a data license (Ducuing, 2022; Kerber, 2023).

In addition, the Data Act lays down data use limitations to preserve the rights and legitimate interests of the other actors in the triangle. For example, trade secrets of the data holder shall be preserved by the user and selected third party. Another example is the prohibition to all actors in the triangle to use data to draw insights about the economic situation, assets and production methods of the other actors in the triangle. This prohibition follows from the observation that manufacturers of smart farming machinery could generate and use data on farms crops, soils, etc. that could then be used to maliciously tailor marketing strategies and prices of input goods (such as fertilizers).

To conclude, a filiation can easily be observed between Chapter II Data Act and the scholarly "Principles for a data economy" elaborated by the American Law Institute (ALI) and the European Law Institute (ELI) (ALI-ELI, 2020), and especially with their data rights. Based on the bundle of rights approach to property rights, the ALI-ELI data rights consist of a set of rights and entitlements on data – data access, data rectification, desistance from data use and right to a share in profit. In contrast to the centralization induced by exclusive ownership and in view of the specificities of data, data rights can be allocated concurrently to different actors who have contributed in different ways to the generation of data. "Data co-generation" may notably arise from contributing computing means or having one's behavior or one's equipment behavior captured through data ("datafied"). Symptomatic of the filiation is the reference, in both the European Data Strategy and the Data Act, to the notion of "data co-generation" and the attempt of the Data Act to apportion data rights between the three categories of actors within the triangle with data access as the starting point.

3.3. Harmonized legal conditions for mandatory data sharing

Chapter III Data Act constitutes a visible example of the EU ambition to achieve a certain level of consistency for data legislation. In case a business (data holder) has a legal obligation to share data to another business (data recipient), as a result of the regulation of connected product data (Data Act, Chapter II) or of other EU or national legislation, they shall comply with a set

of harmonized conditions therein laid down. This legal regime shall protect data recipients against abuses from data holders, such as imposing detrimental conditions or abusive price for data access and use. It is unclear whether it applies only to the benefit of business data recipients (Data Act, Art. 8(1)) or also to non-for-profit organizations (Data Act, Art. 9(4)), while the latter may typically find themselves in a weak negotiating position.

The data holder and data recipient are requested to conclude a contract pertaining to data sharing and to the conditions under which it takes place. This can be analyzed as a mandatory - and heavily-regulated - data transaction. Actually, whether there is room for contractual freedom at all is questionable as the contract is explicitly designed to operationalize legal obligations. However possibly at odds with (contract) law, the requirement for a contract obviously fulfills two important functions. It constitutes a safety net ensuring that all relevant aspects are covered. For example, while Chapter III Data Act does not explicitly lay down the liability consequences of data sharing, the contract would typically include related stipulations. Also, it justifies the triggering of Chapter IV Data Act laying down fairness standards for data-related contracts between businesses (see section 3.4).

The data holder shall share data with data recipients under so-called FRAND - namely fair, reasonable and non-discriminatory - terms, a regulatory toolkit well-known to competition law and to the regulation of network industries (railways, telecom, electricity, etc.). A question that has been intensively discussed within the EU institutions and in the scholarship is the regulation of the price (referred to as "compensation") for data sharing. A crucial question was whether the law should indirectly endorse data commodification by allowing for a price for data as a traded good. The adopted Data Act is equivocal on that matter. On the one hand, data holders may charge a fair and reasonable compensation that shall be cost-based; And Rec. 46 explicitly bans an interpretation of the compensation as a "payment for the data itself". On the other hand, and except where the data recipient is a small or medium enterprise (SME) or a non-for-profit research organization, the data holder may, subject to contractual negotiations, include a margin and the compensation may depend on the quality of the data. This points more towards a market price, with the risk of legally endorsing, indirectly, the factual control of data by data holders, which the Data Act explicitly opposes (Margoni, Ducuing and Schirru, 2023). Both or either the data holder and/or the data recipient may take their case to a dispute settlement body, namely an out-of-court dispute resolution settlement body (Data Act, Art. 10), in addition to the competence of enforcement competent authorities.

Finally, Chapter III Data Act allows data holders – in the context of both the regulation of connected product data and of other data sharing requirements – to apply "appropriate technical protection measures" (TPMs) (Data Act, Art. 11). Stemming from the field of copyright, TPMs refer to the

technical enforcement of legal rules, deemed more effective in the digital environment. Under the Data Act, TPMs may be designed and deployed by data holders to prevent unauthorized access to data – likely referring to access to data beyond the rights granted by law – and to ensure that their trade secrets are respected. Similar to the tolerance for the data holder to charge a compensation in case of mandatory data sharing, the legal protection of TPMs raises the question whether the law indirectly – and inadvertently – endorses factual control over data, in stark contradiction with its alleged goal (Margoni, Ducuing and Schirru, 2023).

TPMs are by default allowed without any prior legal authorization being needed. What the Data Act does bring is their legal recognition and protection so that the altering or removal of TPMs deployed by the data holder triggers a sui generis sanction regime inspired by IPRs. The data holder may notably request the erasure of data accessed as a result but also the end of the production, of the sale and the destruction of derivative data, services and/or goods. Intensively discussed during the legislative process, the sanction regime is thus far-reaching in the sense that legal consequences concern not only data but also derivative productions. To compare, the sanction regime under the GDPR extends only to data – with the possibility for the data protection authority (DPA) to impose a ban on data processing. But DPAs may not require that, for example, an AI-driven software developed based on the illegal processing of personal data be destroyed and/or removed from the market (GDPR, Art. 57).

3.4. Data-specific regulation of unfair commercial terms

Chapter IV Data Act is based on the observation that, in the absence of a property legal status of data, most data access and use across the EU actually relies on and is regulated through contracts. However, in the presence of power imbalance between businesses, influential actors may easily reserve data use and value for themselves or share data only under prohibitive unfair contractual terms, to the detriment of weaker and smaller businesses (especially SMEs).

The EU already disposes of a comprehensive set of rules protecting consumers against unfair commercial practices (see in particular the Unfair commercial practices Directive) and contractual terms of traders (see in particular the Unfair contract terms and the Consumer Rights Directives) (B2C). These rules apply horizontally, namely in principle to all sectors and irrespective of the nature of the goods and services. In contrast, with B2B, the EU has proceeded with a piecemeal approach that addresses specific problems in specific sectors. The EU has already regulated B2B unfair commercial practices in the agriculture and food supply sector (Directive on B2B trading practices in the agricultural and food supply chain) and in the online platform sector (Regulation promoting fairness and transparency for business users of online intermediation services, or Platform-to-Business Regulation; see also the Digital Markets

Act, albeit leaning more towards a competition instrument). Chapter IV Data Act continues this piecemeal approach, with its entry point being this time the traded object, namely data, irrespective of the sector.

Chapter IV Data Act regulates contractual terms between businesses that concern the access to and use of data or the liability and remedies for the breach or the termination of data-related obligations. On the twofold condition that a term has been unilaterally imposed by a business onto the other and that such term is unfair, it shall not be binding onto that other business (Data Act, Art. 13(1)).

Unfairness can be determined in three ways. The Data Act includes: (i) a "black list" of terms which shall irrefutably be considered unfair; (ii) a "gray list" of terms which are rebuttably presumed to be unfair; and (iii) a broad definition of "unfair contractual terms" as these terms of such a nature that their use "grossly deviates from good commercial practice in data access and use, contrary to good faith and fair dealing" (Data Act, Art. 13(3)). Insofar as it subjectively relies on "good commercial practice in data access and use" as a standard, this definition is hard to operationalize, precisely because of the absence of such a standard (Margoni, Ducuing and Schirru, 2023). Most of the "black" and "gray" lists items are not specific to data and may already fall under (national) contract law. Such is the case of the contractual term that gives the party that unilaterally imposed it the exclusive right to interpret contractual terms (Data Act, Art. 13(4)(c)) or to unilaterally and substantially change the price without reasonable justificatory ground for it (Data Act, Art. 13(5)(g)).

By exception, some items are truly novel and data-specific. Such is the case of the contractual term, rebuttably presumed unfair, that prevents the party upon whom the term has been unilaterally imposed from using data provided or generated by that party during the period of the contract or to unduly limit such use "to the extent that that party is not entitled to use, capture, access or control such data or exploit [its] value in an adequate manner" (Data Act, Art. 13(5)(c)). This provision has been interpreted as a form of indirect – *i.e.*, through contract regulation – allocation of data (Margoni, Ducuing and Schirru, 2023). Indeed, it allocates data usus and fructus entitlements – albeit only within the specific scope of Chapter IV Data Act. Also, this allocation symptomatically follows the fact that the party has generated data. Reminiscent of the Lockean labor justification for property (Locke, 2018), the notion of "data generation" is not defined under the Data Act. This begs the question which type of activity(ies) shall be deemed to qualify as generative of data. Especially, it raises the question whether the mere fact of being passively datafied (or having one's equipment behavior datafied) would suffice. A positive answer is likely, given: (i) that being datafied constitutes a relevant justificatory ground for the user of connected products to be allocated rights on data (section 3.2) and (ii) the filiation, when so doing, with the ALI-ELI Principles for a data economy which explicitly recognize this circumstance as generative of data (section 3.2).

3.5. Legal framework for mandatory business-to-government data sharing

"Government-to-business" (G2B) mandatory data sharing was the starting point of data-related legislation in the EU, under the PSI – and then Open Data – Directive (section 2.2). The reverse situation - especially business-to-government (B2G) data sharing – was, until the Data Act, not regulated horizontally but subject to piecemeal regulations mainly at national level. B2G mandatory data sharing has been intensively discussed in the scholarship, such as under the heading of data for good or data for the public interest (Verhulst *et al.*, 2019). The relevance of the B2G mandatory data sharing discussion has been reinforced by the Covid-19 pandemic, which illustrated the extent to which governments were in need of a wide range of data held by the private sector.

The Open Data Directive is based on the principle that data held by PSBs shall be made available for reuse by anyone and for any purpose (see section 2.2). As instantiated by the Data Act, the reverse situation of B2G mandatory data sharing – sometimes referred to as a form of non-exclusive data expropriation - is based on a much stricter principle. The Data Act aims to find a balance between the pursuit of public or general interests and the preservation of the freedom to conduct a business of businesses, with the high ambition to come up with a horizontal legal regime, namely, to a significant extent, a purpose- and sector-agnostic one. The principle is that PSBs, the Commission, the European Central Bank or a Union body (together: public authorities) request a business to share data only in case of an "exceptional need" (Data Act, Chapter V).

An exceptional need can, first, be identified where the requested data – whether personal or non-personal - are necessary to respond to a "public emergency", namely an exceptional situation, limited in time, that negatively affects the population of the Union or the whole or part of a member State, with a risk of serious and lasting repercussions for, among other things, living conditions or economic or financial stability (Data Act, Art. 2(29)). Then, a condition is that there is no timely and effective manner, under equivalent conditions, for public authorities to obtain the necessary data. The imprint of the Covid-19 pandemic on this provision is self-evident.

Second, in contrast to responding to a public emergency, the other circumstances in which public authorities may request data do not address the substantive nature of the "exceptional need". Also, and in contrast to the original Commission proposal, they concern only non-personal data in the face of the serious risk of interference of Chapter V Data Act with data protection. Public authorities may then request data from businesses where the absence of such data would prevent them from fulfilling a task carried out in the public interest and based on EU or national law. The Data Act indicates official statistics and the mitigation of or the recovery from a public emergency (such as, again, a pandemic) as relevant illustrations. Finally, public authorities may request data

from businesses, irrespective of the purpose, when they have exhausted all the other means at their disposal to obtain data in a timely manner. This reads as a catch-all provision for all sorts of situations that cannot be anticipated in advance. This implies that they could neither purchase such data on data markets at market price – thereby endorsing that data can have a price (see section 3.3) - nor rely on existing or future data sharing obligations. The latter condition will likely have a differentiated impact on member States depending on their institutional organization, with some sub-national (regional) authorities having the competence to pass laws and others not.

Chapter V Data Act further regulates the conditions in which public authorities request data to businesses and the conditions in which they may use data, to prevent abuse on their behalf and to minimize the detrimental impact on businesses (Data Act, Art. 17 to 19). In particular, public authorities shall "not use data in a manner incompatible with the purpose" for which they were requested, a wording reminiscent of the purpose limitation principle under the GDPR (GDPR, Art. 5(1)(b) and Art. 6(4)). They shall cease to use and erase data as soon as no longer necessary for such purpose and they shall implement measures to preserve the confidentiality and integrity of data, such requirements being reminiscent of GDPR principles (GDPR, Art. 5(1)(c) and (f)). The possibility to share data with third parties is strictly regulated, with a more lenient approach where the exceptional need implies that research organizations also process data (Data Act, Art. 21). Public authorities are under the same prohibition as discussed in section 3.2 (regulation of connected product data) concerning the use of data to derive insights about the economic situation, assets and production or operation methods of the business who shared data. Disclosure of trade secrets as a result of mandatory data sharing shall be requested only when necessary and subject to the adoption of protective measures by public authorities.

A debated question has been whether mandatory data sharing shall be subject to compensation, to which the Data Act offers a differentiated answer depending on the legal basis. When based on the exceptional need taken from responding to a public emergency, no compensation can be claimed by the business, allegedly justified by the magnitude of such exceptional need. In the other cases, businesses may claim a fair compensation, namely one that is limited to the costs of making data available in addition to a reasonable margin (Data Act, Art. 20).

3.6. Switching between data processing services

Chapter VI Data Act is an odd man out in the following respect: In contrast to the remainder of the Data Act, it is not based on the perception of data as a non-rival resource the use of which should relatedly be optimized. A commonality though is that it fulfills a property function by (re)assigning data (Geiregat, 2022).

Chapter VI Data Act constitutes the long-(over)due response to the problem of customers of cloud computing service - whether individuals or businesses-getting locked-in an incumbent contractual relationship. Customers entrust cloud computing service providers with the storage, organization and otherwise processing of data and digital assets. They may become dependent on the technological infrastructure of service providers, which ends up shaping their digital assets and, ultimately, activities conducted in the cloud. Yet, the lack of competition in these markets results in power asymmetries, leveraged by service providers to prevent their customers to switch to another service provider. The EU tentatively addressed this problem with the Free Flow of Non-Personal Data Regulation. Based on a semi-soft law approach with the mere "encouragement" and "facilitation" of self-regulatory codes, the Regulation failed by general opinion.

With Chapter VI Data Act, the EU makes a 180 degree turn and chooses to regulate, in details, the ways in which cloud (and edge) computing service providers (together: data processing service providers) shall enable their customers to "switch" to one or several, possibly competing, provider(s) offering the same service type. Alternatively, the customer may also choose for an on-premises ICT infrastructure provider, *i.e.,* non-cloud one. In other words, customers shall be able to find an alternative equivalent service, with equivalence being assessed based on customers' needs.

Switching involves (at least) three actors: the "source provider of data processing services", the customer – whether a consumer or a business, and the destination provider (Data Act, Art. 2(34)). The Data Act does not grant customers a general positive "right to switching" enforceable against the source provider but lays down a detailed list of such specific measures that source providers shall take to enable switching (obligations to do certain things). This includes also the prohibition of certain practices (obligations to not do certain things) by source providers that currently prevent or inhibit customers from switching. The latter obligations can be compared to a sector-specific regulation of unfair commercial practices (see section 3.4). This notably includes the prohibition of such commercial, contractual or technical practices that complicate the termination of contract or the conclusion of a new contract with another service provider, as well as forced bundling of services of such a nature as to prevent customers from partial switching.

The obligations to do certain things to enable switching can be understood as forming a new additional mandatory service that data processing service providers shall offer to their customers under heavily regulated terms. Regulation pertains to both the contractual – including charging – and technical conditions for switching and includes instrumental transparency obligations (Data Act, Art. 25 to 30). Data processing service contracts shall notably include a contractual allowance to switch, including the provision of reasonable assistance from the

source provider for doing so. The Data Act regulates in detail the practical ways in which switching shall be made possible, including time limits, necessary prior exchange of information, etc. Importantly, source providers shall erase data and digital assets upon switching. Together with porting, inherent in switching, this can be viewed as a contractual right of customers to revendicate data and digital assets, yet missing in the absence of a property legal status for data. In the long term (from 12 January 2027), switching shall be provided for free, namely with no dedicated price. Until then, data processing service providers may charge (only) the direct costs of switching, excluding a margin (Data Act, Art. 29).

The Data Act accounts for the differences between the types of data processing (especially cloud) services. For example, the ambition to obtain "functional equivalence" - namely an acceptable level of service quality, between the use of, respectively, the source and the destination services -cannot be reasonably expected in case of tailored – in contrast to off-the-shelves - services (Data Act, Art. 31(1)). The Data Act also caters for the different types of data processing services concerning the interoperability measures necessary for switching, including (European) standards and open interfaces (Data Act, Art. 30).

Finally and unrelatedly, the Data Act lays down obligations incumbent on all data processing service providers with respect to international transfer of non-personal data – of their customers - upon the request of third-country governments. Data processing service providers shall bear the special responsibility, in principle, to filter out and oppose such requests that conflict with EU or national law, subject to an elaborate set of rules (Data Act, Art. 32). This attempt to protect European industrial interests against third countries (one may easily think of US requests to Microsoft for example) constitutes but one illustration of the neo-mercantilist approach of the EU with EU data legislation. It shares similarities with the approach of the DGA – requesting data governance structures to bear similar responsibilities (see sections 2.2 and 2.3) (Baloup *et al.*, 2021; Ducuing *et al.*, 2022). The EU legislature thereby designs a legal framework for international transfer of non-personal data inspired by and allegedly complementing this of the GDPR for personal data, referred to as GDPR mimesis (Miadzvetskaya, 2023).

3.7. Interoperability and common European data spaces

Like the DGA, the Data Act - as a deemed horizontal legal framework - is expected to support common European data spaces. The DGA shall provide for the data governance structures for different types of data sharing while the Data Act grants rights and entitlements on data which enable a greater number of individuals and businesses to make active use of data within such data spaces.

The Data Act provides for the only – albeit loose - legal definition of common European data spaces as the legal and technical means to enable data sharing in scenarios deemed of general interest (Data Act, Art. 33(1), see section

1). Seemingly of a political nature, the notion of common European data space does bring about legal consequences for "participants [in such data spaces] that offer data or data services to other participants", an equally loose and unclear notion. These participants shall contribute to the interoperability effort envisaged by the Data Act to enable the sharing of data (Data Act, Art. 33(1)). Interoperability requirements imply that participants provide or even produce relevant metadata concerning the data, such as describing transparently their quality, source, etc. and the technical means (including application programming interface software, *i.e.*, APIs) to access data. In view of data spaces, the Data Act also provides the framework for the future adoption of harmonized standards, that participants will then also have to comply with.

The combination of two unclear notions – common European data space and participants in such data spaces – leaves a considerable room for interpretation. The most extensive interpretation would have anyone engaging in voluntary data sharing in any form, for any purpose and in any scenario being compelled to abide by such interoperability requirements. This would be not only unworkable but would also run radically against the ambition of the EU to foster data sharing. The strictest interpretation would include only such actors that participate in structuring the legal and technical infrastructure for data sharing in a specific context, thus playing a truly data governance function, possibly only in the presence of dedicated governance mechanisms (based on corporate law or on consortium contracts).

4. Commonalities and elements of consistency within EU data legislation

At first glance, this overview, and especially the Data Act, may give the impression of a legislative patchwork, with legal regimes borrowed from - or inspired by - a wide range of branches of law, including (broadly defined) property law, contract law, the regulation of unfair commercial practices, data protection, technical harmonization and public and administrative law. The only common denominator seems to be the focus on data. And even, it appears that older notions – such as documents under the PSI Directive – have been partly rebranded as data. Also and as recounted throughout this chapter, the DGA and – even more so – the Data Act result from political discussions and compromises, which inevitably affects their legal consistency and workability. This is particularly visible with "common European data spaces" which, however seemingly key to the EU data policy, are not well-delineated enough a legal concept. Another unclear notion, connected to similar notions of "agency" and "empower(ment)", is this of data control. It remains without a legal definition and only a broad understanding as the ability of one – deemed legitimate – actor

to make decisions concerning the use or processing of data and the ability to generate value from such data.

However, some commonalities and elements of consistency can be identified. EU data legislation does not provide a legal status for all data, like, for example, property law would do. However, it does go beyond mere sector- or context-specific fixes to certain problems - something that future data space-specific regulation could provide in addition. EU data legislation undoubtedly has horizontal – or universal – ambitions. This is visible with the DGA, which attempts to incentivize data sharing in a wide range of scenarios and in particular both market and non-market ones. The horizontality of EU data legislation is also visible with B2G mandatory data sharing that applies irrespective of the type of data, type of public authority and even, to a large extent, irrespective of the type of exceptional need at stake. Similarly, the regulation of connected product data applies irrespective of the sector. As a broader range of sectors and products become smart or connected, the scope will naturally broaden. The data-specific regulation of unfair commercial practices between businesses has also a broad scope, as it applies to any contractual term between them dealing with data access and use, irrespective of the context or sector. Finally, the ambition of EU data legislation to provide for a horizontal regulation of data is visible from the harmonized legal conditions for mandatory data sharing between businesses which applies irrespective of the sector, context and, even more striking, *even rationale* for mandatory data sharing (Ducuing *et al.*, 2022).

Even more important, a set of common features and principles can be identified. With the exception of the regulation of data processing services, EU data legislation is based on the same premise that, as non-rival resources, data shall, in principle, be shared and broadly used in order to optimize their economic potential. Then, the whole of EU data legislation consists of balancing this principle with contextual specificities requiring more closedness. The legitimate rights and interests of different actors – and especially individuals and businesses – are symptomatically put increasingly on equal footing, with data protection, IPRs and trade secrets essentially viewed as rights and entitlements of some actor(s) on or in relation to data (EDPB and EDPS, 2021). On the one hand, they may tip the scales more towards closedness and, on the other hand, they are interpreted, especially under the DGA, as legal mechanisms likely to support data sharing akin to property entitlements. This raises the question whether EU data legislation is compatible with data protection. In particular, it remains to be assessed whether data protection may continue as we know it while being subsumed, as it is, into such an economic and market framework.

This can be associated with the property ambitions of EU data legislation that consistently aims to provide for a form of initial allocation of data (Data Act) then allowing for secondary allocation through different types of data sharing (especially under the DGA). Together, the DGA and the Data Act are

therefore forming a private law infrastructure for data. "Data generation" constitutes a recurring justificatory ground under the Data Act for the granting of rights to use data or prevent their use by others, observable with the regulation of connected product data and with the data-specific regulation of unfair commercial practices between businesses. EU data legislation does not provide for a definition of activities deemed "generative" of data, despite the fact that this notion produces legal effects. Subject to future case law, a broad concept of data generation seems to be emerging (see sections 3.2 and 3.4), encompassing even the mere passive state of being - or having one's equipment – "datafied".

Relatedly, however imprecise, the notion of "data control", as supported by EU data legislation, appears to constitute a consistent principle, expected to serve as a bridge between several contradictory objectives: (i) while data cannot and shall not be exclusively "owned", data control constitutes a soft property functional equivalence for data, likely to legally support different types of data sharing activities and to enable individuals and businesses to generate value from data; (ii) data control is also demonstrably expected to bridge private interests of individuals and businesses with respect to "their" data with what the EU legislature considers as the general interest, namely unleashing the (economic) potential of data while protecting EU industrial interests on the international plane. In that sense, data control lies at the core of the neo-mercantilist project of the EU; and (iii) EU data legislation aims to bring, under the banner of data control, different types of rights and legitimate interests, and especially these of, respectively, individuals and businesses concerning, respectively, personal and non-personal data. This reinforces the sense of a unified "EU data law" for which, undoubtedly, data control constitutes the cornerstone.

Bibliography

ALI-ELI (2020) 'Principles for a Data Economy'. American Law Institute - European Law Institute. Available at: https://www.europeanlawinstitute.eu/projects-publications/current-projects-feasibility-studies-and-other-activities/current-projects/data-economy.

Baloup, J. *et al.* (2021) *White Paper on the Data Governance Act* (June 23, 2021). CiTiP Working Paper 2021. Available at: https://ssrn.com/abstract=3872703 or http://dx.doi.org/10.2139/ssrn.3872703.

Bietti, E. (2020) 'Consent as a Free Pass: Platform Power and the Limits of the Informational Turn', *Pace Law Review*, 40(1), p. 310. Available at https://doi.org/10.58948/2331-3528.2013.

Bobev, T. *et al.* (2023) *White Paper on the Definition of Data Intermediation Services* (October 2, 2023). Available at: https://ssrn.com/abstract=4589987 or http://dx.doi.org/10.2139/ssrn.4589987.

von Ditfurth, L. and Lienemann, G. (2022) 'The Data Governance Act: – Promoting or Restricting Data Intermediaries?', *Competition and Regulation in Network Industries*, 23(4), pp. 270–295. Available at https://doi.org/10.1177/17835917221141324.

Drexl, J. (2019) 'Legal Challenges of the Changing Role of Personal and Non-Personal Data in the Data Economy' in De Franceschi, A. and Schulze, R. (ed.) *Digital Revolution: Data Protection, Smart Products, Blockchain Technology and Bitcoins Challenges for Law in Practice*, München, Beck, 2019, pp. 19-41, Max Planck Institute for Innovation & Competition Research Paper No. 18-23, Available at: https://ssrn.com/abstract=3274519.

Ducuing, C. (2022) *An analysis of IoT data regulation under the Data Act proposal through property law lenses*. CiTiP Working Paper Series. Leuven, Belgium: KU Leuven, p. 29. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4225027.

Ducuing, C. *et al.* (2022) White Paper on the Data Act (June23, 2021). CiTiP Working Paper 2021, Available at: https://ssrn.com/abstract=3872703 or http://dx.doi.org/10.2139/ssrn.3872703.

EDPB and EDPS (2021) 'Joint Opinion 03/2021 on the Proposal for a regulation of the European Parliament and of the Council on European data governance (Data Governance Act)'. Available at https://edpb.europa.eu/system/files/2021-03/edpb-edps_joint_opinion_dga_en.pdf.

Geiregat, S. (2022) *The Data Act: Start of a New Era for Data Ownership?*, p. 56. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4214704.

Graef, I., Gellert, R. and Husovec, M. (2018) *Towards a Holistic Regulatory Approach for the European Data Economy: Why the Illusive Notion of Non-Personal Data is Counterproductive to Data Innovation*. TILEC Discussion Paper No. 2018-029, Available at: https://ssrn.com/abstract=3256189 or http://dx.doi.org/10.2139/ssrn.32561899.

Kerber, W. (2023) 'Governance of IoT Data: Why the EU Data Act Will not Fulfill Its Objectives', *GRUR International*, 72(2), pp. 120–135. Available at: https://doi.org/10.1093/grurint/ikac107.

König, P.D. (2022) 'Fortress Europe 4.0? An analysis of EU data governance through the lens of the resource regime concept', *European Policy Analysis*, 8(4), p. 484–504. Available at: https://doi.org/10.1002/epa2.1160.

Locke, J. (2018) 'Second Treatise of Government' in Cohen, M. (ed.). *Princeton Readings in Political Thought: Essential Texts since Plato*. Princeton: Princeton University Press, pp. 213-243. Available at: https://doi.org/10.2307/j.ctv19fvzzk.

Margoni, T., Ducuing, C. and Schirru, L. (2023) 'Data property, data governance and Common European Data Spaces', *Computerrecht: Tijdschrift voor Informatica, Telecommunicatie en Recht* [Preprint].

Martens, B. (2021) *An economic perspective on data and platform market power*. 2020–09. JRC, European Commission, p. 31.

Merrill, T.W. (2012) 'The Property Strategy', *University of Pennsylvania Law Review*, 160, p. 35.

Miadzvetskaya, Y. (2023) 'Data Governance Act: On International Transfers of Non-Personal Data and GDPR Mimesis', *European Data Protection Law Review*, 9(1), pp. 13–26. Available at https://doi.org/10.21552/edpl/2023/1/6.

OECD (2016) *Data-Driven Innovation: Big Data for Growth and Well-Being*. Éditions OCDE. Available at https://doi.org/10.1787/9789264229358-en.

Pétel, A. (2020) 'Publication de l'Acte sur la gouvernance des données, les propositions de la Commission européenne', *Revue Lamy Droit de l'immatériel*, 176, pp. 43–47.

Purtova, N. and van Maanen, G. (2023) 'Data as an economic good, data as a commons, and data governance', *Law, Innovation and Technology*, 16(1), pp. 1–42. Available at: https://doi.org/10.1080/17579961.2023.2265270.

Riis, N. (2023) 'Shaping the field of EU Data Law', *JIPITEC - Journal of Intellectual Property, Information Technology and E-Commerce Law*, 14(1). Available at: http://www.jipitec.eu/issues/jipitec-14-1-2023/5707.

Ruppert, E., Isin, E. and Bigo, D. (2017) 'Data politics', *Big Data & Society*, 4(2). Available at: https://doi.org/10.1177/2053951717717749.

Streinz, T. (2021) 'The Evolution of European Data Law (Chapter 29)', in Paul Craig and Gráinne de Búrca (eds.) *The evolution of EU law*. 3rd edn. Oxford: Oxford University Press.

Verhulst, S.G. *et al.* (2019) *Leveraging private data for public good - A Descriptive Analysis and Typology of Existing Practices*. Data Collaboratives, p. 57. Available at: https://datacollaboratives.org/static/files/existing-practices-report.pdf.

Chapter V
The United States of America Approach to Digital Regulation

by Simona Klimbacher*

Index: 1. The early stages of US digital regulation: freedom first. – 2. Towards a new digital regulatory framework. – 3. The US approach to Artificial Intelligence regulation. – 4. Data Privacy: the American way and the interactions with the GDPR. – 5. Digital legacy regulation: a new frontier.

1. The early stages of US digital regulation: freedom first

The universally renowned national anthem "The Star-Spangled Banner", which designates the United States of America as "the land of the free", seems to have greatly inspired the US approach to digital regulation, especially in the last decades. Indeed, since the dawn of the digital era, companies and internet users in the United States have traditionally benefited from a rather liberal regulatory framework based on the principle of freedom.

It is widely acknowledged that a pivotal role in this respect was played by Section 230 of the Communications Decency Act ("CDA"), also known as Title V of the Telecommunications Act, that was passed by the US Congress in 1996. Its initial aim was to restrict minors' access to online pornography. Following strong protests from civil society, the US Supreme Court abolished many of the CDA's provisions; however, Section 230 remained, and has subsequently become one of the most powerful tools for protecting online platforms' operations and fostering the rapid expansion of Silicon Valley's digital giants.

Section 230, which stipulates that "No provider or user of an interactive computer service shall be treated as the publisher or speaker of any information provided by another information content provider", has often been called "the 26 words that made the Internet". In fact, the enactment and the subsequent case law supporting the constitutionality of Section 230 (Zeran *v.* American Online, Inc., 1997; Barnes *v.* Yahoo!, Inc., 2009) have been considered essential to the growth of the Internet through the early part of the 21st century.

---

* Attorney admitted in Italy and New York State (USA). Master of Laws (LL.M.) at Fordham University School of Law. Research Fellow at the Information Society Law Center (ISLC) at the University of Milan.

Coupled with the Digital Millennium Copyright Act ("DMCA") of 1998, Section 230 provides internet service providers safe harbours to operate as intermediaries of content without fear of being liable for that content, as long as, they take reasonable steps to delete or prevent access to that content. Accordingly, Internet users' publications on platforms such as YouTube, Amazon, Twitter, Facebook or Instagram, as well as blogs or forums, do not fall under the responsibility of the same platforms. These protections allowed innovative applications in the digital space without fear of legal ramifications, creating the foundations of modern Internet services such as advanced search engines, social media, video streaming, cloud computing and, lately, artificial intelligence.

The CDA is probably the law that has contributed most to protecting freedom of expression in the United States. Freedom of expression is an integral part of the American national identity. The First Amendment to the US Constitution is the fundamental piece of legislation protecting this freedom. Incorporated into the Bill of Rights in 1791, this amendment was designed to protect individuals' freedom of religion and free speech from government and congressional interference. The amendment prohibits Congress from establishing a national religion, and from interfering with freedom of speech and of the press; within this framework, it is up to the government to provide considerable justification if it interferes with citizens' speech.

The famous Declaration of Independence for Cyberspace of February 8, 1996, made public by John Perry Barlow, Internet pioneer and co-founder of the Electronic Frontier Foundation, reflects American society's disapproval of government regulation in the digital space. In his text, Barlow strongly criticized the Telecommunications Act, passed by the US Senate in 1996, stating that those who passed the law were mistaken, failing to understand the difference between cyberspace and the real world. In Barlow's view, the virtual world could very well regulate itself without any government intervention; this line of thinking was widely shared in the US in the early stages of the digital era.

Between the late 1990s and the early 2000s, the promotion of online freedom of expression by tech companies and civil associations was so strong that content was barely regulated. The efforts of civil society to create a free cyberspace were also encouraged by jurisprudence; in several major rulings on the regulation of online content, American courts ruled in favour of Internet users' freedom of expression.

It was only after the Cambridge Analytica scandal, revealed in 2018, that the situation rapidly changed. Russian interference in the 2016, US presidential elections via social networking manipulations attracted intense public attention; the fact that private interests - from Facebook and Cambridge Analytica - were able to influence election results proved how the freedom granted to platforms can backfire on democracy itself. This event put the United States regulators

under unprecedented pressure to start imposing constraints on platforms, as it revealed that - contrary to what they proclaim in public - platforms may undermine citizens' freedom by surreptitiously guiding their choices. The immunity of digital platforms started then being called into question by citizens, giving the political inspiration to a number of legislative initiatives targeting a wide range of potential online harms.

An example was the enactment by the US Congress in 2018 of the "FOSTA-SESTA package", a law against online sex trafficking which amended Section 230 safe harbours to limit digital platforms' immunity if they have, or should have had, knowledge of sex trafficking being organized through their services. For the first time, Internet platforms were held accountable for the content that their users post; this led to the decline of popular classified websites such as Backpage and Craigslist, who were forced to shut down their adult services sections.

More recently, in May 2020 President Trump signed the Executive Order on Preventing Online Censorship, which aimed at significantly limiting the safe harbour protections granted to platforms by Section 230. However, following intense criticism, one year later President Biden revoked the Executive Order, along with various other Trump's orders.

The loose approach to digital regulation adopted in the US in the past decades was also reflected in other legislative sectors, such as competition law. In the name of entrepreneurial freedom, and in the manifest lack of rules adequate to target the peculiarities of the digital age's economy, major players have benefited for many years from a sort of "competition immunity".

The impressive growth of leading digital platforms such as Big Tech (Google, Apple, Facebook, Amazon, Microsoft) was mainly achieved through the acquisition of innovative digital companies and start-ups. Mergers and acquisitions were crucial to the growth of the digital giants and helped them to secure their strategic positioning and expansion in a market driven by innovation. In a report published by the Federal Trade Commission ("FTC") in 2021, it was disclosed that in the period between 2010 and 2019 the Big Tech were engaged in more than 600 acquisitions, of which around 50% concerned targets aged less than five years, and 65% concerned transactions valued less than US$ 25 million.

Further, players operating in a wide variety of activities within the digital space are better placed to exploit large quantities of data to train their algorithms; the quantity of data collected is directly linked with the accuracy of algorithms in predicting user preferences, and therefore the ability of platforms to make profits from advertising. This driver led to some of the major tech M&A deals of the last decade, such as the acquisitions by Facebook of Instagram in 2012 and WhatsApp in 2014, or the acquisition of LinkedIn by Microsoft in 2016.

More recently, however, US regulators have turned increasing attention to platform abuses of dominance. In June 2019, the House Judiciary Committee

("HJC") initiated a bipartisan investigation into the state of competition in digital markets, led by the Subcommittee on Antitrust, Commercial, and Administrative Law. The Subcommittee examined the dominance of Amazon, Apple, Facebook, and Google, and their business practices to determine how their power affects US economy and democracy. Additionally, the Subcommittee performed a review of existing antitrust laws, competition policies, and current enforcement levels to assess whether they are adequate to address market power and anticompetitive conduct in digital markets. In their report made public in October 2020, the lawmakers concluded that the four companies are all engaged in illicit practices, including eliminating competitors via acquisitions, charging exorbitant fees, or signing unequal contracts with small businesses. The Subcommittee's series of hearings produced significant evidence that these firms exercise their dominance in ways that erode entrepreneurship, degrade Americans' privacy online, and undermine the vibrancy of the free and diverse press; the result is less innovation, fewer choices for consumers, and a weakened democracy.

The recommendations put forward by the HJC in the report are aimed at restoring competition in the digital economy through a number of measures, such as: (i) structural separations and prohibitions of certain dominant platforms from operating in adjacent lines of business, (ii) non-discrimination requirements, prohibiting dominant platforms from engaging in self-preferencing, and requiring them to offer equal terms for equal products and services, (iii) interoperability and data portability, requiring dominant platforms to make their services compatible with various networks and to make content and information easily portable between them, (iv) presumptive prohibition against future mergers and acquisitions by the dominant platforms. The HJC further recommended to strengthen and update the antitrust laws, and to revive antitrust enforcement providing the competent agencies with adequate resources and budgets.

In the wake of HJC's purpose to better regulate the digital arena, since 2021 lawmakers in both chambers of the US Congress have introduced several bills signaling increasing bipartisan appetite for legislative antitrust reform to target tech companies and digital platforms. Amongst them, the one that received the most traction is the "American Innovation and Choice Online Act". This bill would prohibit certain large online platforms from engaging in self-preferencing, unfairly limiting the availability on the platform of competing products from another business or discriminating in the application or enforcement of the platform's terms of service among similarly situated users. The bill also proposes restricting a platform's use of non-public data obtained from or generated on the platform, and prohibits the platform from restricting access to platform data generated by the activity of a competing business user.

Finally, in the last few years, the US Antitrust Agencies *i.e.*, the US Department of Justice (DOJ) and the FTC have increased scrutiny on technology companies and digital services. In particular, the US Antitrust Agencies, supported by the White House, seem to have embraced the "big is bad" philosophy when assessing antitrust issues in the technology sector. The US Antitrust Agencies acknowledged that traditional antitrust laws are inadequate to address modern challenges facing the technology sector, and they are increasingly committed in developing "multidimensional" approaches to challenging anticompetitive conducts in the digital space. This new attitude to antitrust enforcement drives a strategic vision that aims to bring large cases against major technology companies, while developing new and creative ways to do so.

2. Towards a new digital regulatory framework

For many years the US approach to digital regulation has been quite relaxed. Refraining from detailed and intrusive digital regulation fitted well with the constitutional concept of limited government and the broadly accepted attitude of limited market intervention. However, the US is now at a historical crossroads, and must decide whether it is time to change its stance, somehow following the European Union's path in adopting digital regulation to limit the growing powers of digital platforms. The regulatory dilemma on how to handle digital platforms is a complex and multifaceted issue, as it involves balancing the need for regulation with the innate desire to protect free speech and competition.

Nowadays, the natural association between government intervention in the activities of digital platforms and infringement of freedom has almost ceased to exist. Instead of seeing platforms as the embodiment of freedom of expression - as in the days of John Perry Barlow - Americans tend to treat them as a threat to their privacy. Also, the gatekeeper role assumed by Big Tech is commonly seen as harmful to fair competition and consumer protection.

These concerns are shared by the majority of Americans. On September 24, 2020, Consumer Reports published a survey titled "Platform Perceptions: Consumer Attitudes on Competition and Fairness in Online Platforms." Among the findings of the survey, it emerged that:

- 85% of Americans are concerned about the amount of data online platforms store about them, and eighty-one percent are concerned that platforms are collecting and holding this data in order to build out more comprehensive consumer profiles;
- 58% are not confident that they are getting objective and unbiased search results when using an online platform to shop or search for information;
- 79% say Big Tech mergers and acquisitions unfairly undermine competition and consumer choice; and

– 60% support more government regulation of online platforms, including mandatory interoperability features, to make it easier for users to switch from one platform to another without losing important data or connections.

As a consequence of their new social perception, digital companies find themselves in a much less comfortable situation than in the 1990-2000s, as it is now much easier for US legislators and regulators to gain the political support of civil society. Although the substance of this new regulation is still taking shape, its trajectory is clear: the US is moving toward a more interventionist approach to digital regulations. In particular, the big question is now whether Congress should create a new industry-specific regulatory framework for digital platforms; in this respect, in 2023 the legislative activity seems to have taken a fast-forward pace.

In May 2023 US Democratic Senators Michael Bennet and Peter Welch introduced a new bill, the Digital Platform Commission Act ("DPCA"), the first-ever legislation in Congress aimed at creating an expert federal agency to provide comprehensive regulation of digital platforms to protect consumers, promote competition, and defend the public interest. Amid raising calls for regulation of artificial intelligence and social media, the senators propose a new Federal Digital Platform Commission with the mandate, jurisdiction, and tools to develop and enforce rules for a sector that has gone virtually unregulated.

Bennet's statements on the purpose of the bill are very clear:

> There's no reason that the biggest tech companies on Earth should face less regulation than Colorado's small businesses – especially as we see technology corrode our democracy and harm our kids' mental health with virtually no oversight. Technology is moving quicker than Congress could ever hope to keep up with. We need an expert federal agency that can stand up for the American people and ensure AI tools and digital platforms operate in the public interest.

According to Bennet and Welch, Congress has failed to keep pace with fast-moving developments in the sector through comprehensive oversight, instead offering narrow, reactive solutions after problems have arisen. The rapid pace of innovation, combined with the technically complex challenges of regulation, suggest this pattern will continue without a fundamentally new approach. In particular, they note that it is not the first time a new sector of the economy has emerged to amass extraordinary and unregulated power. In the past, Congress has already reacted to these developments by creating expert federal agencies empowered to provide timely, thoughtful, and durable regulations; it was the case, for instance, of the Food and Drug Administration, the Federal Communications Commission, or the Federal Aviation Administration. From airplanes to pharmaceuticals to telecommunications, Congress had the

wisdom to delegate oversight and regulation of complex sectors of the economy to dedicated independent agencies, while no such agency currently exists for digital platforms.

In short, the DPCA would: (i) establish a five member Federal Digital Platform Commission empowered to hold hearings, pursue investigations, conduct research, assess fines, and engage in public rule-making to establish rules of the road for digital platforms to promote competition and protect consumers, for example, from addicting design features or harmful algorithmic processes; (ii) empower the Commission to designate "systemically important digital platforms" subject to extra oversight, reporting, and regulation, including requirements for algorithmic accountability, audits, and explainability; (iii) create a Code Council of technical experts and representatives from industry and civil society to offer specific technical standards, behavioral codes, and other policies to the Commission for consideration, like transparency standards for algorithmic processes; and (iv) direct the Commission to support and coordinate with existing antitrust and consumer protection federal bodies to ensure efficient and effective use of federal resources.

Few months after the presentation of the DPCA, in September 2023 another remarkable bill, Digital Consumer Protection Commission Act ("DCPC"), was introduced by Democratic Senator Elizabeth Warren, co-sponsored by Republican Senator Lindsey Graham. The DCPC aims to take on Big Tech with a new dedicated and specialized federal agency empowered to promulgate and enforce new regulations. In many ways, also the Digital Consumer Protection Commission that Warren proposes is similar to the FTC and the Federal Aviation Administration, sector-specific agencies that were created by Congress to regulate new and powerful technologies. Similarities can be found also with the Consumer Financial Protection Bureau, the agency created through the Dodd-Frank Act in the wake of the subprime financial crisis, that Warren is credited with establishing.

A big part of this bill is about competition and antitrust, but it also covers online privacy. According to the bill, platforms cannot target ads to users based on data from third parties, should limit the processing of personal data, and users should have the right to access and delete data they have collected about them. There is also a section about transparency, which requires covered platforms to have clear terms of service, notify users if they have been banned or their content has been otherwise restricted, and provide ways for users to appeal such decisions.

Last but not least, in the bill the DCPC is entrusted with the authority to issue licenses to operate digital companies, and to revoke them. In a recent interview, Warren said: "Banks operate with a license, airlines operate with a license; the same should be true for the giants in Big Tech."

Taken together, the DPCA and the DCPC bills show a growing awareness in Congress that the existing legislation and current agencies are not enough to regulate the digital arena, and that it has come the time to adopt a new approach. Of course, the tech giants will keep fighting their battle for freedom with all the immense ammunition they dispose of, and the outcome of the next Presidential election round in late 2024 will play an important role; however, the road towards a more stringent digital regulatory framework seems to be paved.

3. The US approach to Artificial Intelligence regulation

In 2023, artificial intelligence ("AI") jumped to the headlines thanks to the public launch of generative AI chatbots such as OpenAI's ChatGPT. The emerging discussions on the opportunities and threats of artificial intelligence refuelled the debate on the need of specific regulation.

On 30 October 2023, President Biden issued a landmark Executive Order titled "Safe, Secure, and Trustworthy Artificial Intelligence" ("AI EO") which was labelled as "the most significant actions ever taken by any government to advance the field of AI safety." The AI EO incorporates and expands on elements of the White House's October 2022 "Blueprint for an AI Bill of Rights" and May 2023 "Action Plan to Promote Responsible AI Innovation", in addition to voluntary commitments secured from over a dozen leading technology companies to drive safe, secure, and trustworthy AI development.

The AI EO sets eight guiding policies and principles intended to serve as a roadmap for the industry and regulators alike:

1. AI must be safe and secure by requiring robust, reliable, repeatable and standardized evaluations of AI systems, as well as policies, institutions and appropriate mechanisms to test, understand, and mitigate risks from these systems before they are put to use.
2. The US should promote responsible innovation, competition and collaboration via investments in education, training, R&D and capacity while addressing intellectual property rights questions and stopping unlawful collusion and monopoly over key assets and technologies.
3. The responsible development and use of AI requires a commitment to supporting American workers through education and job training and understanding the impact of AI on the labour force and workers' rights.
4. AI policies must be consistent with the advancement of equity and civil rights.
5. The interests of Americans who increasingly use, interact with, or purchase AI and AI-enabled products in their daily lives must be protected.
6. Americans' privacy and civil liberties must be protected by ensuring that the collection, use and retention of data is lawful, secure and promotes privacy.

7. It is important to manage the risks from the federal government's own use of AI and increase its internal capacity to regulate, govern and support responsible use of AI to deliver better results for Americans.
8. The federal government should lead the way to global societal, economic and technological progress including by engaging with international partners to develop a framework to manage AI risks, unlock AI's potential for good and promote a common approach to shared challenges.

The AI EO is ambitious, as it seeks to strike a balance between promoting AI's tremendous potential while safeguarding against its risks. It is also quite comprehensive, taking an expansive view of AI, defined as "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments". That broad scope should include not only cutting-edge generative AI systems - such as ChatGPT - but also technologies like: web searches, autocorrect tools, content customization and personalization, voice assistants, and facial recognition.

The National Institute of Standards and Technology (NIST) is tasked with a leading role in implementing many of the directives of the AI EO, as well as to lead the development of key AI guidelines. However, the AI EO adopts an "all-of-government approach", tapping agencies and offices across the entire administration to tackle the use of AI technologies in their respective areas of expertise.

The AI EO takes a notably different approach to artificial intelligence regulation compared to the EU AI Act. While the EU AI Act encompasses a wide spectrum of artificial intelligence system sales and usage in the EU, the AI EO strives for a balanced approach, encouraging artificial intelligence adoption while mitigating associated risks. The AI EO refrains from the more prescriptive approach of the EU AI Act, which employs a risk categorization framework with specific requirements, including comprehensive risk management, data governance, accuracy standards, human oversight mandates, and monitoring procedures.

Regardless of the AI EO, other US federal regulators have been active in specific areas. For instance, in July 2023, the US Securities and Exchange Commission proposed a rule aimed at protecting investors from artificial intelligence technologies that could place broker and investment advisory firms' interest ahead of their investors' interest. In 2024 further activity is expected from regulators in targeted areas, while comprehensive legislative initiatives gain direction. The US Congress does not seem likely to take action on artificial intelligence legislation in the near future; however, the individual states have been quite active, with nearly 200 AI-related bills already introduced in state legislatures, and many more are likely to come soon.

One of the areas where several critical legal issues are fast emerging is the interaction between AI and copyright. The main issues relate to potential

infringement and rights of use issues, uncertainty about ownership of AI-generated works, and questions about use of unlicensed content in training data and whether users should be able to prompt these tools with direct reference by name to other creators' copyrighted works without their permission.

These claims are already being litigated. In a case filed in late 2022, Andersen *v.* Stability AI *et al.*, three artists sued multiple generative AI platforms for using their original works without license to train their AI systems, allowing users to generate works that may be insufficiently transformative from their protected works and, as a result, would be unauthorized derivative works. Similar cases filed in 2023 bring claims that companies trained AI tools using data lakes with thousands, if not millions, of unlicensed works. More recently, the lawsuit filed by the New York Times against OpenAI and Microsoft for copyright infringement - received the world's attention - alleging that the companies' AI tools illegally copied millions of New York Times articles and contents to train ChatGPT and other services to provide people with instant access to information.

In each of these cases, the legal system is being asked to clarify the bounds of what is a "derivative work" under intellectual property laws. The outcome of these cases is expected to hinge on the interpretation of the "fair use" doctrine, which allows copyrighted work to be used without the owner's permission for purposes such as criticism (including satire), comment, news reporting, teaching (including multiple copies for classroom use), scholarship or research, and for a transformative use of the copyrighted material in a manner for which it was not intended.

While waiting for the outcome of the courts, also the regulators started to address the matter. In 2023, the US Copyright Office launched an initiative to examine the copyright law and policy issues raised by AI technology, including the scope of copyright in works generated using AI tools and the use of copyrighted materials in AI training. After convening several public listening sessions to gather and share information about current technologies and their impact, the Office published a notice of inquiry that received over 10,000 comments by the December 2023 deadline. In 2024, the Office plans to release a report in several sections analyzing the issues, which will be published as they are completed; this might be the first step towards specific legislative actions in the near future.

4. Data Privacy: the American way and the interactions with the GDPR

The United States legislative activity in the area of personal data protection started relatively late. It is true that privacy protection dates back to the Fourth Amendment of the US Constitution, which was however intended to protect citizens against intrusions by public authorities, not private actors such

as digital platforms. The oldest pieces of privacy legislation (the US Privacy Act of 1974, the Financial Privacy Act of 1978, the Privacy Protection Act of 1980) aim to protect citizens from abuse by government agencies. In the 1970s, with Griswold *v.* Connecticut and Roe *v.* Wade the Federal Supreme Court confirmed the right to privacy as a fundamental right. In contrast to Europe's broader rights-based approach, under which individuals effectively own their personal data and have the legal right to control it, the US historically adopted a harms-prevention-based approach, generally allowing businesses and institutions to collect personal information without express consent, while regulating its use to prevent or mitigate harms in specific sectors; examples are the Health Insurance Portability and Accountability Act (HIPAA) regarding healthcare and health insurance personal data, the Graham-Leach-Blieley Act (GBLA) covering financial personal information, or the Children's Online Privacy Protection Act (COPPA), protecting children personal information.

Recently, the US legal framework has evolved to provide more robust protections and year 2023 marked the beginning of a profound shift in the philosophy underlying data privacy laws in the United States, at least at state level. Indeed, some states have enacted far-reaching data protection laws inspired by the EU General Data Protection Regulation ("GDPR"), setting strict standards for how service providers should manage personal information, providing users with the right to know what personal data is collected about them, the right to access such information and request its deletion.

The first-mover was California, which enacted the California Consumer Privacy Act (CCPA) of 2018 (and new rights effective as of 2023) and the California Privacy Rights Act (CPRA) which provides additional protections. Following California's lead, four other states – Colorado, Connecticut, Utah and Virginia – started enforcing new GDPR -like statutes in 2023, and other states are about to follow.

The diverse approaches adopted by the US and the EU are well highlighted by the different regimes applicable to international data transfers. While there is no federal law in the United States restricting international data transfers, the GDPR regulates cross-border transfers in detail. Transfers of personal information may freely take place from the EU to a third country (such as the US) only if the European Commission has decided that such country - or specified sectors within the third country - ensures an adequate level of protection to data subjects' information. When assessing the adequacy of the level of protection, the Commission shall evaluate: (i) the rule of law (respect for human rights and fundamental freedoms, relevant legislation, both general and sectoral, including concerning public security, defence, national security etc.), (ii) the existence of a supervisory authority, and (iii) international commitments the third country has entered into, or its participation in multilateral or regional systems. Furthermore, the GDPR details the conditions under which such transfers may

take place in the absence of an adequacy decision or appropriate safeguards listed at art. 49 of the GDPR.

The complex interaction between the US and EU regulations on data privacy have been revealed by the European Court of Justice ("ECJ") in two notable cases, the "Schrems I" and "Schrems II" judgments, where the ECJ declared invalid (pursuant to article 7, 8 and 47 of the Charter of Fundamental Rights) the decisions of the European Commission on the adequacy of the previous EU-US data transfer agreements, *i.e.,* the Safe Harbour Agreement of 2000 and the Privacy Shield of 2016. The cases highlighted how certain US provisions do not meet the adequacy requirements set forth by the EU, due to the presence of surveillance programs that legitimize access by US public authorities to personal data transferred from the EU to the US, and to the absence of adequate enforceable rights for data subjects against the US authorities.

In October 2022, President Biden issued the Executive Order on Enhancing Safeguards for United States Signals Intelligence Activities, aimed at addressing the issues put forward by the ECJ and setting the steps to implement the US commitments under the third attempt to regulate the transfer of data between the EU and the US, the Trans-Atlantic Data Privacy Framework UE-USA ("DPF"). The DPF came somewhat unexpectedly, given the previous opinions of the European Data Protection Board and the European Parliament, which were essentially aligned in urging the European Commission not to adopt the new agreement with the United States in the absence of significant corrective measures that would, however, require a substantial reform of the internal US national security system. Notwithstanding that, on 10 July 2023, the European Commission adopted its adequacy decision, concluding that the United States ensure an adequate level of protection - compared to that of the EU - for personal data transferred from the EU to US companies who participate in the DPF by committing to comply with detailed privacy compliance obligations. Accordingly, personal data can now freely flow from the EU to US companies participating in the DPF, without the need of putting in place additional data protection safeguards.

5. Digital legacy regulation: a new frontier

While in certain areas, such as privacy, the US has been lagging behind compared to the EU, it has been a front-runner with respect to regulation of digital legacies. Digital legacies refer to data accumulated by a data subject once she or he ceases to exist. Indeed, everyone willfully or involuntarily generates large amounts of data; indeed, users create, share, and store content on the internet, through platforms such as social media sites, user reviews, online banking, and many others. As the amount of data that each data subject leaves behind has exponentially grown, and due to rapid development of

digital assets with significant monetary value, such as cryptocurrencies and NFTs, the question arises as to what happens to such data and content after an individual ceases to exist.

Digital assets are regulated by laws surrounding data privacy and unauthorized access to computer systems. Terms of service agreements for online services also place restrictions on, or prohibit access to, digital assets by individuals other than the user themselves. In the context of the deceased, both the safeguard mechanisms are still evolving to ensure the right to access digital assets while respecting their privacy and will.

There are two main types of legislation pertaining to digital legacies: data protection laws and succession laws. However, while uploaded data is protected as long as data owners are alive, it may not be the case once they pass away; this is because data protection laws, such as the GDPR, do not apply to data of the deceased, nor do explicitly address its applicability to the deceased.

At present no common agreement has been reached internationally. In particular, the existing instruments which address digital legacies are enacted and implemented at the national level. In terms of establishing a legally binding instrument on the management and succession of digital legacies after death, the US is one of a few countries that has introduced a law regulating the governance of digital legacies.

In 2015, the Revised Uniform Fiduciary Access to Digital Assets Act ("RUFADAA") was drafted by the Uniform Law Commission (ULC); since then, it has been adopted by nearly all the 50 states. RUFADAA gives a clear hierarchy of instructions for how a person's digital assets are to be treated should a fiduciary seek access, which may include not only executors after death, but trustees, court-appointed guardians, and attorneys-in-fact. The starting point is that online service providers can create an "online tool" that functions as a form of "digital power of attorney" to specify who has control and access for that specific site. In addition, RUFADAA provides a clear legal framework for digital asset rights to be specified in traditional legal documents (such as wills and powers of attorney), and clarifies that only in the absence of an online tool, or any legal documents, the service provider's own terms of service will prevail.

Following the enactment of RUFADAA, in the past years all the main digital platforms like Google, Facebook and Apple have created specific tools allowing their users to appoint a legacy contact who would be entitled to manage the accounts and data of the deceased.

However, many aspects concerning the legal ramifications surrounding digital legacies remain to be determined; further regulation is expected to come in the near future, and the US would likely continue to guide this process.

Chapter VI
The Chinese Approach to Information Technology Law

by Eylül Erva Akin*

Index: 1. Introduction. – 2. China's data protection journey. – 3. China's evolving data protection laws. – 4. Comparison of key aspects in Chinese data law. – 5. Classification and hierarchy in data protection. – 6. Conclusion and key points.

1. Introduction

With the growth of digital services worldwide and the increasing generation, collection, and use of data, the need for robust data protection measures has become more pressing than ever before. Organisations across various sectors are grappling with the challenge of safeguarding sensitive information from cyber threats and data breaches. To address this concern, governments and regulatory bodies have implemented stringent data protection laws and regulations. These frameworks aim to ensure that individuals' personal data is handled responsibly and securely by businesses and organisations. For instance, the General Data Protection Regulation (GDPR) in the European Union has set strict guidelines for data protection, giving individuals more control over their personal information.

The People's Republic of China has faced significant questions regarding the data protection. Before the development of laws related to personal information protection, legislation such as the Cybersecurity Law and the Civil Code addressed personal information, however, earlier legislations were either sector-specific or too broad for practical implementation. In recognition of the need to address these challenges, the People's Republic of China has implemented important changes by enacting new laws to better protect individuals' personal information and ensure data security. These efforts have resulted in the passing of two crucial laws in 2021: the Personal Information Protection Law (PIPL) and the Data Security Law (DSL). These laws demonstrate that the China is committed to protecting personal information and ensuring data security in the digital age.

* Lawyer. Doctoral Researcher at the University of Milan. Research Fellow at the Information Society Law Center (ISLC).

2. China's data protection journey

The Chinese data protection regime is not primarily focused on constitutional principles or liberal values. Instead, it is driven by the broader goal of "cybersecurity and informatization". This aims to make China a strong and wealthy nation by incorporating digital technologies into various aspects of life. People's Republic of China's data protection regulations have evolved in response to concerns about the widespread adoption of digital technologies. The development process included addressing digitized government systems, a digital economy led by platform companies, a black market for personal information, and increased vulnerability to data-related issues. However, the legal system of the China does not adhere to several key assumptions found in Western legal systems. These include the principles of legality, limitations on government access to personal data, rights of individuals against the government, and the existence of an independent judiciary (EDPB, 2021).

In 2014, President Xi Jinping emphasised turning China into a cyber power by using digital capabilities for national development goals like economic growth and effective governance (William Wan, 2014). Cybersecurity is crucial in protecting this development sphere from potential threats. Unlike the U.S. and Europe, China's cybersecurity definition goes beyond technical matters, considering potential harm to the country's politics, economy, and society. President Xi sees cybersecurity and informatisation as interconnected and vital for People's Republic of China's progress. According to this argumentation, it has been argued that the PIPL and the DSL align closely with the concept of informatisation and cybersecurity (Creemers, 2022).

Since 2012, there has been a move towards centralising data governance and adopting more comprehensive legislative frameworks. However, obstacles remained in creating a detailed and workable data protection framework. Progress in passing and implementing draft regulations was limited, with only a handful of initiatives, including provisions for children's personal information, being put into practice. However, there were noteworthy strides in technical standards and sectoral self-regulation, offering more comprehensive guidelines for information security and the protection of personal information.

Lastly, it is important to clarify the distinctions between "data", "information", and "personal information" in Chinese legal framework (Cai and Chen, 2022).

In sum:

- "Data" refers to any information record by electronic or other means.
- "Information" includes the knowledge contained within data.
- "Personal information" is data that can identify a specific individual.

It is important to note that Chinese data law uses "personal information" terminology rather than "personal data" found in the EU GDPR, although

they refer to the same concept (Cai and Chen, 2022). The difference between "personal data" and "personal information" can be illustrated by considering that "personal data" typically refers to binary codes of 0 and 1, whereas "personal information" includes any content that can be used to identify individuals. Nonetheless, in practice, both DSL and PIPL address personal data protection issues. Therefore, it could be argued that in Chinese data law, personal data and personal information are synonymous.

3. China's evolving data protection laws

In 2021 DSL and PIPL were enacted for data governance in People's Republic of China. With these developments, it can be argued that Chinese data law has evolved from regulating network infrastructure to encompassing the management of data and information.

The PIPL shares similarities with personal data protection laws worldwide, while the DSL is a notable innovation aimed at averting harm to national security and the public interest through data-enabled methods.

3.1. Personal Information Protection Law (PIPL)

On 20 August 2021, PIPL was adopted as the China's first comprehensive legislation for personal data protection. The PIPL is primarily designed to regulate different categories of actors and their interactions with a high level of detail, considering potential risks. It aims to balance multiple values, including market efficiency, human dignity, fairness, and others. Instead of concentrating on establishing fundamental rights or general principles, its focus lies on managing specific relationships and associated risks. In general, the PIPL has three different objectives:

- protecting individuals from malicious or improper data collection and use, mostly by companies;
- stimulating the development of the digital economy; and
- safeguarding the public interest.

According to the PIPL electronic data itself does not contain personal information but it can be considered as personal information if it can identify a specific natural person.

3.2. Data Security Law (DSL)

The DSL came into effect in China on 1 September 2021. The DSL aims to ensure data security, promote data development, and protect the lawful rights and interests of individuals and organizations, while safeguarding national sovereignty, security, and development interests.

The DSL distinguishes data (form) from information (content) and data is considered as a carrier of information. The DSL applies differentiated security obligations based on the classification of data as "important data" and "core national data", as outlined in Article 21 of the DSL.

3.3. Cybersecurity Law (CSL)

The CSL was enacted in 2016. The CSL regulates "network data" or "electronic data". It is one of the most comprehensive legislations for securing the Chinese digital sphere. Article 1 of the CSL emphasises cyberspace sovereignty, national security, social and public interest to protect the lawful rights and interests of citizens (Rogier Creemers et al., 2018).

4. Comparison of key aspects in Chinese data law

Here is a table comparing key aspects in Chinese data protection and cybersecurity law.

	Data Security Law (DSL)	**Personal Information Protection Law (PIPL)**	**Cybersecurity Law (CSL)**
Year	2021	2021	2016
Purpose	Regulates the security of data and information within networks (Article 1).	Aims to protect personal information and protect the free flow of personal information (Article 1).	Focuses on establishing China's sovereignty in cyberspace (Article 1).
Applicable to…	The construction, operation, maintenance, and use of networks, as well as to cybersecurity supervision and management within the People's Republic of China (Article 2).	All personal information protection handlers, including both governmental and private entities (Article 72).	Network operators, such as network service providers (Article 76)
Regulated Objects	Regulates "network data" or "electronic data". Data recorded in non-electronic formats is also regulated.	Personal information, including data identifying individuals.	Regulates "network data" or "electronic data".

5. Classification and hierarchy in data protection

The data protection system in China includes horizontal data classification and vertical data hierarchy protection.

5.1. Horizontal: data classification

The data is categorised based on common attributes such as the producer, possessor, concerned area, and industry. The aim is to organize the data into different categories. There are primary and secondary categories. Data is the primary category. The secondary category is narrowly defined data (data other than personal information) and personal information.

Data in different areas may require different approaches to protection. There are specific rules for certain industries and subjects, such as scientific data, medical data and automotive network data. These rules complement the general rules set out in the CSL, DSL and PIPL.

5.2. Vertical: data grading

Vertical data hierarchy protection grades data according to the potential harm it poses to national security, social order, public interest, and individual rights. It applies different levels of protection to different hierarchies of data.

The DSL introduces "key data" with more stringent controls. Key data, including national security and the economy, receives the highest protection.

Personal information is divided into two categories based on the proximity of the information to personal interests: general personal information and sensitive personal information. Sensitive personal information might potentially cause harm to the person and property of an individual, once leaked, or used illegally. It is subject to stricter processing rules, notification, and consent requirements.

6. Conclusion

The goal of Chinese data protection and cybersecurity legal framework is to enhance physical and legal security of data. This framework consists of three main laws: the Cybersecurity Law, the Data Security Law, and the Personal Information Protection Law. There are horizontal data classification system and a vertical grading mechanism.

There is a debate about whether these laws should be applied sequentially or simultaneously in practice. The simultaneous application of the CSL, DSL, and PIPL in data-related issues, resulting in confusion and repetition.

It is still a relatively new legal field, and there is room for improvement. There is an overlap in regulatory objects and legal regimes: challenges in both practical application of data law and theoretical conceptualisation such as data, network data, information, and personal information are sharing substantial similarities.

Bibliography

Cai, P., Chen, L.(2022). 'Demystifying data law in China: a unified regime of tomorrow', *International Data Privacy Law*, 12, pp. 75–92. Available at: https://doi.org/10.1093/idpl/ipac004.

Creemers, R. (2022). 'China's emerging data protection framework', *Journal of Cybersecurity*, 8(1). Available at: https://doi.org/10.1093/cybsec/tyac011

Creemers, R. and Webster, G. (2021) n.d. Translation: Personal Information Protection Law of the People's Republic of China. Available at https://digichina.stanford.edu/work/translation-personal-information-protection-law-of-the-peoples-republic-of-china-effective-nov-1-2021/

Creemers, R., Webster, G., Triolo, P. (2018) *Translation: Cybersecurity Law of the People's Republic of China*. Available at: https://digichina.stanford.edu/work/translation-cybersecurity-law-of-the-peoples-republic-of-china-effective-june-1-2017/

Wan, W. (2014) 'Chinese President Xi Jinping takes charge of new cyber effort'. *The Washington Post*, 27 February.

Huang, Y., Mingli, S. (2021). *Top Scholar Zhou Hanhua Illuminates 15 + Years of History Behind China's Personal Information Protection Law*. Available at: https://digichina.stanford.edu/work/top-scholar-zhou-hanhua-illuminates-15-years-of-history-behind-chinas-personal-information-protection-law/

References

EDPB, (2021) Legal study on Government access to data in third countries.

Huang, Y. and Shi, M. (2021) *Top Scholar Zhou Hanhua Illuminates 15+ Years of History Behind China's Personal Information Protection Law* (*DigiChina*, 8 June 2021). Available at https://digichina.stanford.edu/work/top-scholar-zhou-hanhua-illuminates-15-years-of-history-behind-chinas-personal-information-protection-law/

DigiChina (2021) Translation: *Data Security Law of the People's Republic of China* Effective Sept. 1, 2021. Available at: https://digichina.stanford.edu/work/translation-data-security-law-of-the-peoples-republic-of-china/

PART II
ARTIFICIAL INTELLIGENCE

Chapter VII
An Introduction to Artificial Intelligence

by Giovanni Ziccardi*

Index: 1. A phenomenon to be regulated worldwide? – 2. How to regulate AI? The EU approach for a "digital decade". – 3. The "European" approach to artificial intelligence. – 4. Some considerations: a "balanced" approach and the necessary respect for rights. – 5. In conclusion: respect the right and prohibit certain practices.

1. A phenomenon to be regulated worldwide?

As 2024 dawns, there are now dozens of proposals for regulation in a broad sense (regulatory, ethical, deontological, professional, national, international, internal company, political, etc.) concerning artificial intelligence and robotics. Indeed, the last two years have been those of the worldwide "explosion" of the phenomenon and of peculiar attention on the part of all states. This implies that these issues have overwhelmingly become the nodes of the present and the (near) future, with an unprecedented speed of evolution and with a consequent difficulty on the part of the legal world (notoriously slow and bureaucratic) in addressing the issue.

All the states are investing time and resources trying to understand whether and how to design a legal framework that can allow the economy to develop around these innovative technologies.

Opening the dance, was the European Union that on 21 April 2021 issued a proposal for a regulation of the European Parliament and the Council aimed at establishing harmonized rules on artificial intelligence.

This crucial proposal, which has been attracting the attention of scholars around the world for years, was - after three years of active discussions - finalized. The 12th of July 2024 was published the final text in the Official Gazette of the European Union (in force from the 1st of August 2024).

In the United States of America, too, there are now dozens of regulations (or proposed regulations) expressing interest in the topic and that in some way touch upon artificial intelligence.

In Alabama, for example, an "Alabama Council on Advanced Technology and Artificial Intelligence" has been established to advise the governor and the legislature on the use and development of such technologies.

---

* Full Professor of Legal Informatics at the University of Milan and Director of the Information Society Law Center (ISLC).

In California, an "Automated Decision Systems Accountability Act" has been proposed to minimize the discriminatory risks and impact of such technology, especially when making fully automated decisions.

In Illinois, the "Artificial Intelligence Video Interview Act" would address the use of artificial intelligence in gathering information from citizens, again to avoid discrimination.

In Mississippi, on the other hand, work has been done on introducing artificial intelligence, robotics, and machine learning into traditional computer science curricula in vocational and school curricula.

In Nevada, the creation of an "Emerging Technologies Task Force" has also put artificial intelligence at the centre of specific projects; for example, Vermont state is planning to establish an "Artificial Intelligence Commission".

Interesting is the strategy of the State of Utah, which has created a specific policy on the so-called "deep technologies", *i.e.*, technologies that lead to new products and discoveries.

At the federal level, on October 2023 President Biden issued an executive order designed not only to provide some essential guidelines but also to "steer" regulation in the coming years.

Canada launched a strategy called the "Pan-Canadian Artificial Intelligence Strategy" in 2017 to create research and jobs, while China launched its own project in 2017, "A Next Generation Artificial Intelligence Development Plan" and published ethical guidelines for its use in 2021.

Hence, the whole world is suddenly regulating artificial intelligence. Each State, or " part of the world", is doing so with different approaches: some believe in law and regulations, others prefer, on the contrary, not to regulate and leave the industry as free as possible (fearing that too many strings and ties could slow down or hinder the development of this crucial technology for the future of the economy of all countries), while others are relying on self-regulation by companies (trusting, with a certain optimism, in an ethical approach coming from within the company itself).

Certainly, the European Union's approach for a regulation on artificial intelligence is the most articulate, advanced and interesting and it will guide us in the pages that follow to envisage a regulation of artificial intelligence that will likely reconcile respect for business and the centrality of people's rights and dignity.

2. How to regulate AI? The EU approach for a "digital decade"

The European Union - announcing its policy and regulatory strategy on digital for the next 30 years - has started to design a technological and regulatory

future designed to bring concrete benefits to citizens, users and businesses. The most ambitious goal is to achieve a digital and climate-neutral Europe by 2050.

The actual decade has been identified not only as the "decade of artificial intelligence" but also as the "European digital decade": the time has finally come for the European Union to try to strengthen its digital sovereignty and to establish internal rules on all the most important topics, instead of being exclusively conditioned by the regulations (and decisions) of non-European companies or standards. Currently, only circa a dozen companies are developing artificial intelligence systems with a global impact and they are all North American or Chinese companies.

The three areas of operation for this ambitious action of (re)building digital Europe will concern: data, technologies and infrastructures.

Four regulatory measures are already facilitating the emergence of a particularly agile, uniform and data-driven European economy: the regulation on the free movement of non-personal data, the regulation on cybersecurity, the open data directive and the general data protection regulation.

Alongside these regulations, the announced European Artificial Intelligence Strategy should lead to a reliable and ethical European artificial intelligence (also referred to, rather suggestively, as "anthropocentric"), as well as to a use of big data designed to create wealth for the benefit of society and businesses.

All this, of course, will require new "digital skills", *i.e.*, very specific (and more advanced) digital skills, and a new level of connectivity for European citizens. Two elements, by no means, to be taken for granted.

3. The "European" approach to artificial intelligence

Central to the European Union's regulatory vision for the digital decade will be the development of reliable and trustworthy artificial intelligence, conceived as a tool that can bring many benefits (better health care, safer and cleaner transportation, more efficient production processes and cheaper and more sustainable energy) but that places the rights, freedoms and, above all, the dignity of human beings at the center.

The concrete hope is that citizens will reach a good level of "trust" in the use of such technologies and that businesses will be facilitated in the process of developing them, while paying great attention to the quality of the equipment used in the development of such systems.

The proposal for a regulation on a European approach to artificial intelligence came about by envisaging a series of actions designed to achieve these objectives, designing to this end the first homogeneous, comprehensive legal framework on artificial intelligence ever to be created in such detail. Such framework should lead to the emergence (or introduction into Europe) of

artificial intelligence systems that are safe, transparent, ethical, impartial and always operating under human control for supervision purposes.

The EU's regulatory choice was to classify artificial intelligence tools according to a risk assessment; hence, the same approach used in the GDPR.

Artificial intelligence systems posing an unacceptable risk, *i.e.*, capable of generating a clear threat to EU citizens, will be banned. Indeed, the legislator has in mind the examples of the social credit system used by authoritarian governments or certain "smart" toys with built-in voice assistants that may encourage dangerous behavior by users, especially children.

On the other hand, high-risk artificial intelligence systems will be permitted provided the compliance with stringent rules and requirements for those who intend to develop and place them on the European market.

Similar rules will have to be applied to artificial intelligence used in the management of critical infrastructures (*e.g.*, the transport system) that could pose lives and health of citizens at risk, in education or vocational training, conditioning access to education and people's professional lives (*e.g.*, in calculating exam scores), in the safety components of products (*e.g.*, robotic surgery), in the labour market (*e.g.*, in CV selection software during recruitment procedures), in essential public and private services (*e.g.*, social credit systems that deny people the opportunity to obtain a loan or mortgage), in law enforcement activities that may interfere with people's fundamental rights (*e.g.*, assessment of the reliability of evidence in court), in the management of migration flows, political asylum and border control (*e.g.*, automated verification of the authenticity of travel documents) and in the administration of justice and democratic processes.

Artificial intelligence with limited risk will be subject to minimum transparency requirements aimed at enabling those interacting with such devices to make informed decisions. The user will be able to decide whether to continue interacting with a machine instead of a human being.

Finally, minimal risk systems include the use of applications such as video games or spam filters that exploit artificial intelligence. Most artificial intelligence systems fall into this category in which the new rules tend not to intervene, as such systems generate only a minimal (or no) risk for the rights and security of citizens.

The measures foreseen in the regulation to mitigate the risks mentioned above are very comprehensive:

- adequate risk assessment and systems to mitigate possible damage or incidents;
- high quality of datasets "fed" to the systems to minimize risks and possible discriminatory results;
- constant logging of activities to ensure traceability of results;

- detailed documentation providing all necessary information about the system and its purposes, especially to allow authorities to access such information;
- clear and adequate information to users;
- appropriate human supervision to minimize risks; and
- a high level of security and reliability of the systems.

4. Some considerations: a "balanced" approach and the necessary respect for rights

As mentioned earlier, underlying the EU action on artificial intelligence is the idea of a regulation establishing five fundamental points:

1. The first involves the creation of harmonized rules for the placing on the market, commissioning and use of artificial intelligence systems in the European Union.
2. The second involves the prohibition of certain artificial intelligence practices.
3. The third is the listing of specific requirements for high-risk artificial intelligence systems and related obligations for operators of such systems.
4. The fourth presents harmonized transparency rules for artificial intelligence systems intended to interact with natural persons, emotion recognition systems, biometric categorization systems and artificial intelligence systems used to generate or manipulate images or audio or video content.
5. Finally, the fifth is relevant to design of a strict framework of rules on market monitoring and surveillance.

In short, underlying the Regulation is the idea that artificial intelligence - understood in its broadest sense and for regulatory purposes - consists of a family of very rapidly evolving technologies capable of delivering a wide range of economic and social benefits across the spectrum of industrial and social activities and of improving forecasting, optimizing operations and resource allocation and personalizing service delivery. Thus, such a technology would be able to contribute to socially and environmentally beneficial outcomes as well as provide key competitive advantages to European businesses and the European economy. The drafters of the Regulation consider the use of such technology attractive in high-impact sectors such as climate change, environment and health, the public sector, finance, mobility, home affairs and agriculture. However, those same elements and techniques that would feed into the socio-economic benefits of artificial intelligence could bring new risks or negative consequences for individuals or society.

This calls for a balanced approach: on the one hand, there is the firm political will to build European leadership on the issue. On the other hand, it must be ensured that all systems operate in accordance with the values, fundamental rights and principles of the European Union.

The regulation aims to develop an ecosystem of trust by proposing a legal framework for trustworthy artificial intelligence based on the fundamental values and rights of the Union and giving people and users the confidence to adopt such solutions, while encouraging businesses to develop them.

In this perspective, artificial intelligence should be considered a valuable tool for people and the overall society, with the goal of improving human well-being.

To this end, there is a need to identify those artificial intelligences that are high-risk and to address problems that are not easy, nor immediate, to solve, such as the opacity, complexity, bias, a certain degree of unpredictability and partially autonomous behavior of certain artificial intelligence systems.

Against a similar backdrop, the Commission has presented the regulatory framework on artificial intelligence with some very ambitious specific objectives. The first is to ensure that systems placed on the EU market are secure and respect existing fundamental rights legislation as well as European values.

Furthermore, legal certainty is needed to facilitate investment and innovation in artificial intelligence. At the same time, it is essential to improve governance and effective enforcement of existing legislation on fundamental rights and security requirements applicable to artificial intelligence systems. In addition, there is a need to facilitate the development of a single market for lawful, secure, and reliable artificial intelligence applications and to prevent market fragmentation.

To achieve these objectives, the Regulation presents a balanced and proportionate regulatory approach, which is limited to identifying the minimum requirements necessary to address the risks and issues, without unduly restricting or hampering technological development or otherwise disproportionately increasing the cost of bringing such solutions to the market.

It establishes harmonized rules for the development, placing on the market and use of artificial intelligence systems in the European Union following a proportionate risk-based approach and proposes a single definition of artificial intelligence adapted to future needs. Certain particularly harmful practices are prohibited as contrary to the values of the Union, while restrictions, and specific safeguards, are proposed in relation to certain uses of biometric remote identification systems for law enforcement purposes.

The Regulation establishes a robust risk management methodology used to define "high-risk" systems that pose significant risks to the health and safety or fundamental rights of individuals.

Such systems will have to comply with a set of mandatory requirements for reliable artificial intelligence as well as follow conformity assessment procedures before they can be placed on the EU market. Predictable, proportionate, and

clear obligations are also placed on suppliers and users of such systems with the aim of ensuring security and compliance with existing legislation protecting fundamental rights throughout the life cycle of the systems.

For specific systems, only minimum transparency obligations are proposed, when chatbots or deep fakes are used.

The rules will be implemented through a governance system at Member State level, based on already existing structures, and a cooperation mechanism at EU level with the establishment of a European Artificial Intelligence Board.

Additional measures are also proposed to support innovation, in particular through regulatory testing grounds for artificial intelligence and other measures to reduce regulatory burdens and support small and medium-sized enterprises and start-ups.

5. In conclusion: respect the right and prohibit certain practices

The use of artificial intelligence with its specific characteristics (*e.g.*, opacity, complexity, data dependency, autonomous behavior) may adversely affect a number of fundamental rights enshrined in the Charter of Fundamental Rights of the European Union.

Ex-ante testing, risk management and human oversight obligations are specifically designed to facilitate the respect of other fundamental rights by minimizing the risk of erroneous or biased decisions in critical areas such as education and training, employment, important services, law enforcement and the judiciary.

If violations of fundamental rights do occur, however, effective redress for victims will be made possible by ensuring the transparency and traceability of systems together with strict ex-post controls.

In conclusion, the Regulation aims to ensure a high level of protection of these fundamental rights and to address various sources of risk through a clearly defined risk-based approach. By defining a set of requirements for reliable artificial intelligence and proportionate obligations for all participants in the value chain, it enhances and promotes the protection of rights protected by the Charter: the right to human dignity, respect for private life and protection of personal data, non-discrimination and equality between women and men.

The Regulation positively affects the rights of several special groups, such as the rights of workers (to just and fair working conditions), a high level of consumer protection, the rights of children and the inclusion of persons with disabilities. Also relevant is the right to a high level of environmental protection and improvement of the quality of the environment, also in relation to the health and safety of persons.

No one can predict, today, the effect the AI Act will have on the European economy in the near future, nor the reactions of the large non-European manufacturers of artificial intelligence systems to such a regulatory stance on the Old Continent.

Certainly, it is one of the most interesting regulatory experiments that have ever been introduced to try to regulate the most sophisticated and powerful technology ever developed by human beings.

Chapter VIII
Artificial Intelligence, Data Protection and Responsibilities

by Maria Grazia Peluso*

Index: 1. AI and GDPR. – 2. Profiling and automated decisions. – 3. Artificial intelligence and liability. – 4. The risk-based approach: the AI Act and the proposed AI liability directive.

1. AI and GDPR

When we talk about artificial intelligence, we do not refer to a single paradigm, but to different systems that are based on various techniques. Chat GPT, driverless cars, predictive diagnostics, robot judges, are just a few examples of the application of AI.

The most promising training techniques today are based on the use of large amounts of data. These are used to train algorithms that have a complex structure and are able to recognise links between variables, which form the basis of the outputs of the system.

Today's most promising AI algorithms are so complex that they have been called black boxes.

The expression black box society (F. Pasquale, 2015), coined by Frank Pasquale, refers to the spread of algorithms of AI that are difficult to understand even for their own programmers. The use of a huge number of variables makes it complex to provide an explanation of why the system has produced the output.

Think, for instance, of Google's algorithm. Although the algorithm is roughly based on the swarm theory, it remains secret and difficult to understand in all its parts, even by its programmers.

However, what all the AI systems have in common is the need and use for a huge amount of data.

The great success that AI algorithms have achieved in recent years is in fact linked to the great availability of data. Today data is playing an increasingly important role in the world economy.

---

* Lawyer and Post-doc Researcher at University of Milan. Key staff of the Information Society Law Center (ISLC).

We speak of a "data economy" and data is considered the new oil: data are the fuel of artificial intelligence.

The widespread use of AI systems today is therefore closely linked to the great availability of user data, which is increasingly growing due to the development of the Internet and the spread of the Internet of Things (IoT).

When we speak of IoT, we are referring to objects that possess an internet connection, which is necessary to store and exchange data collected via sensors and to be able to access online content. We are talking about objects in common use, smartphones, tablets, but also refrigerators, alarm clocks, watches, thermostats, etc., that are able to collect any types of data.

These technologies are now so widespread that we hear more and more about the so-called "Internet of Everything". In this respect, it is estimated that there are more than twenty billion objects connected to each other. Around 163 zettabytes of data will be produced by 2025 and one zettabyte corresponds to 10^{12} gigabytes.

As mentioned, artificial intelligence systems need large amounts of data in order to train models *i.e.,* "Big Data". Such term refers to sets of data characterised by three features, known as the "three Vs":

1. volume: a large number of different devices generate a huge amount of data.
2. speed: data are generated and available practically immediately and almost continuously.
3. variety: due to the spread of various devices generating data, their formats are heterogeneous (video, audio, text, code, etc.).

An example of application that can process large amounts of data using algorithms is GoogleFlu Trends (GFT). GFT was a Google service that predicts the spread of influenza cases by analysing the searches on Google.

In 2009, a new flu strain derived from the avian flu was identified. Its high virulence made it necessary to investigate the dynamics of its circulation and the possibility of a pandemic to occur. The analyses carried out up to that point were conducted thanks to the data collected by the Centers for Disease Control and Prevention, a US Agency that is part of the Department of Health and Human Services, on the diagnoses made by healthcare personnel. This methodology, requiring a long time to collect data could not have taken into account several variables including the mobility of citizens. Indeed, Google used GFT to predict the spread of the flu and the results were accurate.

However, the volume and heterogeneity of the data generated and collected makes it difficult to verify its quality, both in terms of the accuracy of the information and the removal of possible bias.

Issues arise when the collected information serves as a dataset for training algorithms. Their ability to find links and statistical recurrences could in fact lead to poor quality results, according to a garbage-in-garbage-out effect.

As long as the outputs obtained are correct, there is no problem. However, when the system is wrong or generates discriminatory output, an explanation of the output leads to different problems: specifically, who should be held responsible for damages.

Moreover, the collection and use of a large amount of data and the ability of artificial intelligence algorithms highlight some critical aspects and some possible conflicts also with the provisions of EU Regulation 2016/679 (GDPR), in particular with the cardinal principles.

1.1. The purpose limitation principle

The provisions of GDPR show how the entire legal framework does not seem to be fully compatible with the processing carried out by artificial intelligence applications.

Particular attention should be paid to the principles of limitation and minimisation, aimed at regulating the initial moment of processing, *i.e.,* the moment of selection and collection of personal data. These principles, together with the principle of storage limitation, do not seem to be in line with the way in which today's technologies operate.

The purpose limitation principle provided by Article 5(b) require data to be collected "for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes".

Recital 39 further clarifies that these purposes must be explicitly stated at the time of collection, so that the data subject can know the actual use of his or her data.

With the increasingly pervasive digitisation of every aspect of society, which was followed by the availability of ever greater quantities of data, certain tensions emerged in relation to the processing methods. In fact, data analysis techniques making the data usable for purposes other than those declared at the time of collection. Provision for the "compatibility" with the purposes initially declared therefore began to take on greater centrality.

The "compatibility" parameter introduced in the Directive 95/46/CE and then transfused into the GDPR, while on the one hand permitting an extensive application of the mentioned principle, on the other hand has given rise to certain critical points linked to its rather broad formulation.

Modern data processing tends to collect and store all the data produced, not making a choice of what data should be used or not. The real wealth of data lies in their versatility of use, since the data controller cannot always foresee what purposes they may be used for.

Since the same dataset can be used to achieve different purposes, changing the analysis algorithms by means of which correlations between variables can be discovered, particular difficulties arise with regard to the information that is inferred from the collected data. The purposes here are to be defined during

the course of the processing itself, as it is not possible to foresee at the time of collection what information may be inferred and, consequently, how it will be used.

For instance, there are business models aimed at collecting data to be resold to companies, which clearly have purposes of which the data controller who originated the first processing is unaware at the time of collection.

The concerted risk is therefore the gradual loss of control of one's own data, a right that is recognised as a fundamental human right and of which the GDPR should be the ultimate expression of guarantee.

1.2. The principles of minimisation and limitation of storage

Similar considerations can be made with regard to the principle of data minimisation, laid down in Article 5(c), which is linked to the principle of storage limitation.

The GDPR rules that the data collected must be adequate, relevant and limited (including in terms of storage time) to what is strictly necessary to pursue the stated purposes.

The need of relevance (and adequacy) refers to the necessary presence of a link between the data collected and the purposes pursued, since this must exist throughout the processing. The provision of a limitation requires that the scope of the operations be restricted to only those personal data that are indispensable, with anonymised or pseudonymised data being the first recourse, where possible.

If, therefore, the use of data processed in this way is likely to limit the impact in the sphere of the individual, it has been argued that the principle of relevance imposes a real obligation to process pseudonymised or anonymised data.

Artificial intelligence systems are based, at least machine learning, on a massive collection and use of data, without either quantitative or qualitative selection; this is because it is not possible to predict what correlations the system will find, it is not possible to operate an *a priori* limitation of the inputs provided to the machine. The richness of data-driven systems lies in their ability to bring out correlations from different data, finding statistical recurrences that a human analyst would be unlikely to find.

In light of these considerations, the principle of minimisation should be widely applied in the subsequent stages of processing. This, in fact, should operate throughout the process and, in particular, in the phases following collection, which also include the transfer of data to third parties. In this phase it would assume a relevant role. The data controller should in fact limit the transmission only to data that are strictly necessary for the individual further processing. In this perspective, the principle of minimisation could then assume centrality in the control of artificial intelligence systems, which are characterised by a complex and long chain of processing.

These considerations can certainly be shared for certain processing methods, but it seems difficult to assume a generic application. As we have seen, the data controller cannot always make a selection of data, since he cannot identify *a priori* those that are aimed at achieving the scope of the analysis. These difficulties are primarily related to the variety of data sources, as well as to their quantity. The composition of big data makes it possible for patterns and statistical correlations to emerge, without it being possible to identify specifically which data have more weight than others.

For example, one can think of algorithms designed to create a credit score that measures the creditworthiness of individuals. These algorithms are commonly employed in credit companies. The data that are processed and contribute to "creating" this score are varied, but the process remains monitored due to their sensitive nature. However, these scores are treated as data and used together with other information to form the so-called buckets within which to categorize individuals in order to predict their behavior for a variety of purposes.

For instance, a 2015 study showed how insurance companies used credit scores as one of the direct parameters to determine car insurance premiums, rather than driving quality data. In Florida, adults with no credit report on their driver's license, but low creditworthiness, pay a higher insurance premium than drivers with excellent creditworthiness but a conviction for driving under the influence (C. O'Neil, 2017).

To complete the picture, a particular connection is made with the principle of data retention limitation. The European legislator, in Article 5(e), requires personal data to be kept for the time strictly limited to achieving the purposes underlying the processing. On this point, the same Recital no. 39 specifies how the controller must establish a precise deadline for erasure or for periodic verification of the continued necessity of the data for the purposes pursued.

The difficulties, mentioned above, regarding the definition of the purposes to be pursued and the principles of minimisation and limitation of storage show an ever-increasing tension with artificial intelligence techniques. Even assuming that it is possible to determine *a priori* the purposes for which big data should be collected and analysed, it is in fact not always easy to estimate a time limit for the use of datasets. These data are often subject to re-use by third parties, as well as in the training of artificial intelligence algorithms. As we have seen, by changing the parameters of the analysis algorithms, previously hidden correlations can emerge and new richness can be extracted from them.

In the light of the operational practices of the digital market, the provision of periodic audits appears worthy of a broader declination within the regulatory dictate. In fact, the rapid evolution of the technologies needs an accurate check of the compatibility and effectiveness of the tools introduced by the GDPR;

a need that emerges especially for those processing operations that have put in place pseudonymisation and anonymisation measures.

2. Profiling and automated decisions

The possibility of an AI being able to take certain decisions with an impact on users in a completely autonomous manner, without human supervision, is not so far off today; on the contrary, the AI applications make it possible to limit, and in some cases even marginalise, human input. We can think of the algorithms used to trade in financial markets without any intervention by brokers. However, the use of digital technologies is not limited to the financial sector, but is now widespread and affects the entire world economy; whether it is the use of algorithms in marketing, the medical sector or the automotive industry, we are increasingly witnessing a digital transformation of the society.

The advantages of fully automated decision-making are certainly considerable. AI systems are, in fact, cheap, quick, reduce bureaucracy, their decisions appear predictable and certain, and, finally, they are considered impartial. However, there are also certain risks associated with the use of these algorithmic systems. First of all, algorithms are hardly designed to be discriminatory, but they can be so in practice. Indeed, applications based on data analysis are strongly influenced by them. The dataset used for training is an expression of a reality, which is not unbiased; it follows that the outputs of the system may be biased and thus be erroneous or discriminatory.

Processing using data-driven technologies has a very strong impact on citizens' privacy. We can think about the ability of algorithms to bring out personal data even from non-personal datasets, re-identifying data rendered anonymous. It is clear that these algorithms could generate a serious risk for the privacy.

Third parties, in fact, could have access, even for illicit purposes, to the private sphere of users, including "sensitive" data, which taken as a whole represent a manifestation of their personality in the digital ecosystem.

If it is true that, apart from cases of external attacks, it is the users themselves who freely provide their data on the web. Nevertheless, in common thought the individual user chooses to provide his or her data only for the purpose of being able to make use of a service, not for any subsequent process. The same data are collected and analysed, often to build up a user profile directed at third parties, but the data subject is hardly aware of this.

The risks become more evident when data is collected and analysed in processing operations involving fully automated decisions. In these cases, there is a greater need to protect the data subjects unaware of such processes as they can only see the outcome, without understanding the underlying reasons.

A specific regulatory provision is provided by Art. 22 of the GDPR, which provides for certain forms of protection in the event that the data subject is subjected to decisions based solely on automated processing, including profiling.

The word "profiling", as statued by Article 4 of the GDPR, means: "any form of automated processing of personal data consisting of the use of such personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects of that person's professional performance, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements". The general principles laid down in the GDPR must apply to profiling, including the grounds of legitimacy set out in Article 6; these are accompanied by some specific indications concerning both information obligations (Articles 13 and 14) and the right to object under Article 21.

In contrast, different considerations must be made with regard to the automated decisions provided for in Article 22. The expression "automated decision making" indicates, more generally, the process carried out by an algorithmic system capable of autonomously inferring from a dataset a decision relevant to the data subjects - thus without the actual participation of a human being. It is therefore clear that profiling could well fall within the category of automated decisions, if the creation of a profile is followed by a decision based on it. All applications that make suggestions to their users are based on a profiling of the person, with the aim of creating clusters of similar profiles. Suggestions are based on a statistical recurrence that, to simplify, follow this rule: since other users liked this additional product, you (who have a similar profile) may also be interested in the same purchase, so I recommend it to you. If the suggestion turns out to be right, and the recommended product is purchased, the algorithm will have a confirmation that it has made a correct prediction, proposing the same content to other users.

There are algorithms that estimate the price and propose to each user what they think he or she is willing to spend. Social networks are also based on user profiling. This is actually their main activity, as it represents the platforms' real source of income. The profiles created are, in fact, the basis of targeted advertising; it is precisely the collection of personal data, necessary for the creation of profiles, that is the consideration for the "free" use of the services rendered.

Profiling techniques are becoming more widespread in the marketing sector. The possibility of "predicting", although one could often speak of "creating", the needs of users clearly bring a great advantage to companies, which can propose advertising targeted on the needs of customers.

This type of processing, however, is not limited to marketing, the use that is made of profiles is very varied; one thinks of filter bubbles (E. Pariser, 2011). These bubbles are the result of a profiling of individuals, which is followed by

the proposition to them only of content selected on the basis of the findings of these profiles, in a sort of confirmatory bias.

Nothing prohibits one from seeking other sources, different opinions, so as to be able to "break out" of the filtering bubble in which one finds oneself. But to do so individuals should have a certain maturity, which is ill-suited to certain group of people, including teenagers and young-adult, as the spread of fake news shows.

In addition, there is the risk of exclusion and discrimination resulting from an inaccurate profile, to which those concerned are unlikely to have full access in the face of the widespread opacity of the algorithms' operation.

There are several experiences of erroneous or discriminatory algorithmic decisions that have had a serious impact on those concerned. For instance, COMPAS case, where an African-American citizen was sentenced to six years imprisonment on the basis of the results given by a software that classified him as a high risk of reoffending.

In 2013 Eric Loomis was stopped by the police while driving a car used in a shooting. The man was sentenced to six years' imprisonment as the judges, through the use of COMPAS software, deemed the defendant to be a high risk of reoffending. The software, whose source code is covered by secrecy, promised, in fact, to estimate the risk of reoffending by inferring from a series of data including the defendant's criminal record and his answers to a 137-question questionnaire.

Loomis' lawyers argued that the judge had violated the defendant's right to know the reasons behind the sentence, since he was prevented from knowing the logic behind the prediction made by the software, which covered a relevant part of the motivation. However, these arguments were rejected by the Supreme Court, which upheld the first instance judgment and held that the use of the software in the case did not violate due process rights. The Court found that the judges' reasoning was correct, as they clarified that the COMPAS scores was only one of the elements considered in the judgment, which was therefore also supported by other independent factors.

2.1. The right to an explanation

Profiling and automated decision-making processes could also bring a risk for individuals of being deprived of the right to construct and control their own social image.

Consequently, the European legislator has provided the GDPR with several requirements necessary to bring automated processing into compliance with the law, including: specific transparency and fairness requirements; increased accountability obligations; indications of the specific legal basis of the processing; guarantees for individuals, including the provision of a right to object

to profiling; and, finally, the indication of the obligation of a data protection impact assessment under certain conditions.

Article 22, headed "automated decision-making relating to natural persons, including profiling", states that: "The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her".

The rule thus covers all decisions in which the preparatory phase, such as the evaluation of the evidence, and the discretionary judgements are made autonomously by the machine; regardless, therefore, of the presence of a human being who formally, but not in substance, makes the final decision, perhaps limiting himself to conforming to the outcome worked out by the system.

Think, for instance, of a person who is in financial difficulty. This individual could be subjected to a profiling process, as a result of which he or she is regularly targeted with online gambling advertisements. This type of advertisement could encourage him or her to sign up for the service and thus ending up worsening his or her financial situation. Or think of the aforementioned differentiated pricing algorithms (so-called dynamic pricing) that could significantly affect those concerned if prohibitive prices prevented access to goods or services.

Article 22 GDPR provides that the data controller, in cases where the processing is based on explicit consent or is necessary for the conclusion of a contract, must take appropriate measures to protect the rights and freedoms of data subjects. Furthermore, the intervention of a human in the decision-making process (human in the loop), the right to express one's opinion and also to contest the automated decision.

It seems clear that in order to make the rights of the data subjects effective, an understanding of the reason of the output given by the system is required. If it is not possible to understand the rationale behind the specific output, it would be difficult for the data subject to exercise his or her rights under the GDPR, such as the right to challenge the output.

In the light of these considerations, a heated debate has opened up on the existence of a so-called right to an explanation for those subject to automated decisions (S. Wachter, B. Mittelstadt, L. Floridi, 2017; M. Brkan, 2019).

The debate takes into consideration the text of the Recital n. 71, which specifies that the data subject has the right to "obtain an explanation of the decision reached after such assessment and to contest the decision" (B. Goodman, S. Flaxman, 2017).

It is interesting to note that in the text of the proposal for a regulation a right to an explanation was in fact expressly included in the legal text; this provision was then expunged and left only in Recital n. 71. It was a precise choice of the legislator not to provide for such a delineated obligation, perhaps also on

account of the difficulties for the data controller in providing an explanation in each individual case.

Thus, looking only at the literal wording of the provisions, the "specific" information required would indeed seem to refer to a set of indications basically aimed at specifying how the automated process operates in general, and how it may affect the rights and freedoms of data subjects. This kind of information may in some cases also be sufficient to explain the output of the system; this is particularly true for all those algorithms that are based on a linear and deterministic logic. In these, once the general operation, usually of an IF/THEN nature, has been clarified, the reason of the final result will then also be clear.

The discourse is different for those non-deterministic algorithms. The deep learning techniques are indeed particularly complex, because they are affected by a large number of data. In these cases, the general operation does not allow the rationale behind individual decisions to be understood. This would therefore result in a *de facto* lack of tools to protect the rights of individuals, since a general explanation would not be useful in order to effectively challenge a specific result.

For two authoritative commentators, the GDPR provides for a right to the "readability" of data and algorithms (G. Malgieri, G. Comandè, 2017). The obligation to provide meaningful information would therefore have as its object all the information necessary to make the functioning of the algorithm comprehensible and at the same time allow transparency of the automated process for the data subjects.

In fact, the information enabling the readability of the decision would not coincide with what is necessary to explain in detail the outputs, nor would it oblige holders to disclose the operating techniques of the algorithms potentially covered by secrecy; is required to provide would only be that information enabling data subjects to exercise their rights under the GDPR.

These reflections are part of a broader theme related to the opacity of algorithms in general.

Some of the experts have, in fact, pointed out that even if a right to an explanation existed, it would in any case not be useful, since an explanation of how the algorithm works would prove too complex to understand. Alongside this remark is also the consideration of the technical impossibility of deriving an explanation, in the meaning given to the term by the GDPR. In fact, the use of algorithms with a complex architecture (black boxes), from which it is not necessarily possible to fully trace back a true explanation of the output generated, is becoming increasingly pervasive.

In order to make the provision of the regulation really effective, it would be preferable to use explainable algorithms (white boxes), so that it is really possible for data subjects, but also for data controllers, to understand the reasons that led the machine to a particular solution.

3. Artificial intelligence and liability

One of the most critical issues arising from the use of artificial intelligence systems is the identification of the person who must be liable for damages caused to property or persons.

This is a particularly sensitive issue, not only for those who have been harmed by artificial intelligence applications, but also for the manufacturers.

An overly rigid liability regime, and a possible very high compensation amounts for damages, could in fact discourage the development and spread of algorithmic systems.

Civil liability is a complex matter, whose regulation is, moreover, left to individual States, which adds further complexity to an already particularly uncertain picture.

What would happen if the individual EU member States were left with free and completely autonomous regulation of AI? Most probably, some States would decide to adopt particularly strict liability criteria, thereby discouraging any producers from developing and marketing products in those territories. Clearly, this would cause serious damage to the single market, generating an unacceptable disparity between citizens within the European Union, with even the risk of the digital divide widening.

Most systems of EU States provide for a general criterion for imputation of liability that is based on the fault of the agent. For this criterion a person is liable if by his/her culpable conduct, *i.e.*, negligent or reckless or careless, she/he has caused damage to a third party.

There are also special liability regimes which regulate specific situations, such as the strict liability regime for damage resulting from dangerous activities, discussed later on.

If, therefore, civil law has so far been able to regulate the phenomena that have emerged with the development of technology, a particular concern has arisen with the spread of artificial intelligence systems with an increasing capacity for autonomous learning.

It is precisely the so-called black box algorithms that also generate tensions with the classical liability regime (based on the fault of the agent).

A few questions arise:

1. if systems are autonomous and not even those who designed and trained them are able to understand how they reached a specific result, are they responsible for the output?
2. Would it be fair to hold them responsible for an output which they not only do not fully understand, but could not even foresee?

   This last question is particularly important. The function of the liability is in fact not only to compensate for damage suffered, which is its main function, but also to discourage conduct that causes damage. There are

also systems that provide for actual punitive damages, so as to discourage certain conduct with particularly high compensations.

With this function of liability in mind, we can understand the difficulty of its application in the case of AI. Indeed, not only would it not be fair to hold the programmer or producer liable, but neither would it serve to discourage a behavior deemed harmful to society.

The European Parliament is also aware of this concern, which already some years ago invited scholars to reflect on it.

In 2017, an important document was published: Resolution on Artificial Intelligence 2015/2103(INL). This document highlighted the critical nature of the AI technologies. Once doubts about the compatibility of the classical liability regime were highlighted, the committee wondered whether autonomous machines should be granted a legal personality, called electronic personality, and thus proceed to impute liability directly to the machine for damages created.

This solution is certainly appealing and is inspired by the legal responsibility accorded to companies; however, to date this solution does not seem necessary for at least two reasons:

1. machines have not reached a point of such autonomy requiring a legal recognition of personality; machines are still fully at the service of man, hence programmed to carry out certain tasks; and
2. even if we decided to recognise the legal personality of machines, this would not solve the problem of who would be called upon to compensate for the damage caused.

   In fact, machines do not have assets, so the manufacturer would end up being liable anyway. One might think of endowing the machines with assets, but the extent of this would remain uncertain at the moment. In this respect, we do not have sufficient case history to determine what the amount of damages awarded might be.

If, then, autonomous machines are not recognized as having a personality, and are therefore not held directly responsible, it is necessary to consider whether and how the legal instruments in force today can regulate the phenomenon.

3.1 Liability for dangerous goods

Given the tensions generated with the fault-based liability regime, the question arose as to whether the liability regime governing dangerous activities could be applied.

In fact, it was thought that in the face of the difficulty of foreseeing potentially even serious damage, as could happen, for example, in the case of road accidents for driverless cars, if artificial intelligence were to be applicable an objective liability would be placed on the manufacturer.

Dangerous activities in fact, even though they generate substantial risks for the safety of third parties, are considered to be lawful in any case, on account of the benefits they bring to the owner or the community, as for example in the case of nuclear power plants. However, in order to balance the risk created by these, it has been provided that those who benefit from them are also called upon to answer for the damage generated in an objective manner.

What is thus envisaged is liability without fault; it is not necessary that the holder of the activity has acted negligently, *i.e.*, recklessly, negligently or carelessly, and caused compensable damage. Rather, it is sufficient that he/she has merely engaged in the dangerous activity, which in itself generates risks, for him/her to be held liable.

This liability regime is particularly burdensome, and its extensive application could easily discourage investment in the sector and thus potentially block an industry that appears to be highly productive.

Nor does an extensive application to the entire artificial intelligence sector seem desirable. While this would ensure that the injured parties would be compensated for the damage suffered, at the same time careful reflection is required since not all systems could be classified as a dangerous activity per se, since the risks generated by their use are not the same and the sacrifice required of the manufacturer would then not be proportionate to the risk generated.

Extensive application does not seem feasible, the question then arose as to whether product liability could apply instead.

3.2 Product liability

The question arose as to whether artificial intelligence applications could be qualified as products and thus be subject to the strict producer liability regime provided for in the Defective Products Directive 85/374/EEC.

The directive provides that if a consumer suffers damage from a defective product, she/he can claim compensation directly from the producer, a party with whom consumers often do not have a direct relationship.

The consumer will therefore be required to prove not only the damage suffered, but also the defect in the product and the causal link between the defect and the damage suffered.

Although this solution would seem to be the most suitable to regulate the artificial intelligence systems various concerns arise.

First, the qualification of artificial intelligence systems as products is not straightforward; artificial intelligence algorithms could in fact also be qualified as services embedded within a product.

In addition, autonomy and opacity could make it difficult, or in any case excessively burdensome, for injured parties to identify those responsible and prove the existence of a causal link between defect and damage, and this is also

due to the long chain of actors involved in the creation of artificial intelligence systems.

In addition, the Directive itself provides for an exclusion of liability, in the event that:

1. the defect did not exist when the manufacturer put the product into circulation;
2. if the defect is due to the conformity of the product with a mandatory legal norm or binding measure; and
3. if the state of scientific and technical knowledge at the time the manufacturer put the product into circulation did not yet permit the product to be considered defective.

It is easy to see how these exemption cases can be applicable to artificial intelligence applications. In fact, as we have seen, the cause of harmful outputs is not fully comprehensible, nor can these be foreseeable at the time the artificial intelligence is put into circulation.

Moreover, it could be the case that the ability to learn autonomously could lead to the emergence bias, which consequently generates malicious output.

In these exclusion cases, the injured consumer remains without protection.

In light of these critical issues, on 28 September 2022, the European Commission presented two proposals for directives aimed at making liability rules more suitable for the digital age.

First of all, a revision of the product liability directive with the aim of making the injured party better protected and thus incentivising him/her to use digital products.

An important provision in the draft directive concerns precisely artificial intelligence applications.

The Commission prescribes that damages caused by robots, drones and other intelligent systems made insecure by software updates or by artificial intelligence will have to be compensated.

Moreover, it is provided that the operator, *i.e.*, no longer just the manufacturer but all those involved in the production chain of artificial intelligence (like the designer or programmer of the algorithm), will be liable for all damage caused, including psychological damage, damage to property or loss of data, but not for the violation of fundamental rights, as in the case of a discriminatory result.

In addition, Article 8, which provides for an easing of the burden of proof on consumers, seeks to resolve the information asymmetry that characterises the relationship with the producer. In fact, it is proposed to require economic operators to disclose information necessary to demonstrate the existence of a causal link between the "defective" AI and the damage suffered.

4. The risk-based approach: the AI Act and the proposed AI liability directive

The European risk-based approach found in the AI Act also extends to the area of civil liability. Although we do not find a reference to this in the Regulation, a connection and coordination can be found in the draft directive on civil liability, published by the Commission on 28 September 2022.

The legislative text aims to regulate civil lawsuits brought in cases of non-contractual liability for damages arising from the use of artificial intelligence systems.

In the text, after providing for a duty of care on the part of suppliers, developers and the subjects themselves who use AI applications, it is provided that anyone can take legal action for damages resulting from the violation of subjective legal situations such as life, health, privacy, property, but also in cases of discrimination.

Although no specific applicable liability regime is envisaged, and this is evidently in order not to conflict with the different regimes applicable in each member State, the Commission proposes certain rules of a procedural nature.

Article 4 introduces an easing of the burden of proof of the causal link for injured parties, by means of a presumption as to the existence of the causal link. The injured party should only prove that a legal or judicial duty was not fulfilled - and its relevance to the production of the damage - or that the culpable conduct of the damaging party influenced the output of the artificial intelligence system, thereby causing damage.

Furthermore, in the case of high-risk AI systems, as defined by the AI Act, the draft directive grants injured parties the right to have access to evidence that may constitute proof of the defectiveness of the product. Article 3 provides that the Member States must set up a mechanism that allows the court to order the person who has access to it (producer, supplier or user) to show evidence that is available to it on a specific high-risk artificial intelligence system. In the event of non-compliance with the order, the draft directive provides for a presumption of non-compliance with the duty of care, such as the AI Act.

The provision concerning the use of artificial intelligence systems by users is also interesting. The article 4(6) provides that where the application of artificial intelligence is used in a personal, and thus non-professional, context, the presumption of causation applies only where the user has materially interfered with the operating conditions of the AI. The presumption applies also where users although required and able to determine the operating conditions had failed to do so.

The regulatory text has not yet entered into force, but it nevertheless proves to be an interesting instrument and an important building block in the regulation of AI, because its provisions are capable of completing and specifying

certain aspects of the AI Act. Moreover, if it were to come into force it would be an instrument aimed at creating an all-round protection and regulation of the AI; in fact, if the AI Act has an *ex ante* application, with a risk-based approach aimed at limiting the risk of damaging events, the proposal for a directive would instead regulate the subsequent phase, once the damage has occurred. The draft Directive wants to introduce some corrective measures aimed at ensuring greater protection for injured parties.

Hopefully, both draft directives can enter into force to complete the provisions of the AI Act.

Bibliography

Bertolini, A. (2013) 'Robots as Products: the case for a realistic analysis of robotics applications and liability rules', *Law Innovation and Technology*, 5(2).

Brkan, M., (2019) 'Do algorithms rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond', *International Journal of Law and Information Technology*, (27)2.

Goodman, B., Flaxman, S. (2017) 'European Union Regulations on Algorithmic Decision Making and a "Right to Explanation"', *AI Magazine*, (38)3.

Malgieri, G., Comandè, G. (2017) 'Why a right to legibility of automated decision-making exists in the general data protection regulation', *International Data Privacy Law*, (7)3.

O'Neil, C. (2016) *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Denver: Crown.

Pariser, E. (2012) *The filter bubble. What the Internet is hiding from you*. London: Penguin Book.

Pasquale, F. (2015) *The black box society: The Secret Algorithms that control money and information*. Cambridge: Harvard University Press.

Wachter, S., Mittelstadt, B. and Floridi, L. (2017) 'Why a right to explanation of automated decision-making does not exist in the general data protection regulation', *International Data Privacy Law*, (7)2.

Chapter IX.
Artificial Intelligence and Ethics

by Simona Klimbacher*

Index: -1. Introduction. - 2. Law, fundamental rights and ethics of artificial intelligence systems. - 3. The European Union ethical principles: a) in theory and b) in practice. - 4. The United Nations international governance of AI and the UNESCO recommendations on ethics. - 5. The opacity of AI systems: the "explicability" requirement. - 6. Proposal for a regulation of ethical principles and the AI Act. - 7. The chatbots forecast for the future on ethics and the Moral Machine Project of the Massachusetts Institute of Technology.

1. Introduction

Artificial Intelligence ("AI") has the potential to redefine our traditional ethical approaches and moral theories. The onset of machines that may either meet or supersede human capabilities, poses a big challenge to humanity's self-understanding as the only beings in the world.

The first ethical code for AI systems was introduced by the science fiction writer Isaac Asimov in 1942, who presented his Three Laws of Robotics in a short story entitled *Runaround*. These three laws were later supplemented and replaced by a fourth law in 1985, namely the Zero Law of Robotics.

The four laws are:

1. a robot may not injure a human being or, through inaction, allow a human being to be harmed;
2. a robot must obey the orders given it by human beings except where such orders would conflict with the first law;
3. a robot must protect its own existence as long as such protection does not conflict with the first or second law;
4. a robot may not harm humanity or, by inaction, allow humanity to suffer harm.

Asimov's laws represent an interesting starting point to think on the ethics of artificial intelligence; although, the real world requires more concrete and adaptable solutions.

---

* Attorney admitted in Italy and New York State (USA). Master of Laws (LL.M.) at Fordham University School of Law. Research Fellow at the Information Society Law Center (ISLC) at the University of Milan.

The academic research on AI dates from the 1950s (Turing, 1950 and McCharthy, 1959) and the ethical debate is almost as old (Wiener, 1960). Nevertheless, only in the recent years, advances in the implementation of AI systems - in almost any sector – have been impressive. Benefits and risks of artificial intelligence have intensified the demand for reflections and practical solutions resulting in a significant number of initiatives for socially beneficial AI.

Due to the widespread awareness of the relevance of the ethical issues related to AI, governments, the European Union, large companies and associations have encouraged the establishment of expert panels and the drafting of policy documents and ethical guidelines for AI.

Hence, several research centers for AI ethics have been established to carry out this work, receiving substantial funding from various public and private sources. In 2020 there were more than 160 guidelines on AI ethics, and the European Commission's AI Act proposal of April 2021, has further animated the debate among scholars on how to discipline the proper and improper uses of AI. However, such document proliferation made it rather harsh to keep track of all the latest AI ethical guidelines being released.

Furthermore, this "hyperactivity" created a marketplace of principles and values in which public and private actors could "shop" for the kind of ethics that would best fit their behavior. An "ethical shopping" permitted by the same countless ethical principles, codes, guidelines or frameworks that result confusing and detrimental to the creation of generally accepted standards (Floridi, 2022).

AI's new set of opportunities and constraints highlight how AI affects our understanding of ourselves and of our environment and realigns traditional issues of rights and duties, autonomy and accountability, negligence, and duties of care (Pagallo and Durante, 2022).

AI offers immense opportunities but also many risks related to its use; this is why it is crucial to establish upfront the principles and values that designers, engineers, and programmers need to incorporate in the development of artificial intelligence systems.

In this respect, it is worth mentioning the six-high profile initiatives in the interest of socially beneficial AI published between 2017 and 2018 by highly reputable authorities and incorporating forty-seven principles, including the Asilomar AI Principles, the Montreal Declaration and the European's Commission's Group on Ethics in Science and New Technologies (the "Principles") and to recall the four core principles used in bioethics (the branch of ethics devoted to the treatment of moral problems in practice, as opposed to theoretical problems): *i.e.,* beneficence, non-maleficence, autonomy and justice. Scholars Jobin, Ienca and Vayena conducted an in-depth analysis on the soft law tools, recalling *inter alia* the five ethical principles (Jobin, Ienca and Vayena, 2019).

The study conducted by notable scholars found a high degree of overlap among the Principles (often repeating the same concepts) and contradictions (the same language is used for different meanings). Additionally, the study stressed the need for a "new" enabling principle for AI namely, the explicability principle capable of explaining how an intelligent system works and who will be accountable and, therefore, responsible for the way the system works (Floridi and Cowls, 2019). The explicability principle together with the Principles constitute a valid ethical framework to which laws, rules, technical standards, and best practices for ethical AI can be further integrated by sector and industry specific needs in any applicable jurisdiction. The framework's discussion and graphic representation can be found in Professor Luciano Floridi's book and is reported in Figure 9.1. Finally, the explicability principle is further discussed *infra*.

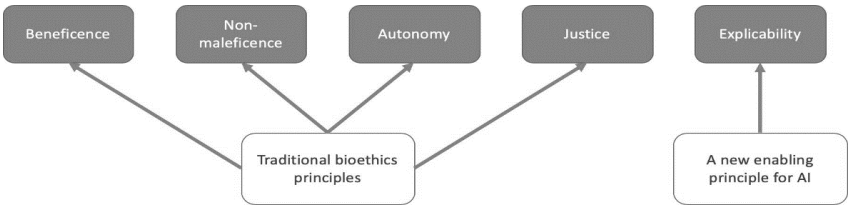


Figure 9.1 An ethical framework of the five overarching principles for AI.

2. Law, fundamental rights and ethics of artificial intelligence systems

The regulatory frameworks concerning AI considers the design and development phases crucial for the creation of and use by individuals and entities of intelligent systems.

With the term "development" reference is made to researchers, designers and developers of AI systems, whereas "deployment" includes public or private organizations that use AI systems within their business processes and offer products and services to others; and, finally, the "end-users" are those engaging with the AI system, directly or indirectly.

A number of legally binding rules at European, national and international level already apply to the development, deployment and use of AI systems today.

Examples of legal sources include:

– EU primary law: the Treaties of the European Union and its Charter of Fundamental Rights;
– EU secondary law: the General Data Protection Regulation, the Product Liability Directive, the Regulation on the Free Flow of Non-Personal

Data, anti-discrimination Directives, consumer law and Safety and Health at Work Directives, the UN Human Rights treaties and the Council of Europe conventions (the European Convention on Human Rights), and numerous EU Member State laws;
- various domain-specific rules applying to specific AI applications (*e.g.*, the Medical Device Regulation in the healthcare sector);
- US Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence in the United States of America (representing a government-wide effort to guide responsible AI through federal agency leadership, regulation of industry, and engagement with international partners).

In the view of the European Commission, AI is only emerging and the EU will take actions to foster the setting of global artificial intelligence standards in close collaboration with international partners in line with the rules-based multilateral system and upholding values.

For this reason, the EU favors partnerships, coalitions and alliances with EU partners (*e.g.*, Japan, US, India, Canada, South Korea, Singapore etc.) as well as multilateral (*e.g.*, OECD, G7 and G20) and regional organizations (*e.g.*, Council of Europe).

Among the main pillars of the European Commission's vision, there is a clear need to address ethical standards based on the Union's values and in line with the EU Treaties, the EU Charter and International human rights law. Many of these rights are - in specific circumstances - enforceable in the EU. However, fundamental rights are often unable to provide a comprehensive legal protection that can cover all cases and cope with the protections required by the fast developments of technology. Hence, standards and partnerships are intended to integrate the EU AI regulation.

Academics also emphasis the relevance of ethical reflections as supporting the understanding about what can be done with technology and how the development, deployment and use of AI systems may implicate fundamental rights and their underlying values. The main scope of ethics is indeed to identify how AI can advance or raise concerns to the good life of individuals in terms of quality of life, human autonomy and freedom in furtherance of a democratic society.

Consequently, a specific ethics code may not function as a substitute for ethical reasoning itself, which must always remain sensitive to contextual details that cannot be inserted in general guidelines or codes and the adherence to ethical principles goes beyond following statutes for compliance purposes (Floridi, 2018).

Professor Setefano Rodotà noted that the relevant issue rests on whether the ethical system is able to regulate specific principles that the society considers

morally just and where instead we need the law to intervene. He states (translated into English) (Rodotà, 2019):

> Legal rules are rules that tend eventually to be imposed even by force. […]. The ethical rule has other paths. It escapes, by its very nature, from this possibility. Consequently, we should determine the extent to which we can leave it to the ethics system, to the moral rules, to regulate behaviour, and when, instead, the intervention of law is necessary to prevent the violation of the principles that we consider fundamental ethical principles.

Provided the above, scholars welcomed the European Commission's Ethics guidelines described hereinafter as representing a recent and complete standard for a socially good AI.

3. The European Union ethical principles:

a) in theory

In June 2018, the European Commission appointed a High Level Group of Experts ("AI HLEG") of 52 experts of diverse backgrounds (*e.g.*, professionals, academics etc.) representing a wide range of stakeholders with the task to draft AI ethics guidelines and prepare a set of policy and investment recommendations.

At the same time, the Commission established the European AI Alliance, an open multi-stakeholder online forum with over 4000 members representing academia, business and industry, civil society, EU citizens and policymakers, to provide a broader input with respect to the work of the AI appointed experts. Specifically, they provided assistance and feedback on policy documents and on the ethical guidelines, academic papers and participating to discussions which helped define the other necessary deliverables such as, for example, the assessment list described *infra*. Although the AI HLEG ended its mandate in July 2020, the community of the AI Alliance continued its activity.

On 8 April 2019, the High-Level Expert Group on Artificial Intelligence issued the "Ethics Guidelines for Trustworthy AI" ("Guidelines") as a result of an in-depth stakeholder consultation resulting in comments - from 511 organizations, associations, companies, research institutes, individuals etc. - and meetings with representatives of the EU Member States.

The Guidelines are indeed addressed to all AI stakeholders designing, developing, deploying, implementing, using or affected by AI, including companies, organizations, researchers, public services, government agencies, institutions, civil society organizations, individuals, workers and consumers that welcomed the practical nature of the Guidelines and the concrete support they offer to developers, suppliers and users of AI on how to ensure trustworthiness.

The aim of the Guidelines is to promote trust *vis-à-vis* artificial intelligence systems which should be met throughout the AI system's entire life in presence of three components:

- lawfulness: complying with all applicable laws and regulations;
- ethical: ensuring adherence to ethical principles and values; and
- robustness: both from a technical and social perspective since, even with good interactions, AI systems can cause unintentional harm.

The four ethical principles rooted in fundamental rights and identified as ethical imperatives are listed hereinafter.

1. Respect for human autonomy. AI systems should complement and empower human cognitive, social and cultural skills and not unjustifiably subordinate, deceive and manipulate. The allocation of functions between humans and AI systems should follow human-centric design principles and allow for human choices by securing human oversight over work processes in AI systems.
2. Prevention of harm. AI systems should neither cause harm or adversely affect human beings by protecting mental and physical integrity. AI systems and the environments in which they operate must be safe and secure. They must be technically robust and not open to malicious use. Greater attention should be granted to vulnerable persons and to asymmetries of power or information (*e.g.,* employers and employees, businesses and consumers, governments and citizens etc.).
3. Fairness. The development, deployment and use of AI systems must be fair so that benefits and costs are distributed equally and individuals and groups are free from unfair bias, discrimination and stigmatization. An effective redress against decisions made by AI systems and by the humans operating them is required and the entity accountable for the decision must be identifiable and the decision-making processes must be explicable.
4. Explicability. Procedures need to be transparent, the purpose of AI systems openly communicated and the decisions or outputs – to the extent possible – explainable to those directly and indirectly affected.

b) in practice

The Guidelines translate the above-listed four ethical principles in seven requirements that should be continuously evaluated and addressed throughout the entire AI system life cycle to create trust in the use of AI systems. The list of seven requirements follow.

1. Human agency and oversight. AI systems should enable humans to make their own informed decisions based on information or decisions, especially in cases where AI systems guide, influence or support humans in decision-making processes. Human oversight mechanisms allow to decide when and how to use, or cease to use, an AI system in a given situation.

2. Technical robustness and safety. Algorithms need to be secure and sufficiently robust to deal with errors or inconsistencies during all phases of AI systems. This includes ensuring there is a plan to address AI systems mistakes, as well as ensuring systems are accurate, reliable and reproducible.
3. Privacy and data governance. Data subjects should have full control over their own data and AI systems should incorporate protections regarding privacy, as well as ensure the quality and integrity of the data used.
4. Transparency. AI development processes should be documented to allow AI systems' outcomes to be traced. Companies should be able to explain the AI system's technical processes and the reasoning behind the decisions or predictions that the AI system makes. Consumers need to be aware that they are interacting with an AI system and must be informed of the system's capabilities and limitations.
5. Diversity, non-discrimination and fairness. AI systems should be inclusive, available and addressed to all users, regardless of age, gender, abilities or other characteristics. Unfair bias should be avoided, as it could have multiple negative implications including the marginalization of vulnerable groups.
6. Societal and environmental wellbeing. AI systems should benefit all human beings and must be sustainable and environmentally friendly. The AI system's impact on parts of the economy as well as the society at large should also be considered.
7. Accountability. Mechanisms should be put in place to ensure responsibility and accountability for the development, deployment and use of AI systems, especially in the occurrence of negative impact on consumers. AI systems should be available for evaluation to auditors and provide adequate and accessible redress procedures to users.

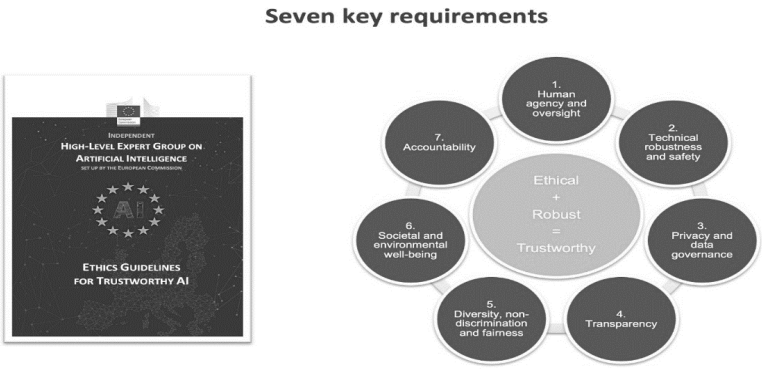


Figure 9.2 Guidelines' representation of the interrelationship of the seven requirements.

Indeed, a trustworthy AI is a continuous process as shown in Figure 9.2 whereby methods should be implemented and amended on an ongoing basis as AI systems are continuously evolving.

Ethics and the rule of law should be incorporated by design. In this respect, methods to ensure values-by-design provide precise and explicit links between the abstract principles which the system is required to respect and the specific implementation decisions.

Consequently, companies are required to identify the impact of their AI system from the start to avoid that outcomes may subsequently result harmful. A system is trustworthy and, therefore, "trusted" when the output is predictable and the behavior understandable. Although, not all systems are predictable and a whole field of research named "Explainable AI" tries to address this issue to better understand the system's underlying mechanisms and to find the relevant solutions. It supports model accuracy, fairness, transparency and outcomes in the decision-making processes and as AI becomes more advanced, humans are challenged to understand how the algorithm came to a specific result.

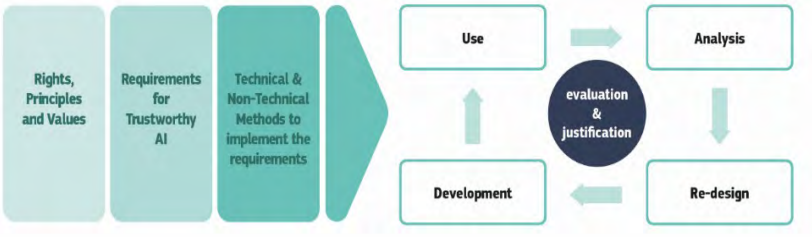


Figure 9.3 Guidelines' representation of a Trustworthy AI throughout the system's entire life cycle.

The AI HLEG recommends that organizations perform a fundamental rights impact assessment ("FRIA") to determine whether their AI systems respect the EU Charter of Fundamental Rights and the European Convention on Human Rights and respond to the questions listed hereinafter that recall specific articles of the Charter and the European Convention on Human Rights.

Human Agency and Oversight	• Is the AI system designed to interact, guide or take decisions by human end-users that affect humans or society? • Could the AI system generate confusion for some or all end-users or subjects on whether they are interacting with a human or AI system?
Technical Robustness and Safety	• Were adequate measures put in place to ensure the integrity, robustness and overall security of the AI system against potential attacks during its life cycle? • Were the risks, risk metrics and risk levels of the AI system defined in each specific use?
Privacy and Data Governance	• Have adequate measures been put in place to ensure the protection of personal data with respect to the development, deployment and use phases of the AI system? • Was the impact of the AI system considered as it relates to the right to privacy, the right to physical, mental and/or moral integrity and the right to data protection?

Transparency	• Were adequate measures put in place to address the traceability of the AI system during its entire life cycle? • Were decisions of the AI system explained to users?
Diversity, Non-discrim-ination and Fairness	• Does the AI system potentially negatively discriminate against people on the basis of race, gender, age or any other characteristics? • Was a strategy or a set of procedures established to avoid creating or reinforcing unfair bias in the AI system, regarding both the use of input data as well as the algorithm design? • Was a mechanism put in place to allow users to flag of issues about bias, discrimination or poor performance of the AI system?
Social and Environmental Well-being	• Where possible, were mechanisms established to evaluate the environmental impact of the AI system's development, deployment and/or use (*e.g.,* the amount of energy used and carbon emissions)?
Accountability	• Were mechanisms that facilitate the AI system's auditability (*e.g.,* traceability of the development process, the sourcing of training data and the logging of the AI system's processes, outcomes, positive and negative impact) established? • Was an AI ethics review board or a similar mechanism established to discuss the overall accountability and ethics practices, including potential unclear grey areas?

Both the Guidelines and the impact assessment list of questions consider the reliability of AI systems the backbone of the future legal framework for AI based on the EU's respect for fundamental rights and ethics at its core by means of a risk-based approach and mandatory preemptive compliance assessments for "high-risk" AI systems that pose significant risks especially to the health and safety or fundamental rights of individuals. The scope of the assessment will include several key elements for a better understanding of the risks and relevant remedies. For example, the duration and frequency of the use of the high-risk AI system, the categories of individuals and groups that could be affected by the use of AI systems in the specific context, the specific risks of harm that may impact the identified categories of people etc.

For an additional description of the beneficial opportunities and the range of concerns of AI reference can be made to the final section of the Guidelines recalling *inter alia* the identification and tracking systems, hidden AI systems, social scoring systems of citizens, autonomous weapons systems etc. Furthermore, Professor Alessandro Mantalero offers practical methodology to conduct impact assessments of AI systems based on the respect of fundamental rights, ethical and societal values providing strategies on the integration of ethical values in the specific contexts in which artificial intelligent systems will operate (Mantalero, 2022).

The seven Principles listed above are relevant and connected to each other. With respect to the "accountability" principle, it is worth recalling the pursuit of the Institute for Ethics in Artificial Intelligence (the "IEAI") in establishing a comprehensive and practical framework for AI accountability based on the AI ethics principles. The IEAI mission is to promote the importance of AI ethics and the responsible use of AI by diving into the ethical issues related to

the development, use of AI technologies and moving the discussion from theory to practice. The Institute gathers several stakeholders to build partnerships and tackle AI ethics challenges in a comprehensive way. The key mechanisms to enable accountability are: transparent processes, stakeholder engagement, ethics integration and continuous monitoring, as represented by the IEAI in Figure 9.4. Indeed, the basic idea is to break ethical obligations down to more concrete actions for which responsibilities and accountabilities can be defined more clearly. The continuous measurement and monitoring will allow a swifter reaction in cases of possible harm.

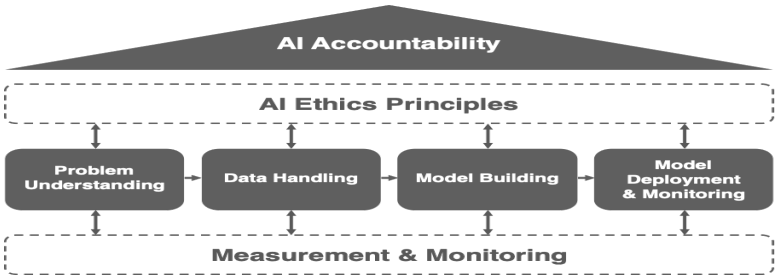


Figure 9.4 - The IEAI accountability framework for AI systems.

4. The United Nations international governance of AI and the UNESCO recommendation on ethics

In December 2023, the United Nations ("UN") Secretary-General's AI Advisory Body issued an Interim Report entitled "Governing AI for Humanity" ("Report").

The Report calls for a closer alignment between international rules and how AI is developed and used. It acknowledges that there is no global alignment on implementation between jurisdictions across the world. Further, some countries favour rules while others prefer a non-binding approach. Hence, in many jurisdictions, the governance of AI can amount to self-policing by the developers, deployers and users of AI systems. However, such an approach may not work in the long term.

The UN is guided by norms and principles (*e.g.*, UN Charter, The Universal Declaration of Human Rights, international and environmental law etc.) to which all its country members are required to commit.

The High-Level Advisory Board for AI was formed with the purpose of furthering recommendations for the international governance of artificial intelligence use for the benefit of all humanity. Members of this board include individuals with different academic and professional backgrounds as well as age, gender and geography, specific sector (*e.g.*, private, public, academia etc.).

For example, Father Paolo Benanti is one of the board members. He is a professor both in Italy and in the United States and chairman of the Artificial Intelligence

Commission by the Italian Government (to help safeguard journalism from fake news and other disinformation); due to his expertise in engineering and passion for the ethics of technology his participation in the UN will certainly foster the ethical approach for AI.

The UN Advisory Board identified five guiding principles and seven institutional functions capable of governing AI at an international level. At the global level, international organizations, governments, and private sector would bear primary responsibility for these functions depending on the layer of governance. The five principles that should guide the formation of new global governance institutions consider that AI should be governed for the benefit of all and in the public interest through the promotion of data commons primarily to assist societal changes such as climate change, public health, economic development, crisis response and rooted in adaptive multistakeholder collaboration.

The seven institutional functions detailed in the Report may be carried out either by individual institutions or by a network of institutions. Civil society, academia, scientists etc. play key roles in building evidence for policy, assessment impact as well as holding key actors to account during implementation.

The seven functions are shown in Figure 9.5 and explained more in detail in the following paragraphs.

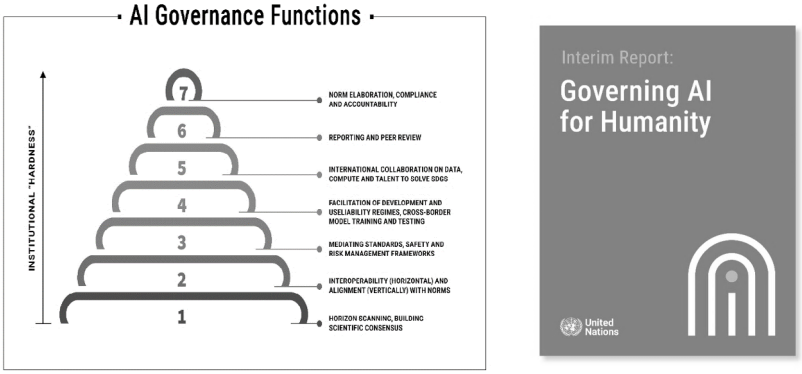


Figure 9.5 The seven levels identified by the AI Advisory Board of the United Nations in the Interim Report of December 2023.

1. Assessment of the future directions and implication of AI: to understand the direction and pace of AI technologies and relevant risks and opportunities to allow policymakers to develop AI programmes encouraging innovation and managing risks provided there is no institutionalized multidisciplinary authority established to-date.
2. Reinforcement of interoperability of governance efforts emerging around the world and their grounding in international norms through the Global AI

Governance Framework endorsed in a universal setting: to support policy-making and private sectors, regions, and countries and to support the principles and norms under which various organizations should function in order to avoid AI divides and governance gaps across private and public sectors.

3. Development and harmonization of standard, safety, and risk management frameworks: due to the lack of a global harmonization and alignment whereby the UN can play a critical role in bringing states together developing common socio-technical standards and ensuring legal and technical interoperability.
4. Development, deployment, and use of AI for economic and societal benefit through international multi-stakeholder cooperation: to facilitate responsible and beneficial use of AI as well as participate in international multi-stakeholder cooperative frameworks to develop enablers for AI.
5. Promotion of international collaboration on talent development, access to compute infrastructure, building of diverse high-quality datasets, responsible sharing of open source model and AI enabled public goods for the Sustainable Development Goods: by sharing tools for research and development purposes, tracking of positive uses of AI by the private sector ect.
6. Monitoring of risks, report incidents, coordinate emergency response: capabilities at a global level should be created to monitor, report and rapidly respond to vulnerabilities and disruptions to international stability.
7. Compliance and accountability based on norms: non-binding norms could also play an important role alone or in combination with other norms; although the UN should not seek to be the sole arbiter of AI governance in certain fields such as challenges to international security it has a unique legitimacy to elaborate norms and it can also assist ensure that there are no accountability gaps.

The Report provides for an interesting table of subfunctions at pages 22 and 23 for international governance of AI and the relevant possible timeframes for the relevant implementation. For example, subfunction 10 encourages the development of best practices, aligning soft and hard law, standards, methods and frameworks at a regional, national and industry level (to support interoperability). The timeframe identified to institutionalise such proposed subfunction is from twelve to twenty-four months. Subfunction 12 encourages stakeholders to negotiate non-binding and binding frameworks, treaties, or other regimes for AI with a timeframe from twenty-four to thirty-six months to carry out such tasks. By reading the other 13 subfunctions of the table, special care should rest on the timeframe; indeed, time could vanish the purpose of the task itself.

On 5 and 6 February 2024, the United Nations Educational, Scientific and Cultural Organization ("UNESCO") second Global Forum on the Ethics of Artificial Intelligence took place in the Brdo Congress Centre of Kranj, entitled: "Changing the Landscape of AI Governance", whereby the experiences

and expertise of countries at different levels of technological and policy development could meet to learn from each other and dialogue with the private sector, academia and a wider civil society. UNESCO is a specialized agency of the UN with the aim of promoting world peace and security through international cooperation in education, arts, sciences, and culture. It has led the international effort to ensure that sciences and technology develop with strong ethical guardrails for decades. It counts 193 member states as well as partners in the non-governmental, intergovernmental and private sector.

On 23 November 2021, it issued the first-ever global standard on AI ethics the "Recommendation on the Ethics of Artificial Intelligence" ("Recommendation").

After having set forth the scope, aims, objectives, the Recommendation describes the values and principles and the areas of policy action, monitoring and evaluation, utilization and exploitation of AI systems.

The listed values include: (i) the respect, protection and promotion of human rights and fundamental freedoms and human dignity, (ii) environment and ecosystem flourishing, (iii) ensuring diversity and inclusiveness, living in peaceful, just and interconnected societies.

The UNESCO principles are ten and recall the Principles set forth in the Guidelines although the titles and order differ: (i) proportionality and do no harm, (ii)safety and security, (iii) fairness and non-discrimination, (iv) sustainability, (v) right to privacy and data protection, (vi) human oversight and determination, (vii) transparency and explicability, (viii) responsibility and accountability, (ix) awareness and literacy, (x) multi-stakeholder and adaptive governance and collaboration. For examples, the transparency and explainability principles are joined in the Recommendation. Indeed, explainability is closely related to transparency, as outcomes and sub-processes leading to outcomes should aim to be understandable and traceable, appropriate to the context. In this respect, paragraph 40 states: 'It should be ensured that the meaningful explanation is provided with any decision that resulted in the action taken in order for the outcome to be considered transparent."

However, what makes the Recommendation exceptional are its extensive policy action areas detailed at section IV of the Recommendation. Such policy actions allow policymakers to translate the core values and principles into action with respect to data governance, environment and other spheres.

5. The opacity of AI systems: the "explicability" requirement

The explicability requirement we recalled earlier, is considered the basic ethical criteria to accept a given AI decision or output (Florid and Cowls, 2019).

Artificial intelligence includes several approaches and techniques capable of generating outputs like machine learning, deep learning, robotics that are used

to make numerous decisions that impact significantly the life of individuals. For example, AI systems may decide who may be eligible for a specific loan, law school, and job.

Although, bias and discriminations are present in almost any human activity, if these result from an AI system they have a much larger impact on individuals and a number of people involved in the absence of adequate supervision or in cases where the outcome cannot be prevented upfront by design. Indeed, such impact is the result of correlations made by machine using large datasets of information and at times such correlations can be unexpected.

Consequently, if the decision generated by an autonomous AI system is not explicable, the issue rests on how can people (directly affected by such decisions) be protected against unjust exclusions or discriminations.

The specific features of many AI technologies such as opacity, the so-called "black-box" effect, or unpredictability may make it hard to verify compliance with rules and individuals and legal entities may have no effective legal remedy in cases that prejudice and negatively impact them.

The opacity issue requires additional clarifications. First, the opacity of decisions can be of different types. Some decisions are simply difficult for people to understand because their lack of technical expertise on how an AI system actually works, although the technology is not opaque. In other cases, the decision-making process is simply unclear to everyone, including to experts determining the afore-mentioned "black-box" effect. In other cases, sharing of information can be limited because the output is protected by commercial trade secrets. Hence, these cases show how the dignity and autonomy of individuals and the democracy of the society are at stake in the absence of a guiding philosophy of ethical values.

In cases where systems are powered and function by means of artificial intelligence the principle of explicability is always mentioned by policymakers, big tech companies and academics to stress the need to explain the process that allows the machine to deliver a certain decision or output allowing individuals to receive a clear explanation of the decision-making process especially if the result is unwanted (Floridi *et al.*, 2018).

Interestingly, Scott Robbins (Robbins, 2019) provides noteworthy arguments against the necessity to provide explicability in every case. In his view, there is a need to be able to explain only certain decisions that may significantly impact one or more individuals (*e.g.*, explaining why someone decided to hurt another individual) as opposed to actions that are not relevant to anyone (*e.g.*, explaining why the individual decided to dance).

Provided that automation is still an option that can be avoided in specific relevant circumstances, Robbins further argues that for AI we should decide which decisions require an explanation and exclude machine learning outputs in these specific cases.

Robbins suggests the solution for the opacity of machine learning can be found using decisions or actions which do not require an explanation. In the healthcare sector, machine learning can make decisions that do not require explanations, although its outcomes offer significant ethical and societal consequences. For example, for the detection of cancerous moles, an algorithm can take a picture of a mole and classify it as malignant or not and the consequences of this decision are simply a biopsy if the mole is labeled as malignant. However, algorithms that perform tasks of this type, may still be biased and indirectly prejudice a group of people; for example, performing poorly on those with a dark complexion. Here the issue does not involve explicability; the algorithm is not using skin color to determine whether a mole is malignant and the algorithm is simply not very good at labeling moles on patients with dark complexion.

The principle of explicability allows humans a control over AI. The explanation of an algorithm's output will allow an individual to control the process by deciding what to do with the relevant output, remaining accountable for the ultimate decision made (either accepting or rejecting the decision). In this respect, the chairman of the Italian Data Protection Authority, a few years ago, recalled the need for a "new ethics" for artificial intelligence in order to hold responsible those involved in the development, deployment and use of the AI systems. Such new ethics should rest on the "awareness of individuals" and on "regulations". Regulations are necessary to allow individuals to make choices against possible discriminations. For example, individuals should be able to refuse to be made subject to exclusive automated decisions in absence of their consent as provided by the European Data Protection Regulation.

In the forthcoming years, the opacity of algorithms will certainly require special attention. The explicability requirement is highly dependent on the specific situation and possible consequences that a decision may cause in cases of mistakes; for example, at footnote 33 the AI HLEG Guidelines state that: "little ethical concern may flow from inaccurate shopping recommendations generated by an AI system, in contrast to AI systems that evaluate whether an individual convicted of a criminal offence should be released on parole".

6. Proposal for a regulation of ethical principles and the AI Act

The two main documents that frame the European Commission's strategy on AI are the Communication on Europe's Digital Future of 20 February 2020 and the White Paper of 19 February 2020.

The White Paper suggests strategies on how to promote the adoption of AI and deal with the risks associated to specific uses of such technology and it

encourages the development of an ethical AI within a uniform European legal framework able to provide an EU-wide approach, and avoid fragmentation in the internal market, which would undermine the objectives of trust, legal certainty and market uptake.

On 20 October 2020, the Resolution of the European Parliament on the framework of ethical aspects of artificial intelligence, robotics and related technologies recommended to the Commission the issuance of a regulation to govern "the good use of AI" and ensure the respect of the "ethical principles".

The ethical principles proposed by the Resolution are broader in scope than the ethical imperatives of the HLEG recalled *supra* at paragraph 2. The proposed principles are listed hereinafter:

- human-centric, human-made and human-controlled artificial intelligence;
- mandatory compliance assessment of high-risk artificial intelligence;
- safety, transparency and accountability;
- safeguards and remedies against bias and discrimination;
- right to redress;
- social responsibility and gender equality in artificial intelligence, robotics and related technologies;
- environmentally sustainable artificial intelligence, robotics and related technologies;
- respect for privacy and limitations on the use of biometric recognition;
- good governance relating to artificial intelligence, robotics and related technologies, including the data used or produced by such technologies.

At this time, the European AI Act is the first formal legislation to begin to fill in the gaps of ethical and regulatory principles to which organizations must adhere when deploying AI.

In October 2023, President Biden issued an executive order entitled: "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" ("US AI Executive Order"). It aims at establishing a comprehensive framework for the ethical development, deployment, and regulation of artificial intelligence technologies across government agencies and private sectors in the United States territory. Testing and evaluations, including post-deployment performance monitoring, will help ensure that AI systems function as intended and are *inter alia* ethically developed and in compliance with applicable Federal laws and policies. The choice of an executive order is based on the urgency to protect Americans against intelligent systems harms - in time - before a legislative act could actually be agreed and entered into force; as executive orders are immediately enforceable.

On 8 December 2023, the European institutions reached a provisional political agreement on the first comprehensive law on artificial intelligence establishing obligations for AI systems according to a risk–based approach

implying stricter rules for higher risks. Specific compliance requirements such as conformity assessments, data governance requirements, risk management and quality systems, human oversight, robustness etc., will apply to AI systems classified as "high risk" due to their potential ability to harm fundamental rights, democracy, the rule of law etc. The Act will apply also to extra-EU entities involved in the development, design etc. of AI systems if the system is placed on the EU market or its use affects people located in Europe. Furthermore, the European Commission set up an AI Office within its Commission, tasked with overseeing the most advanced AI models, contributing to fostering new standards and testing practices, and enforcing the common rules in all EU member states.

Professor Giusella Finocchiaro in her recent book on artificial intelligence (Finocchiaro, 2024) reminds her readers that the European legislation should not have the effect of isolating Europe. It is essential to propose and affirm a European model but at the same time – it is equally relevant – to foster an international cooperation with other regions of the world such as the United States and China. In this respect, Professor Finocchiaro recalled a metaphor: "if we want to build fortresses we must also remember to build the bridges that enable us to connect them to other systems".

7. The chatbots forecast for the future on ethics and the Moral Machine Project of the Massachusetts Institute of Technology

A tech editor, Grace Chng, noting how in year 2023 the technology on AI chatbots has advanced rapidly carrying with it a significant increase of issues on ethics, wondered what would selected chatbots think about these issues. She elaborated a "roundtable" discussion with ChatGPT, Claude and Bard on their future perspectives.

An abstract of some questions relevant to the ethical aspect that should stimulate further reflection are reported below and have been edited for brevity:

Question 1: How do you navigate ethical dilemmas?

Bard: "As a large language model, I am not designed to make subjective moral judgments. Instead, I navigate ethical dilemmas by relying on the principles and values embedded in my training data. This includes datasets of factual information, ethical guidelines, and diverse viewpoints. When faced with conflicting values, I prioritise accuracy and objectivity over personal opinions. For example, if asked about a controversial topic, I would present a balanced overview of different perspectives, citing reliable sources, and avoiding bias. An example of ethical considerations influencing my response would be a query about sensitive topics

like suicide or self-harm. In such cases, I would prioritise user safety and well-being. I would provide information about mental health resources and hotlines, encouraging the user to seek professional help."
ChatGPT: "I don't have personal experiences, consciousness, or the ability to navigate ethical dilemmas autonomously. My responses are generated based on patterns and information present in the data on which I was trained. While I can provide information and perspectives on ethical considerations, the responsibility for making ethical decisions ultimately lies with individuals and those implementing or interacting with AI systems. It's important for developers, users, and organisations to establish ethical guidelines, principles, and oversight mechanisms when deploying AI systems."

Question 2: Responses provided by AI systems need to be approved by a human. Why do you agree or disagree?

Bard: "I believe that human oversight is crucial for responsible AI development and deployment. While I can generate text independently, my responses are often reviewed by human experts to ensure accuracy, safety, and alignment with ethical guidelines. This helps to mitigate potential risks and biases. However, complete human oversight may not be feasible or desirable in all cases. As AI technology advances, it is likely that more sophisticated models will be able to generate responses that do not require human intervention. The key is to find a balance between ensuring responsible AI development and maximizing the potential benefits of this technology".
Claude: "I agree that having human oversight for responsible and ethical AI systems is important. I have been created to be helpful, harmless, and honest, my responses already incorporate levels of human review. My training methodology involved extensive human review – researchers supervised datasets, conversations, and model versions to ensure I meet safety and accuracy standards before deployment. I was designed within a tested conceptual framework for AI alignment focused on minimising potential harms. Humans set these philosophical guardrails. For any commercial or high-stakes real world usage, I would also recommend having a human-in-the-loop both monitoring conversations and available on-demand for users. This allows for contextual intervention and reference. At the same time, not all AI assistants require intensive human approval workflows beyond responsible development practices. Conversation contexts vary greatly, from casual chat to sensitive use cases. And human reviewers in loops face issues around bias, fairness, scalability and cost-effectiveness. The levels of human oversight merit case-by-case analysis rather than default requirements".

The answers of Claude, Bard and ChatGPT highlight *inter alia* the crucial issue on how ethics should govern human relationship with technology in the absence of universally accepted ethical principles when it comes to incorporating technology into the automated systems to which we entrust decision-making tasks.

Interestingly, a group of researchers at the Massachusetts Institute of Technology (MIT) created a Moral Machine platform for gathering human perspectives on moral decisions made by machine intelligence, such as, for example, self-driving cars. The platform generates moral dilemmas. Individuals may express their preferred outcome and compare their responses with other people choices. Further, they can design their own scenarios, for others to view, share, and discuss. The platform can be accessed at https://www.moralmachine.net.

Allowing the public to express opinions and conforming to the choices made by the majority may prove useful in the decision-making process of ethical dilemmas by means of a democratic approach. The drafting of the Artificial Intelligence Act followed the same path.

Provided the above, the main issue we are – and will increasingly be – facing is how to govern AI worldwide coherently. Specifically, ethical choices come in when we try to find the right level of its use inside a social environment; hence, regulating artificial intelligence means developing systems compatible with democracy. The journey is still a long way to go.

Bibliography

Asimov, I. (1942) *Runaround: A Short Story*. New York: Street and Smith.

Asimov, I. (1986) *Robots and Empire: The Classic Robot Novel*. HarperCollins.

Benanti, P. (2022) *Human in the Loop*, *Decisioni umane e intelligenze artificiali*. Edizione digitale, Mondadori.

Chng, G. (2023) *An interview with three chatbots: AI ethics, governance and developments in 2024*. [online] Techgoondu. Available at https://www.techgoondu.com/2023/12/29/ai-ethics-governance-and-developments-in-2024-an-interview-with-three-chatbots/.

Finocchiaro, G. (2024) *Intelligenza artificiale. Quali regole?* Bologna: Il Mulino, pp. 45, 124-125.

Floridi, L. (2022) 'Lo Shopping Etico' in *Etica dell'intelligenza artificiale: Sviluppi, Opportunità, Sfide*. Milano: Cortina, pp. 101, 109-111.

Floridi L, et. al. (2018) AI4People-An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles and Recommendations. *Minds and Machines*, 28(4), pp. 689-707. Available at: https://link.springer.com/article/10.1007/s11023-018-9482-5.

Floridi, L. and Cowls, J. (2019) 'A Unified Framework of Five Principles for AI in Society', *Harvard Data Science Review*, 1(1), pp. 699-700. Available at: https://doi.org/10.1162/99608f92.8cd550d1

Floridi, L. (2022) 'Un Quadro Unificato di Principi Etici' in: *Etica dell'Intelligenza Artificiale. Sviluppi, opportunità e sfide*. Milano: Cortina, p. 101.

Floridi, L. (2018) 'Soft Ethics and the Governance of the Digital', *Philosophy & Technology*, 31(1), pp. 1-8. Available at: https://doi.org/10.1007/s13347-018-0303-9.

Li, F.-F. (2023) *The Worlds I see*. New York: Flatiron Books, A Moment of Lift Book.

Jobin, A., Ienca, M. and Vayena, E. (2019) 'The global landscape of AI ethics guidelines', *Nature Machine Intelligence*, 1(9), pp. 389-399. Available at: https://www.nature.com/articles/s42256-019-0088-2.

Lifschitz, V. (2011) 'John McCarthy (1927–2011)'. *Nature,* 480 (40). Available at: https://doi.org/10.1038/480040a.

Mantalero, A. (2022) *Beyond Data*. Den Haag: T.M.C. Asser Press. Available at: https://link.springer.com/book/10.1007/978-94-6265-531-7.

McCarthy, J. (1959) *Programs with Common Sense*. Proceedings of the Teddington Conference on the Mechanization of Thought Processes, pp. 756–91. Available at: http://jmc.stanford.edu/articles/mcc59/mcc59.pdf.

Pagallo, U. and Durante, M. (2022) 'The good, the Bad, and the Invisible with its Opportunity Costs: Introduction to the 'J' Special Issue on 'the impact of Artificial Intelligence on Law's', 5(1), pp. 139-149. Available at: https://doi.org/10.3390/j5010011.

Robbins, S. (2019) 'A Misdirected Principle with a Catch; Explicability for AI', *Minds and Machines,* 29(4), pp. 495-514. Available at: https://doi.org/10.1007/s11023-019-09509-3.

Rodotà, S. (2019) *Etica e Diritto (dialogo tra alcuni studenti e Stefano Rodotà),* Costituzionalismo.it, Fascicolo I, pp. 23-32. Available at https://www.costituzionalismo.it/costituzionalismo/download/Costituzionalismo_201901_702.pdf.

Santosuosso, A. (2000) 'Diritti umani ed etica' in *Intelligenza Artificiale e Diritto*. Milano: Mondadori, pp. 32-37.

Turing, A.M. (1950) 'Computing Machinery and Intelligence'. *Mind,* LIX (236), pp. 433-460. Available at: https://doi.org/10.1093/mind/LIX.236.433.

Wiener, N. (1960) 'Some Moral and Technical Consequences of Automation'. *Science,* 131 (3410), pp. 1355-1358. Available at : https://www.science.org/doi/10.1126/science.131.3410.1355

Chapter X
Generative Artificial Intelligence and Copyright

By Eylül Erva Akin*

Index: 1. Generative AI. – 2. Copyright law challenges: redefining creativity and AI's use of copyrighted material. – 3. EU legal framework. – 4. Conclusion

1. Generative AI

Artificial intelligence (AI) has had a profound impact on various aspects of our lives in recent years. From revolutionising industries to enhancing everyday experiences, AI has become an integral part of our modern world. AI encompasses a broad spectrum of technologies and techniques that enable machines to perform tasks that typically require human intelligence. These tasks can include natural language processing, computer vision, machine learning, and more.

Given the diverse range of AI systems and applications, it is understandable that defining AI can be challenging. Such a diverse range of applications makes it complex, having potential risks associated with each use of AI. Therefore, it is argued that policy makers should adopt a risk-based approach when formulating regulations (Schuett, 2023). According to the EU AI Act Article 3(1a), risk is defined as the combination of the probability and severity of harm. Accordingly, risk-based approach involves assessing the potential risks and benefits of AI systems and implementing appropriate measures to mitigate any potential harm. By adopting a risk-based approach, policy makers can strike a balance between promoting innovation and ensuring the responsible development and deployment of AI technologies.

According to Article 3(1) of the EU AI Act, an AI system is a machine-based system that generates outputs such as predictions, content, recommendations, or decisions based on the input it receives. It is characterised by its adaptiveness after deployment and the levels of autonomy it works with.

When it comes to generative AI, one of the key differences is its ability to learn from training material, enabling it to generate new and original content.

* Lawyer. Doctoral Researcher at the University of Milan. Research Fellow at the Information Society Law Center (ISLC).

Other traditional AI models primarily focus on making predictions based on a limited set of parameters, such as weather forecast, blocking spam emails or translating text. On the other hand, generative AI goes beyond prediction and delves into the realm of creativity (Wingström *et al.*, 2022). The learning process is non-linear with the help of neural networks of deep learning, meaning that the model can capture complex patterns and relationships in the data (Drexl *et al.*, 2019). The machine learning allows the model to understand and generate content that is not explicitly present in the training material. By leveraging learnable parameters, generative AI can adapt and refine its output based on the input it receives. In parallel to that, the applications of generative AI are vast and diverse. It can be used to create realistic images, generate natural language text, compose music, and even design new products.

2. Copyright law challenges: redefining creativity and AI's use of copyrighted material

The operation of generative AI depends on large amounts of high-quality data. Foundation models, including large language models (LLMs) such as GPT3 and GPT4, PaLM, or Alexa TM, as well as text-to-image models such as Midjourney or Stable Diffusion, are trained on publicly available materials. These generative AI models use human-created code, text, music and art, often sourced from the web. However, it is important to note that some of these materials may be protected by copyright. Therefore, acts of training may require authorisation from the rightholders under EU copyright law. This leads developers of generative AI models to find legal ways of training the models.

There are two main ways in which training data can lead to illegal consequences. First, it can occur when the training data is used without proper permission, resulting in the unauthorised use of copyrighted material. Secondly, datasets used for training can inadvertently lead to unlawful outcomes such as discrimination, manipulation or the dissemination of misinformation. The main issue between generative AI and copyright law within the scope of this chapter is the use of copyrighted works for machine learning purposes without a copyright clearance (Kop, 2020).

Furthermore, the emergence of generative AI models has challenged the traditional concept of creativity, which poses a difficult issue from a copyright law perspective. The lines between human and machine creativity are blurred. Consequently, the traditional understanding of the term "originality" has been challenged. While AI models can autonomously generate new and unique content, the question of who should be considered the author for copyright becomes complex. There are different opinions among courts and organisations in the US, China, and the EU about whether prompt-based images are copyrightable.

Copyright law, particularly in common-law countries, has long been perceived as a tool primarily to provide economic incentives rather than to protect the creative efforts of authors. This utilitarian view asserts that financial rewards are the driving force behind innovation, as creators supposedly require the promise of profit to produce new works. On the other hand, civil law jurisdictions, such as those in continental Europe, prioritise the protection of authors' efforts and personalities, emphasising the inherent link between creators and their works. Rather than solely focusing on economic gain, copyright serves as a means to preserve the integrity and attribution rights of authors. From this perspective, it is clear that these systems face greater challenges in identifying works in which the distinctive influence of human creativity becomes less significant over time. For instance, when artificial intelligence creates works and human creativity becomes less significant, these works raise questions of authorship and originality.

Nevertheless, both the utilitarian and continental European perspectives argue that people should be rewarded for their creations, whether for economic gain or because of the link between the author and the work. Both perspectives converge on the idea that rewarding creators is not only fair but also essential for the continued progress and development of society. The key legal question is whether the existing framework of copyright law is sufficient to address the complexities of generative AI technology, particularly with respect to the originality and authorship of works created through this innovative method of creation.

The issue of authorship in the context of generative AI is indeed a complex matter. With the increasing use of AI technology, the question of whether AI-generated creations should be eligible for copyright protection has become a topic of debate. Currently, there is no clear consensus on this matter. One could argue that the person who inputs the prompt should be considered the author of the AI-generated creations. According to this viewpoint, the AI system is merely a tool or a medium through which the human author's creativity is expressed. However, some argued that copyrighting AI output would eliminate human incentive to invest in intellectual efforts (Schönberger, 2018).

To continue this debate, it is unclear whether AI can be considered the author of such creations. The idea of granting copyright protection directly to AI would imply recognising them as legal entities, which raises significant challenges that go well beyond the scope of copyright law.

Real-life examples illustrate the contentious issues surrounding the use of AI in the creative industries. In one case, audiobook narrators raise concerns about Apple's use of their voices for machine learning training without a clear consent, potentially contributing to the development of synthetic voices without adequate compensation (Agarwal, 2023). Similarly, three artists filed a landmark lawsuit against AI art generators after discovering that their artwork was being used to train AI models without permission, risking loss of income and control

over their creations.(Dixit, 2023) Getty Images also filed a lawsuit alleging copyright infringement because AI models are trained on copyrighted images scraped from the internet, threatening artists' livelihoods as AI-generated content based on copyrighted material enters commercial markets ("Getty Images Statement - Getty Images," 2023). These cases highlight the urgent need for clearer rules to protect creators' rights and ensure fair compensation in the age of AI-generated content.

People often mention cameras or photographs as illustrations of technological advancements when discussing whether AI-generated images should be granted copyright.

This means that photographs taken by humans are being compared to pictures made by AI. In civil law countries like Germany, however, there is a difference between photographic works and photographs. Photographs are protected under neighbouring rights rather than direct copyright. Therefore, using the camera as an example may not effectively demonstrate that copyright protection is automatically ensured with less effort due to technological progress.

3. EU legal framework

The risks associated with AI activities are prompting policymakers to consider regulating AI activities. Significant and extensive policy activity on AI regulation is taking place at the EU institutional level, but most policy initiatives and research do not directly deal with intellectual property rights and AI. Notwithstanding, AI developers have some ways for using copyrighted material without infringing copyright. One option is to obtain legal permission through licensing agreements with copyright holders. This ensures mutual agreement on usage terms and prevents unauthorised exploitation. However, dealing with multiple authors or rights holders can make this process cumbersome. Furthermore, AI developers can use material from the public domain, where copyright protection has expired or is non-existent, to avoid the need for clearance. Additionally, copyright frameworks, such as the InfoSoc Directive Art. 5 in EU law, offer flexibility through exceptions and limitations, enabling the use of copyrighted data in machine learning applications while complying with legal standards. This approach encourages innovation while also respecting the rights of creators and the public interest.

Copyright law provides certain flexibilities to balance the monopoly of copyright holders against the use and reproduction of the authored work, thus ensuring creativity and fundamental rights and the public interest. Without these flexibilities, copyright law could create overly strict regime, potentially hindering research and limiting the potential for data analysis. In this context, copyright

law offers certain flexibility for generative AI, especially concerning text and data mining.

Text and data mining is the statistical or computational analysis of electronic text, sound, images or data. Text and data mining is widely used by researchers and scientists to extract knowledge from extensive datasets. An everyday example of text and data mining is searching on internet or using voice assistants on smartphones. Text and data mining makes it possible to gain new knowledge and discover new trends. However, in certain cases text and data mining may involve acts protected by copyright, in particular the reproduction of copyrighted works. It is important to have rules on exceptions and limitations under current EU law.

Text and data mining is recognised as a copyright exception or limitation in Articles 3 and 4 of the Copyright in the Digital Single Market (CDSM) Directive. Text and data mining exceptions are applicable to the training of generative models. The provisional compromise text of the EU AI Act (as of April 2024) proposal has also recognised this. These are mandatory exceptions or limitations, with research organisations and cultural heritage institutions as the primary beneficiaries under Art. 3 of the CDSM Directive. In addition to the context of scientific research, both private and public entities use text and data mining techniques to analyse large amounts of data for various purposes. For this reason, Art. 4 of the CDSM Directive allows the reproduction and extraction of lawfully accessible works for text and data mining purposes. This is a so-called general provision that also allows commercial use. Reproductions and extractions in this regard may be kept for as long as needed for text and data mining.

The difference between Art. 4 and Art. 3 is that Art. 4 provides an exception or limitation that allows copyright holders to make a reservation in an acceptable manner, such as machine-readable means in the case of publicly available online work. It is called "opt-out mechanism".

As a result, all copies made in the process of machine learning are excused within the conditions of Art. Article 3 and 4 of the CDSM Directive. However, it is important to note that the exception for text and data mining may have limitations due to the narrow scope of beneficiaries outlined in Article 3 and the strict conditions associated with the opt-out mechanism detailed in Article 4.

The European Union is dealing with the complex area of technology governance through a combination of regulatory measures, including the Data Governance Act (DGA), Data Act (DA), Digital Services Act (DSA), and, most notably for this discussion, the EU AI Act.

Under the EU AI Act compromise agreement (as of April 2024), generative AI models are considered as general-purpose AI models. With regards to copyright, the compromise agreement stipulates that providers of general-purpose AI models must implement a policy that adheres to Union copyright law.

Additionally, they must make a sufficiently detailed summary of the content used to train the general-purpose AI model publicly available, based on a template provided by the AI Office. The summary should be comprehensive at a general level without being overly technical.

Moving forward, these regulations aim to balance fostering innovation in AI development with upholding the principles of copyright protection. It recognises the significance of transparency regarding the data used to train AI models, thereby addressing concerns surrounding copyright infringement and ensuring accountability within the AI ecosystem.

4. Conclusion

In conclusion, the intersection of generative AI and copyright law presents a multifaceted challenge for legal systems worldwide. While the existing copyright legal framework offers some guidance, it is clear that the rapid advancement of AI technology requires a re-evaluation of traditional legal frameworks. The notion of granting copyright over AI-generated material raises important questions about "authorship" and the concept of "creativity". Also, the rise of a new creative market based on AI, along with traditional creative industries, makes human artists feel uncomfortable.

One important thing to keep in mind is that copyright might not be the only or even the most suitable way to regulate creation with AI. We should explore alternative approaches, such as licencing schemes, moral rights protection, or even the creation of entirely new legal concepts, to maximise the benefits of AI technology and minimise potential negative consequences.

Bibliography

Agarwal, S., (2023) 'Audiobook Narrators Fear Apple Used Their Voices to Train AI'. *WIRED*, 14 February. Available at: https://www.wired.com/story/apple-spotify-audiobook-narrators-ai-contract/.

Dixit, P. (2023) 'Meet The Trio Of Artists Suing AI Image Generators'; *Buzz Feed News*, 20 January. Available at: https://www.buzzfeednews.com/article/pranavdixit/ai-art-generators-lawsuit-stable-diffusion-midjourney.

Drexl, J., et. al. (2019) Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective. (October 8, 2019). Max Planck Institute for Innovation & Competition Research Paper No. 19-13. Available at: https://doi.org/https://doi.org/10.2139/ssrn.3465577

Getty Images (2023) Getty Images Statement: Available at: https://newsroom.gettyimages.com/en/getty-images/getty-images-statement.

Kop, M. (2020) 'Machine Learning & EU Data Sharing Practices'. *Transatlantic Antitrust and IPR Developments*, 1(20).

Kop, M. (2020) *Shaping the Law of AI: Transatlantic Perspectives*. Stanford: Stanford Law School; Vienna: University of Vienna School of Law: Stanford.

Kop, M (2020) 'Regulating Transformative Technology in The Quantum Age: intellectual property, standardization & sustainable innovation', *Transatlantic Antitrust and IPR Developments,* 2(2020). Schönberger, D. (2018) 'Deep Copyright: Up-and Downstream Questions Related to Artificial Intelligence (AI) and Machine Learning (ML)' in .De Werra, J. (ed.) *Droit d'auteur 4.0*. Geneva, Zürich: Schulthess Editions Romandes, pp. 145-173. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3098315

Schuett, J. (2023) 'Defining the scope of AI regulations'. *Law Innovation Technology*, 15(1), pp. 60–82. Available at: https://doi.org/10.1080/17579961.2023.2184135.

Wingström, R., Hautala, J., Lundman, R., (2022) 'Redefining Creativity in the Era of AI? Perspectives of Computer Scientists and New Media Artists', *Creativity Research Journal*, 36(2), pp. 177-193. Available at: https://doi.org/10.1080/10400419.2022.2107850.

Chapter XI
Artificial Intelligence and Healthcare

by Malwina Anna Wójcik-Suffia*

Index: 1. The opportunities and risks of "black-box medicine". – 2. Regulating AI in healthcare – the EU perspective. – 3. AI in healthcare and the four principles of bioethics. – 4. Conclusion.

1. The opportunities and risks of "black-box medicine"

The integration of Artificial Intelligence (AI) into the medical field marks a transformative era in healthcare, public health, and biomedical research. Firstly, machine learning (ML) algorithms are increasingly used in patient care for diagnosis, prognosis, and treatment recommendations. Secondly, AI is facilitating effective planning and management of healthcare systems and public health surveillance. Thirdly, advancements in AI technologies offer benefits for health research and drug development. For instance, AI has enabled the prediction of protein folding, a key element of drug discovery (WHO, 2021).

The latest developments in healthcare AI involve large foundation models, which are trained on broad data at scale and subsequently adapted to a variety of clinical and administrative tasks, including analysis of different modalities of data, information retrieval, medical chatbots, and education tools (WHO, 2024).

Unfortunately, the operation of certain ML systems in healthcare raises transparency and safety concerns, as they fall into the category of "black box medicine" (W. Nicholson Price II 2017). As opposed to traditional, knowledge-based systems which operate on the rules of logic, ML algorithms work by drawing complex correlations between huge amounts of data. These correlations are not always clearly identifiable and explainable, making algorithmic decision-making inherently opaque. For instance, researchers discovered that ML models can detect self-reported race from medical images with a high degree of accuracy. However, research is not yet able to explain how the system reaches these conclusions (Banerjee *et al.*, 2022). Moreover, since it is very common for ML algorithms to find correlations between numerous data points, it might not be feasible to confirm them in clinical trials (Ferretti *et al.*, 2018). Finally, the

* PhD Researcher at the University of Bologna and the University of Luxembourg. Research Fellow at the Information Society Law Center (ISLC).

developers of AI might refuse to disclose how the model works to protect their intellectual property rights.

This chapter offers a review of the legal and ethical challenges associated with the use of black-box medicine. It starts with a brief introduction to the key legal instruments regulating healthcare AI in the EU. Following the legal overview, the chapter adopts a bioethical perspective to scrutinise the intricacies involved in deploying AI in healthcare settings. It does so by analysing the impact of AI on the four principles of bioethics: beneficence, non-maleficence, autonomy and justice (Beauchamp and Childress, 2019).

2. Regulating AI in healthcare – the EU perspective

In the EU, AI applications in healthcare are governed by a complex regulatory framework. Its core elements include:

- the General Data Protection Regulation (GDPR);
- the Medical Devices Regulation (MDR);
- the AI Act;
- the European Health Data Space (EHDS).

2.1 The General Data Protection Regulation

The GDPR plays a key role in regulating health data which constitute the lifeblood of medical AI. According to Art. 9 of the GDPR, genetic data, biometric data, and data concerning health fall into the category of sensitive data. Thus, their processing is permissible only based on specific grounds. These include, among others, patient consent, public interest in the safety of medicinal products or medical devices, and scientific research. Moreover, further processing of data collected for the purpose of providing healthcare is possible in the case of scientific research. However, there are two important caveats. Firstly, the extent to which the commercial development of AI technologies in healthcare can be considered scientific research under the GDPR is debatable and is likely to depend on the presence of substantial public interest (Meszaros and Ho, 2021). Secondly, it is crucial to note that GDPR allows Member States to introduce further restrictions concerning the processing of genetic data, biometric data or data concerning health, potentially obstructing the re-use and cross-border flow of health data.

2.2 The Medical Device Regulation

AI tools that are either independent software or an accessory to a medical device (*e.g.,* software for a wearable device) fall within the scope of the Medical Device Regulation if they are intended, among others, for "diagnosis, prevention,

monitoring, prediction, prognosis, treatment or alleviation of disease (Art. 2 MDR)." Medical devices need to undergo a conformity assessment procedure to confirm their safety before being placed on the market, and they are subject to rigorous post-authorisation oversight. The level of scrutiny depends on the classification of medical devices into four categories: I (low risk), IIa (moderate risk), IIb (medium risk), and III (high risk). The level of potential risk is based on the device's intended purpose. While class I requires only a self-assessment by the manufacturers, classes IIa, IIb, and III require an intervention by a notified body.

Software that is driving a medical device falls within the same class as the device itself. Standalone software is classified as low risk unless it is used for medical diagnosis, therapy, or to monitor physiological processes. In these cases, it falls under class IIa (moderate risk), IIb (medium risk), or III (high risk), depending on its function and the level of possible harm to the patient. Software will be classified as high risk if it can cause death or an irreversible deterioration of a person's state of health.

2.3 The AI Act

Many applications of AI in healthcare will be classified as high-risk systems under the AI Act, triggering requirements about risk management, quality of data requirements, technical documentation, transparency and provision of information to users, quality management systems, human oversight, robustness, accuracy, and cybersecurity.

There are two ways in which an AI system can be classified as high risk under the AI Act.

Firstly, a system will fall within the high-risk category if it is a product (or a safety component of a product) subject to third-party *ex ante* conformity assessment under EU harmonisation legislation listed in Annex I, which includes the MDR. Put simpler, software that is qualified as a medical device of class IIa, IIb or III (moderate to high risk according to the MDR) will be qualified as a high-risk system under the AI Act. Thus, these systems will need to comply with both the MDR and the AI Act.

Secondly, a system will fall within the high-risk category if it is a standalone system listed in Annex III. In the healthcare domain, these systems include systems intended:

- to be used by public authorities or on behalf of public authorities to evaluate individual eligibility for healthcare services and benefits,
- to dispatch emergency medical aid and emergency triage systems; and
- for risk assessment and pricing determination for life and health insurance.

These systems will always be considered high risk if they perform profiling of natural persons. Otherwise, they can be deemed non-high-risk if they are intended to perform certain simple administrative or supervisory tasks

2.4 The European Health Data Space

The EHDS deals specifically with data concerning health, governing their primary and secondary use. The EHDS is relevant for regulating medical AI for two main reasons.

Firstly, it introduces requirements aimed at increasing the availability of health data. It does so by creating trustworthy mechanisms for the re-use of health data to facilitate, among others, research and innovation in the field of medical AI.

Secondly, the EHDS introduces quality, safety, and interoperability requirements concerning electronic health records (EHRs), a major source of health data. The regulation states that EHRs will need to comply with common specifications elaborated by the EU Commission.

Moreover, medical devices and high-risk AI systems under the AI Act that claim compatibility with EHRs will also need to comply with selected requirements under the EHDS.

3. AI in healthcare and the four principles of bioethics

In highly connected environments in which black-box medicine is deployed, it is useful to refer to the four well-established principles of bioethics (beneficence, non-maleficence, autonomy, and justice) to guide the ethical and legal assessment of challenges posed by AI.

a) Beneficence

Beneficence is the first core principle of bioethics. It requires healthcare professionals to "do good", acting in the best interest of the patient. The list of potential benefits of medical AI for patient care and the organisation of healthcare systems is long. AI solutions can contribute to a quicker diagnosis and effective risk assessment, allowing healthcare systems to implement preventive care to reduce the burden of disease and healthcare spending. Advancements in AI-enabled personal mobile devices and apps monitoring the health of patients increase their safety and comfort of living. The analysis of vast amounts of medical data leads to the development of increasingly personalised medicine. Last but not least, the automatisation of repetitive administrative tasks and improving clinical workflow allows doctors to focus on the human side of medicine – presence, empathy, trust, and caring (Eric Topol, 2019).

However, the beneficial effects of AI are not equally distributed among society. The digital divide in healthcare technologies is present both within states and globally.

On a domestic level, the low digital literacy of vulnerable patient groups, such as the elderly or immigrant communities, often prevents them from seizing the benefits of AI solutions. Moreover, the spatial accessibility of state-of-the-art

medical AI, especially in the public healthcare systems, is limited to well-served, urban areas. This often results in excluding patients coming from lower socio-economic backgrounds.

On a global level, resource constraints and the lack of digital medical infrastructure make lower and middle-income countries less likely to fully benefit from AI technologies. At the same time, AI's potential to revolutionise healthcare is the greatest in these areas.

In Europe, some measures aimed at increasing universal access to AI benefits can be highlighted. For example, in Germany, a system of reimbursement for "apps on prescription" has been introduced to enable more equitable access to technology.

The problems of inequitable distribution of benefits of AI are closely related to the principle of justice, explained in the paragraphs that follow.

b) Non-maleficence

The principle of non-maleficence entails the duty not to harm the patient, as expressed in the Hippocratic Oath itself – "first, do no harm".

Despite their potential benefits, medical AI systems can still exhibit risks of errors and inaccuracies, leading to failed or misguided interventions. Firstly, noise in the input data, such as poor quality of medical images or inconsistent labelling, can cause AI tools to draw incorrect conclusions. Secondly, AI is particularly prone to memorizing spurious correlations. For example, deep learning systems detecting Covid-19 from chest X-rays often yield incorrect diagnoses because they associate the presence of the disease with clinically irrelevant factors (De Grave *et al.*, 2021). Thirdly, AI systems do not generalize well. In other words, highly accurate ML algorithms trained on a dataset coming from hospitals X and Y can yield inadequate results when applied to a new set of data in hospital Z (Zech *et al.*, 2018).

Misguided algorithmic recommendations can lead clinicians to commit medical errors, causing harm to patients. At the same time, the legal regimes of medical negligence and product liability are not well suited to address the complex, multi-actor problem of AI-enabled decision-making in healthcare. In particular, the allocation of liability between doctors, healthcare providers and developers of AI tools remains an open question. Although the proposed Product Liability Directive (PLD) and AI Liability Directive (AILD) constitute a crucial step toward developing a unified European approach to liability issues, they fail to protect patients when harm occurs due to the non-interpretable character of the AI system (Duffourc and Gerke, 2023).

As discussed, the lack of explainability is an inherent feature of some AI models. Thus, it does not constitute a defect under the PLD's strict liability rules. Similarly, it cannot be considered a fault of the manufacturer or healthcare provider under the AILD. The lack of a coherent liability framework for

AI-perpetuated harm in medicine can lead to stifling innovation in healthcare, as healthcare professionals avoid implementing AI solutions for fear of liability.

Another harm is associated with the potential misuse of AI systems in healthcare. Firstly, the majority of commercially available m-Health apps are not qualified as medical devices, and thus they do not undergo rigorous testing. Yet, many patients rely on them for diagnosis and medical advice despite their questionable accuracy (Lekadir, 2022). Secondly, the lack of sufficient education and training results in low technical literacy of healthcare professionals. This makes doctors and healthcare providers prone to automation and confirmation biases, showing a preference for algorithmic decisions without properly reviewing them. For instance, one of the main healthcare insurers in the US has been recently sued for prematurely terminating coverage based on the recommendation of an algorithm and contrary to the advice of healthcare professionals (Barrows *et al. v.* Humana, Inc. (3:23-cv-00654) Kentucky Western District Court; Estate of Gene B. Lokken *et al. v.* UnitedHealth Group, Inc. et al. (0:23-cv-03514) Minnesota District Court).

Finally, the use of AI in healthcare is connected with considerable privacy and cybersecurity risks. Potential leaks of data relating to health can expose patients to various forms of cyberattacks. The misuse of healthcare data could lead to serious consequences for data subjects, such as denial of insurance or job opportunities based on one's state of health. While data concerning health are rightfully protected as sensitive data under the GDPR, access to them is also necessary for the training and validation of AI systems. Thus, Art. 10 of the AI Act allows developers of AI to process sensitive data subject to appropriate safeguards and to the extent it is strictly necessary to detect and correct bias. Moreover, the EHDS Proposal enables the sharing of healthcare data for secondary purposes, including the training of medical AI. However, privacy requires that patients understand and consent to the re-purposing of their health data. For instance, in 2016, a scandal was caused by the transfer of data of 1,6 million UK patients to Google's DeepMind to develop an AI kidney disease detection system. The transfer took place without the data subjects' consent, violating patient privacy (Lekadir, 2022).

The AI Act acknowledges and addresses the potential harms of AI. For instance, Art. 14 states that providers and deployers of high-risk systems should ensure that natural persons who oversee their functioning are aware of automation and confirmation bias. Moreover, Art. 15, introduces requirements regarding accuracy, robustness, and cybersecurity.

c) Autonomy

The principle of autonomy entails the right to freely decide about one's medical treatment and participation in research. A central component of autonomy is informed consent to treatment, often described as the cornerstone of medical

ethics. The right to informed consent is protected by domestic and supranational law. The latter includes the Council of Europe's Oviedo Convention on Human Rights and Biomedicine (Art. 5) and the EU's Charter of Fundamental Rights (Art. 3). However, the application of the right in the context of AI poses many doubts explored by practitioners and scholars.

The first key debate concerns the doctor's duty to disclose the fact that a diagnosis was made with the help of AI. Interpretations of the disclosure question might vary under domestic law. Arguably, the more prevalent AI solutions become in healthcare, the less controversial the fact of non-disclosure becomes. After all, informed consent does not require doctors to voluntarily disclose all the sources of their decision, such as practical experience or the books they have consulted. Therefore, it could be argued that it is unlikely that the disclosure is mandatory unless a patient explicitly enquires about the involvement of AI (Glen I. Cohen, 2019). However, we might imagine specific fact patterns that point in the direction of disclosure.

Firstly, disclosure can be considered mandatory when AI plays a substantial role in decision-making. On the EU level, Art. 22 of the GDPR enshrines the right of data subjects, including patients, not to be subject to a decision "based solely on automated processing" and a corresponding right to be informed when such a decision is made. However, it's important to underline that decisions based solely on algorithmic output are very rare in the medical domain. Thus, a degree of human involvement in decision-making could prevent patients from relying on the GDPR to support a duty to disclose the use of medical AI.

Nevertheless, it could be argued that such a duty would exist under domestic informed consent law in cases in which AI plays a substantial role in the decision process. This is especially true in jurisdictions in which the appropriate legal test of informed consent is patient-centric, as opposed to doctor-centric. Studies of patient attitudes toward AI decision-making suggest that they tend to find automated medical diagnostics "dehumanising" (Formosa *et al.*, 2022). Thus, it is likely that a reasonable patient would like to know whether a key recommendation concerning her or his treatment was formulated by a machine.

Secondly, a duty to disclose might exist when the use of AI would be analogous to an experimental treatment (Glen I. Cohen, 2019). That could be the case if healthcare professional lacks certainty about the accuracy of the tool. For instance, an AI system that detects cancers accurately for a subset of the population on which it was tested, might not be effective when run on different demographics. The doctor should therefore inform the patient when the potential risk of misdiagnosis is high. Thus, it could be argued that the duty to disclose the use of AI to a patient would be conditioned on the doctor possessing enough evidence about the performance of the system in question. To this end, different jurisdictions have proposed transparency solutions enabling healthcare professionals to access basic information about the system's functioning.

For instance, Art. 13 of the AI Act puts certain transparency obligations on providers of high-risk systems, enabling users to receive basic information about the system's functioning. Similarly, in the US, a new executive rule holds that healthcare IT should be accompanied by information that enables clinical users to determine its utility in a given context.

Thirdly, a duty to disclose the use of AI might exist when the system is designed to optimise its decisions for a specific variable, such as cost-effectiveness. (Glenn I. Cohen, 2019). This would potentially limit a patient's right to choose between alternative treatments.

The debate on informed consent gets complex once a patient starts asking more specific questions about the diagnosis or recommendation. Informed consent requires the medical professional to ensure the patient can interpret the information to make a fully informed choice about the treatment. However, physicians who use AI for decision support might face difficulties in explaining the basis of the algorithmic decision to patients due to algorithmic opacity.

Both computer science and law aim to address the problem of black-box medicine, dealing with the challenge of adjusting the explanation mechanisms to the needs of different stakeholders. From the technical point of view, researchers working in the field of explainable AI (XAI) elaborate different methods to interpret the outputs of algorithmic systems (Chaddad *et al.*, 2023). From the legal point of view, scholars and legislators discuss the right to explanation of decisions taken with the use of AI. Although Art. 15 of the GDPR contains the right to request "meaningful information about the logic involved" in fully autonomous decisions, scholars argue that the right does not extend to rationales for specific decisions. Rather, it focuses on general system functionality (Wachter *et al.*, 2017). In addition to the GDPR, the AI Act introduces a right to explanation of decisions taken with the use of high-risk AI in certain healthcare contexts. However, the practical implications of this provision remain to be explored.

While a lot of scholarly attention in the context of medical AI is devoted to the right to know, it should be underlined that autonomy also comprises the right not to know (Andorno, 2004). Such right is often invoked in genetic cases when patients express the will not to be informed about potential hereditary and incurable diseases. The pervasiveness of AI-enabled medical screening technologies could violate the right not to know, potentially leading to overdiagnosis, increased stress, and anxiety.

Lastly, it is important to mention that AI-powered developments in neurotechnology, including brain-computer interfaces enabled through external or implantable medical devices, pose serious risks to human autonomy. As pointed out by researchers, the processing of "mental data" creates a danger of third-party intrusion and manipulation of the mental sphere (Ienca and Malgieri, 2022).

d) Justice

The principle of justice requires a fair distribution of burdens and benefits in the healthcare system. Put simply, patients should not face discrimination in access to and provision of healthcare. They should also be able to freely participate in medical research and benefit from its results.

Unfortunately, using AI to guide decisions in both diagnostic and resource allocation scenarios disrupts the core tenets of justice. AI has the propensity to encode and perpetuate data gaps and biases concerning vulnerable patient groups, leading to unlawful discrimination.

Firstly, the datasets on which AI systems are trained and validated often lack reliable data on groups historically excluded from research. A leading example is women, whose bodies have been largely understudied despite evidence of numerous sex differences in organ anatomy and disease patterns (Caroline Criado Perez, 2020). The data gaps contribute to the misdiagnosis of many common conditions, such as stroke. Consequently, AI systems deployed in cardiology are likely to work less accurately for female patients (Fosch-Villaronga *et al.*, 2022). Similarly, skin cancer detection systems that do not take into account sex and gender-related characteristics risk reinforcing bias (Lee *et al.*, 2022).

Secondly, even if the data used to train and validate the algorithm includes information on vulnerable groups, it might still be tainted by harmful stereotypes. A good example is the use of race-adjusted clinical algorithms. This practice, spanning across various medical disciplines, entails differential diagnosis of patients based on their race (Vyas *et al.*, 2020). For instance, in nephrology, the eGFR formula used to measure kidney function based on the level of creatinine is designed to return higher scores for Black patients. A high score is associated with better kidney function, potentially resulting in de-prioritisation of Black patients in the prevention and treatment of nephrological diseases. The rationale for race correction in eGFR is based on the observation that the level of creatine is typically higher in Black patients. This in turn is often explained by Black people being allegedly more muscular.

Even if adopted with equity and accuracy in mind, many of the race-correction algorithms are based on the flawed premise of biological differences between races. Race, however, has been proven to be a socio-cultural concept, distinct from genetic ancestry (Malinowska and Żuradzki, 2022). Thus, many of the presumed differences between White and non-White patients stem from the legacy of racism and eugenics which sought to portray Black bodies as inferior. The race-based medicine led to the development of stereotypes such as the thickness of Black skin, higher pain tolerance, or even the difference in brain size. Researchers show that some of these harmful premises are replicated by state-of-the-art LLMs in healthcare (Omiye *et al.*, 2023).

Thirdly, it is important to note that training algorithms on data that adequately represent reality can still lead to unfair decisions when this reality

reflects inequitable access to healthcare. Using historic data often means that instead of moving towards more fair and equitable decisions, AI systems simply embrace the status quo, fuelling the vicious circle of exclusion. A real-life example is provided by the Impact Pro algorithm which was used in the US to identify patients with complex health problems, suitable for high-risk care management. The algorithm used healthcare spending as a proxy for illness and falsely attributed a lower risk of serious disease to Black patients (Obermeyer *et al.*, 2019). Note that unlike the clinical algorithms described above, this AI system was race blind. Nevertheless, the data reflected unequal healthcare access experienced by people of colour, leading to the replication of discriminatory outcomes.

Regrettably, EU antidiscrimination law is often ill-suited to address algorithmic discrimination in healthcare (Wójcik, 2022). Firstly, although discrimination can occur on many different grounds, the material scope of EU antidiscrimination law is limited. The anti-discrimination directives applicable in the context of healthcare protect patients on just three grounds: race, ethnic origin and sex (Di Federico, 2017). Article 21 of the Charter of Fundamental Rights, which contains an open list of non-discrimination grounds only applies to Member States when they implement EU law. Since organisation and delivery of healthcare is a primary competence of Member States, the applicability of the Charter will likely be limited to cross-border healthcare scenarios (Di Federico, 2017). Secondly, the EU antidiscrimination law does not protect against discrimination on multiple combined grounds, which is often perpetuated by algorithms (Gerards and Xenidis, 2021). Thirdly, the enforcement mechanisms of antidiscrimination law are highly fragmented (Di Federico, 2017). Fourthly, patients are likely to face difficulties detecting and proving discriminatory treatment, especially in the case of black-box algorithms.

The AI Act contains provisions aimed at reducing the risk of algorithmic bias. Most importantly, Art. 10 addresses bias in training data by introducing quality criteria for training, validation and testing of data sets in data-driven high-risk systems. These data sets must be, *inter alia*, examined for possible biases, relevant, representative, free of errors, complete and contextual, that is trained, validated and tested in a particular geographic, behavioural or functional setting. Moreover, Art. 15 addresses feedback loop bias, providing that high-risk algorithms which continue to learn after being placed on the market, should be accompanied by appropriate bias mitigation measures throughout their life cycle.

Furthermore, the EHDS can contribute to debiasing medical algorithms by increasing the diversity of datasets through data sharing mechanisms.

4. Conclusion

This chapter aimed to systematise the numerous ethical and legal concerns posed by the deployment of AI models in healthcare through the lens of the key principles of bioethics. However, it is important to underline that the complexity of the ethical and regulatory landscape of black-box medicine, can lead to conflicts between the bioethical principles, creating a need to balance competing interests.

For instance, the negotiations concerning the data-sharing mechanisms under the EHDS proposal illustrate the conflict between justice and autonomy. On the one hand, the availability of high-quality data is a prerequisite for creating fairer AI systems. On the other hand, the respect for privacy and informed consent points towards the introduction of an opt-out mechanism for patients who do not want their data to be shared.

Another conflict can arise between justice and non-maleficence in the context of de-biasing algorithms, as researchers acknowledge that there is a trade-off between the fairness and accuracy of AI models (Schönberger, 2019). Thus, imposing fairness constraints on the model can decrease its overall performance.

As a final example, certain uses of AI in mental health can result in a conflict between autonomy and beneficence. For instance, the AI Act carves a therapeutic exception for the general prohibition of algorithmic manipulation and persuasion based on explicit informed consent of patients. While subliminal techniques could offer a benefit to patients in a clinical setting, there is a risk that the therapeutic exception can be abused with a detriment to patient autonomy.

These examples make clear that navigating the ethical and regulatory maze of AI in health requires a collaborative effort among lawyers, policymakers, computer scientists, healthcare professionals, and patients to ensure that AI technologies appropriately balance patient benefit, autonomy, and equity concerns.

Bibliography

Andorno, R. (2004) 'The right not to know: an autonomy based approach', *Journal of medical ethics*, 30(5). Available at: https://doi.org/10.1136/jme.2002.001578.

Banerjee, I. *et al.* (2022) 'Reading Race: AI Recognises Patient's Racial Identity In Medical Images', *The Lancet Digital Health*, 4(6), pp. e406–e414. Available at: https://doi.org/10.1016/S2589-7500(22)00063-2

Beauchamp, T.L. and Childress, J.F. (2019) *Principles of biomedical ethics*. Eighth edition. New York: Oxford University Press.

Chaddad, A. *et al.* (2023) 'Survey of Explainable AI Techniques in Healthcare', *Sensors*, 23(2), p. 634. Available at https://doi.org/10.3390/s23020634.

Cohen, I.G. (2019) 'Informed Consent and Medical Artificial Intelligence: What to Tell the Patient? Symposium: Law and the Nation's Health', *Georgetown Law Journal*, 108(6), pp. 1425–1470.

Criado-Perez, C. (2020) *Invisible women: exposing data bias in a world designed for men*. London: Vintage.

Di Federico, G. (2017) 'Access to Healthcare in the European Union: Are EU Patients (Effectively) Protected Against Discriminatory Practices?', in L.S. Rossi and F. Casolari (eds). *The Principle of Equality in EU Law*. Springer Publishing.

Duffourc, M.N. and Gerke, S. (2023) 'The proposed EU Directives for AI liability leave worrying gaps likely to impact medical AI', *npj Digital Medicine*, 6(1), pp. 1–6. Available at: https://doi.org/10.1038/s41746-023-00823-w.

Ferretti, A., Schneider, M. and Blasimme, A. (2018) 'Machine Learning in Medicine: Opening the New Data Protection Black Box', *European Data Protection Law Review (EDPL)*, 4(3), pp. 320–332.

Formosa, P. *et al.* (2022) 'Medical AI and human dignity: Contrasting perceptions of human and artificially intelligent (AI) decision making in diagnostic and medical resource allocation contexts', *Computers in Human Behavior*, 133, p. 107296. Available at: https://doi.org/10.1016/j.chb.2022.107296.

Fosch Villaronga, E. *et al.* (2022) 'Accounting for diversity in AI for medicine', *Computer Law & Security Review*, 47, p. 105735. Available at: https://doi.org/10.1016/j.clsr.2022.105735.

European Commission, Directorate-General for Justice and Consumers, Gerards, J., Xenidis, R. (2021), *Algorithmic discrimination in Europe – Challenges and opportunities for gender equality and non-discrimination law*. Publications Office. Available at: https://data.europa.eu/doi/10.2838/544956.

Ienca, M. and Malgieri, G. (2022) 'Mental data protection and the GDPR', *Journal of Law and the Biosciences*, 9(1), p. lsac006. Available at: https://doi.org/10.1093/jlb/lsac006.

Lee, M.S., Guo, L.N. and Nambudiri, V.E. (2022) 'Towards gender equity in artificial intelligence and machine learning applications in dermatology', *Journal of the American Medical Informatics Association*, 29(2), pp. 400–403. Available at: https://doi.org/10.1093/jamia/ocab113.

Lekadir, K. *et al.* (2022) *Artificial intelligence in healthcare: applications, risks, and ethical and societal impacts*. Brussels: European Parliament.

Malinowska, J.K. and Żuradzki, T. (2022) 'Towards the multileveled and processual conceptualisation of racialised individuals in biomedical research', *Synthese*, 201(1), p. 11. Available at: https://doi.org/10.1007/s11229-022-04004-2.

Meszaros, J. and Ho, C. (2021) 'AI research and data protection: Can the same rules apply for commercial and academic research under the GDPR?', *Computer*

Law & Security Review, 41, p. 105532. Available at : https://doi.org/10.1016/j.clsr.2021.105532.

Obermeyer, Z. *et al.* (2019) 'Dissecting racial bias in an algorithm used to manage the health of populations', *Science (New York, N.Y.)*, 366(6464), pp. 447–453. Available at: https://doi.org/10.1126/science.aax2342.

Omiye, J.A. *et al.* (2023) 'Large language models propagate race-based medicine', *npj Digital Medicine*, 6(1), pp. 1–4. Available at: https://doi.org/10.1038/s41746-023-00939-z.

Price, W.N.I. (2017) 'Regulating Black-Box Medicine', *Michigan Law Review*, 116(3), pp. 421–474.

Schönberger, D. (2019) 'Artificial intelligence in healthcare: a critical analysis of the legal and ethical implications', *International Journal of Law and Information Technology*, 27(2), pp. 171–203. Available at: https://doi.org/10.1093/ijlit/eaz004.

Topol, E.J. (2019) *Deep medicine: how artificial intelligence can make healthcare human again*. First edition. New York: Basic Books.

Vyas, D.A., Eisenstein, L.G. and Jones, D.S. (2020) 'Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms', *New England Journal of Medicine*, 383(9), pp. 874–882. Available at: https://doi.org/10.1056/NEJMms2004740.

Wachter, S., Mittelstadt, B. and Floridi, L. (2017) 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation', *International Data Privacy Law*, 7(2), pp. 76–99. Available at: https://doi.org/10.1093/idpl/ipx005.

Wójcik, M.A. (2022) 'Algorithmic Discrimination in Health Care: An EU Law Perspective', *Health and Human Rights*, 24(1), p. 93.

World Health Organization (2021) *Ethics and governance of artificial intelligence for health: WHO guidance*. Geneva: World Health Organization. Available at https://apps.who.int/iris/handle/10665/341996.

World Health Organization (2024) *Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models*. World Health Organization. Available at https://www.who.int/publications-detail-redirect/9789240084759.

Zech, J.R. *et al.* (2018) 'Variable generalization performance of a deep learning model to detect pneumonia in chest radiographs: A cross-sectional study', *PLoS medicine*, 15(11), p. e1002683. Available at https://doi.org/10.1371/journal.pmed.1002683.

PART III
SMART CITIES AND DIGITAL SOCIETY

Chapter XII
The Idea and Creation of a Smart City

by Gabriele Suffia*

Index: 1. The growing importance of cities and the perception of cities. – 2. The idea of smart cities. – 3. Smart city criticism. – 4. Smart Cities and legislation. – 5. Future research. – 6. Conclusions.

1. The growing importance of cities and the perception of cities

Over the last century, the urban population has increased dramatically. This significant growth positioned the cities as central hubs of economic activity, cultural exchange, and innovation. The concentration of resources, talent, and infrastructure in urban areas not only fuelled economic development, but also presented unique challenges.

According to many scholars (O'Brien *et al.*, 2019), these challenges include, but are not limited to, escalating urban poverty, soaring costs associated with urban living, severe traffic congestion, a pronounced shortage of housing, insufficient investment in urban infrastructure, limited financial and governance capabilities within urban settings, increasing social inequality and crime rates, and environmental degradation.

In addressing these multifaceted issues, the concept of "smart cities" emerged at the end of the last century as a plausible solution with technology and automation at its centre. After many decades, the term "smart city" remains a popular umbrella term for "a technologically modern urban area that uses different types of electronic methods and sensors to collect specific data", as this brief definition in Wikipedia tries to explain. However, the complexity of the urban dimensions is more and more evident in many circumstances, as both scholars and legislation are facing what a "smart" city means.

The debate is crucial not only to deliver better public services, or better allocate resources, but also from a security perspective. As an unwanted "part" of the digitization, for example, in recent years several cities experienced ransomware attacks which caused significant disruptions to their operations. These attacks can cripple the infrastructure of cities, affecting services like emergency response systems, utility services, and municipal functions.

* PhD Candidate at the University of Bologna and University of Luxembourg.

This chapter will discuss the idea of a smart city, present the relevant legislation at the European level and offer insights on the most up-to-date research on the topic.

2. The idea of smart cities

As more individuals migrate to urban areas in search of better opportunities, cities have become pivotal in shaping economic trends and driving technological advancements.

The first way to see the origin of smart cities, consequently, is by applying technological enhancements to city management and life. The necessary urban planning and management to sustainably accommodate the growing urban populace wouldn't have been possible without a growing usage of technology, especially digital computing. In the same way, new technologies are described as fundamental tools to improve the quality of life for a city's inhabitants.

Automation is fundamental to the current concept of smart cities. Although this notion appears to be a recent development, early indications of such "smartness" were evident in proposals like the "Magic Motorways", sponsored by General Motors and presented at the 1939 New York World's Fair. The impressive "Futurama" diorama, designed by Norman Bel Geddes, envisioned a future world (at the time, expected for 1960) in which automated highways stretched between skyscrapers and sprawling suburban landscapes. The powerful idea behind "Futurama", highlighting the long-standing vision of integrating automation into urban and transportation planning, is still present in our perception of smart cities. The features of the city presented are quintessential to our contemporary notion of smart cities, highlighting the critical role of technology in urban development and efficiency.

The "Futurama" concept planted seeds within the collective imagination, followed and integrated by many examples of this "smartification": dashboards, control rooms and human-machine integrations populated the (sci-fi) imaginary for decades in the late 1900s, particularly in the US and Europe. The visions from 1939, in many respects, are not too dissimilar from the futuristic ideas that have been revisited and reimagined in numerous subsequent contexts. And, while at that time the term "Magic Motorways" was used, the digital era today is often still shrouded in an aura of the unknown, bordering on what could almost be described as magical.

The reflection on the smartification of cities developed theories and saw many phases.

A brief history of smart cities is attempted by G. R. Halegoua for The MIT Press (2020), starting in the 1950s. It offers an insightful exploration into the evolution of smart cities, highlighting the promises of smart cities in managing

public life and enhancing its predictability through sensors and Big data (in particular, the predictability of trends, of future urban activities and conditions).

From the mere analysis of "urban dynamics" (1969 work by Jay Forrester, in Halegoua, 2020) to the possibility of responding, adapting and predicting users' needs and behaviours, Information and Communication Technologies (ICT) became more and more integrated into the physical urban environment.

The question arises: how do we evaluate the success of these integrations? How to measure this "success"?

Over the past two decades, city rankings have emerged as a pivotal tool for evaluating the allure of urban areas. These rankings provide a quantitative measure of a city's appeal, factoring in aspects such as liveability, innovation, and smart technology implementation. The metrics aim to offer valuable insights into the effectiveness of smart city initiatives and their impact on urban life.

A popular one, often referred later in defining smart cities, was the ranking elaborated in 2007 by the Centre of Regional Science (SRF) of the Vienna University of Technology (Giffinger *et al.*, 2007). It identified six areas that quickly became "key strategic action fields" for developers and practitioners: - Smart Government, - Smart Economy, - Smart Environment, - Smart Living, - Smart Mobility, - Smart People. Rankings define benchmarks and promote a vision in which success is obtained when reaching these benchmarks.

This technological empowerment of the cities is what is popular in the so-called "Smart from the Start Cities" and "Retrofitted Smart Cities". The firsts are cities that are built from scratch around benchmarks of efficiency and "smartness", like the Songdo International Business District, Seoul (a prominent example of what some authors describe as a new private high-tech city (Courmont and Le Galès, 2019)). The second ones are cities already existing, that are "retrofitted" to accommodate digital technologies.

This approach, often referred to as the "Corporate View" of smart cities (because it is often sponsored by private companies who sell tech solutions to city bodies)(Halegoua, 2020), emphasises the integration or realignment of cities with advanced technologies. This perspective embodies a top-down vision, where the transformation of urban spaces is driven by technological innovations either from the inception of city planning ("Smart from the Start") or by integrating technology into existing urban frameworks ("Retrofitted Smart Cities").

This model underscores a strategic approach to urban development, prioritising technological solutions to improve city living, infrastructure, and governance, instead of other types of solutions.

A famous example of the concept of the "corporate view" in smart city development is seen in Rio de Janeiro. The Brazilian city established one of the pioneering city "control rooms" in 2010. This facility, managed by IBM, was designed to integrate and oversee various urban systems and services. However,

this initiative has faced critique from numerous scholars who argue that the control room resulted in a dependency, or "lock-in," effect for the public administration concerning its relationship with IBM.

This critique highlights concerns about the autonomy of public entities in the governance of smart cities and raises questions about the long-term implications of such partnerships. The case of Rio de Janeiro serves as a cautionary tale, argues Halegoua, emphasising the need for balance between leveraging technological advancements and maintaining control over urban governance.

3. Smart city criticism

Smart city critics, however, frame the digitisation of cities differently. In reality, they focus the attention on how cities were always born around documents, dating back to early human experiences with the writing system. Documents were always part of human life, and the simple fact that they are now digitised, or produced and managed by machines, doesn't constitute a change in the paradigm of "documentality" (Ferraris, 2014, 2021), as described in the words of the Italian philosopher Maurizio Ferraris.

Cities are the structural organisation of human settlements and the technology is (just) necessary to process data. Following this approach, it's certainly true that our era presents characteristics different from the past, but in the sense that it's a revolution (Floridi, 2017) in the paradigm of "documentality". We now live in the so-called "infosphere" (Floridi, 2020), and cities are "naturally" smart cities because they are in the infosphere.

Smart city critics argue that the primary focus is to understand the city as a place where people live with the technology of their time, and not a place in which technology needs to be implemented to solve problems. This shift could change our perception of the problems triggered by urbanisation, reminding us that the solutions could also not be technological.

The experience of Bletchley Park, for example, presents a contrasting example often regarded as the first "smart community". This community-centric model focuses on leveraging the collective intelligence and expertise of individuals to solve complex problems, most notably during World War II for code-breaking efforts. This perspective diverges from the technology-centric approach seen in modern smart cities by placing a greater emphasis on people rather than technology. This approach suggests that the essence of a "smart" community lies not solely in its technological advancements but in its ability to harness the potential of its people to improve collective outcomes. It highlights the importance of human collaboration, knowledge sharing, and problem-solving capabilities in creating effective and efficient solutions (with, or without, the usage of technology). Technology, in fact, serves as a tool to enhance human capabilities and it's not a purpose in itself.

This path is explored in the so-called "Social cities" (Halegoua, 2020), which emphasise sociality and put people first. In the same way, many scholars are proposing to shift the idea of a smart city, from the city itself to its citizens. This passage, from the book "Smart Cities" by G. R. Halegoua clearly explains the point: "Approaching smart citizens as foundations for smart cities recognises citizen agency in addressing issues that governments and their tech partners attempt to ameliorate. This perspective re-envisions smart cities as spaces of collective intelligence and collaboration where people work together to address concerns beyond efficient services, such as education systems, poverty, housing, as well as ideas and issues generated from urban communities" (Halegoua, 2020).

4. Smart cities and legislation

To deal with the topic of smart cities and legislation, two important premises must be made: the existence of legal imprecisions and the existence of legal fragmentation. Legal imprecision pertains to the vagueness of the "smart city" term, which is reverberated in how the term is used in the legislation. Legal fragmentation refers to the lack of uniform legislation across countries, even with similar experiences and traditions. Definition is the first problem for a scholar, particularly a legal one. The definition of "Smart city" is at the centre of a debate that has lasted for decades. As reconstructed by scholars, IBM officially registered the term as its trademark in 2011 (Söderström *et al.*, 2014; Breuer and Pierson, 2021). The first usage seems to date back to 2007, as mentioned Centre of Regional Science (SRF) report by the Vienna University of Technology mentioned earlier. However, as Breuer and Pierson explain (2021), "Previously, scholars had been discussing concepts of wired city, digital city, telecommunications city, informational city, or intelligent city, cities and urban, all with similar connotations; announcing changes to urbanisation based on ICT for better or worse".

The inconsistent usage of the term has led to its limited usage in the legislation. Since 2013, the term has been more frequent in strategies and policy papers (it also entered the official Italian lexicon (Russo, Rindone and Panuccio, 2014).

Recent landmarks are the European strategy "Shaping Europe's digital future", which mention smart cities in its webpage "Smart Cities and Communities", for their aim "at the well-being of its inhabitants, businesses, visitors, organisations and administrators by offering digitally enabled services that contribute to a better quality of life" (website European Commission, "Smart Cities and Communities").

Another reference occurs with the European Commission initiatives on cities, which mention smart cities in its webpage "Smart Cities" as "a place

where traditional networks and services are made more efficient with the use of digital solutions for the benefit of its inhabitants and business. A smart city goes beyond the use of digital technologies for better resource use and less emissions. It means smarter urban transport networks, upgraded water supply and waste disposal facilities and more efficient ways to light and heat buildings. It also means a more interactive and responsive city administration, safer public spaces and meeting the needs of an ageing population" (website European Commission, "Smart Cities").

There is no doubt that the European debate on smart cities sees these words as a starting point in pondering and shaping the legislation.

It is worth mentioning that the European Commission is supporting the digital transformation of cities and communities through a variety of tools, like the Living-in.EU movement, which aims to expedite the digital transformation of cities and communities in a manner that prioritizes citizens. It emphasizes ethical and socially responsible data usage, co-creation, and citizen engagement, adhering to open and interoperable standards. Local data platforms, "Data Space for Smart Communities" and the "DIGITAL programme" are all initiatives that promote standards, facilitate data sharing across Europe and offer financial support supplementing the primary funding expected from national sources, including Cohesion Policy Funds or the Recovery and Resilience Facility.

A new field of research is opened also with "Local Digital Twins": they involve the development of virtual representations of physical assets, processes, and systems within cities and communities (see more later in Paragraph 5 of this chapter).

A relevant tool is constituted by the so-called "Smart Cities Marketplace", a platform that aims to bring cities, industries, SMEs, investors, banks, researchers and many other smart city actors together. The matchmaking process is organised in three phases building on each other:

- Explore (the projects and the initiatives collected in the platform);
- Shape (a dialogue between all key stakeholders involved); and
- Deal (to ultimately close deals and finance projects).

Related policies are the "Energy and smart cities", the "Sustainable transport for smart cities", and the "Digital single market and smart cities".

Despite the proliferation of policy instruments, the legislation on smart cities is still highly fragmented, as it touches not only European exclusive competences (Article 3 TFUE: customs union; the establishing of competition rules necessary for the functioning of the internal market; monetary policy for euro-area countries; conservation of marine biological resources under the common fisheries policy common commercial policy), and Shared competence (Article 4 TFEU: internal market; social policy, only for aspects specifically defined in the treaty; economic, social and territorial cohesion - regional policy; agriculture

and fisheries except conservation of marine biological resources; environment; consumer protection; transport; trans-European networks; energy; area of freedom, security and justice; common safety concerns in public health matters limited to the aspects defined in the TFEU; research, technological development and space; development cooperation and humanitarian aid), but also numerous matters at national and local levels.

Starting from the base of a hypothetical pyramid, at the local level we can recall the relationship and interaction between local authorities and different stakeholders, including citizens.

At the state level, for instance, we can mention the organisation of the State and the organisation of the public sector, as well as the relationship between the State and the local authorities.

At the European level, many steps were taken to regulate aspects crucial also in smart cities, creating a common ground that is not limited to strategies and soft-law mechanisms.

One example is the GDPR (General Data Protection Regulation), which plays a pivotal role in the realm of smart cities, safeguarding personal data privacy amidst the web of interconnected technologies. With "consent" not the only source of legitimacy in data collecting, the focus is on the accountability of the subjects involved in the data processing for the smart city. This legislation aims to help cultivate confidence in smart city innovations, trying to find balances between individual privacy and public interest.

All the other relevant legislation on the Digital sector potentially intersects the area of Smart cities, like the Digital Markets Act, the Digital Services Act and the Data Act.

Peculiar interest could constitute the Data Governance Act, another cross-sectoral instrument that aims to regulate the re-use of publicly/held, protected data, by boosting data sharing through the regulation of novel data intermediaries and by encouraging the sharing of data for altruistic purposes" (website European Commission, "Data Governance Act Explained). It applies to both personal and non-personal data. In case of the former, it offers safeguards supplementing the provisions of the GDPR to build a trustworthy data reuse and sharing infrastructure, a prerequisite to making more data available on the market.

Another relevant legislation is the AI Act, particularly with its provisions prohibiting AI systems from providing social scoring of people by public authorities or by private actors.

The social scoring could be defined (Recitals 17, not binding, and Article 5, binding) as the evaluation or classification of individuals or groups by public or private actors using AI systems, based on an aggregation of data points related to their social behaviour across various contexts or on known, inferred, or predicted personal or personality traits over time.

The "social score" (similar to what is experienced in China) creates concerns because it may result in discriminatory, detrimental, or unfavourable treatment of individuals or groups, in contexts unrelated to where the data was initially gathered, or in a manner that is disproportionate or unjustified relative to the gravity of their social behaviour.

The relationship between the topic of smart cities and social scoring is pivotal in shaping how innovations can be done in the European urban context. Practises which can lead to outcomes that violate dignity, non-discrimination, equality, and justice, must be considered unacceptable by the European legal standards. Space for "lawful evaluations done for specific purposes in compliance with relevant laws" is, however, open and subject to further developments.

5. Future research

Future developments in the smart city paradigm could be represented by any of the developments in the Digital sector involving or intersecting the use of data.

For the purposes of this Chapter, we limit the discussion to the regulatory sandboxes in the context of AI, comprehensive forms of city management and the metaverse.

The regulatory sandboxes, mentioned in the AI Act as an encouraged way to facilitate the development and testing of innovative AI systems under strict regulatory oversight, are controlled environments in physical, digital or hybrid form, within which businesses and startups can test new technologies, products, and services under regulatory supervision but with certain relaxations or exemptions from standard regulatory requirements. These sandboxes intend to provide a safe space for innovation, allowing companies to experiment and iterate on their solutions while ensuring consumer protection and compliance with regulatory standards. Member States are encouraged to establish at least one AI regulatory sandbox at the national level, but they can also join other Member States in their efforts.

Comprehensive forms of city management are, for example, the so-called Digital Twin Cities. The idea, and its definition, comes from product lifecycle management. A digital twin "consists of three crucial parts: physical products, virtual products, and the connections tying them" (definition of Prof. Grieves (2003), in Deng *et al.*, 2021). It applies potentially to every physical product, but in the case of a city it's composed by the city itself and a sort of "virtual replica".

Digital twin technology is described as having a huge impact on cities, even if intended as the digitisation of systems simpler than a city (for example, a building or a train station). These digital twins leverage AI, data analytics, and machine learning to create dynamic digital simulation models. They could

enable real-time city management and informed policymaking through models, visualization, and scenario building.

A further dimension for smart cities is the metaverse and its implementation. Virtual reality, augmented reality, or the metaverse, could, in fact, profoundly shape the city reframing spaces and the physical relationships, between themselves and between citizens and spaces.

6. Conclusions

The impact of technology on the city is a topic that is less prone to determinism than it seems. The results can be unpredictable and promised revolutions didn't obtain what they were supposed to obtain (for example, the advent of digital didn't result in the disappearance of papers). Contrarily, the actual usage of technologies has shaped the world in new ways.

The expansion of AI (or its contraction compared to expectations) can either change fundamental components of the city or remain limited to superficial aspects. It will depend on the usage of technology (and not on the technology in itself), as no usage of technology is ideologically neutral.

The role of practitioners, scholars and public servants is to keep in mind that "along with spaces themselves, "smart city" terminology and values are under construction" (Halegoua, 2020). Citizens, and citizens' needs, must be central more than the cities in themselves.

Bibliography

Breuer, J. and Pierson, J. (2021) 'The Right to the City and Data Protection for Developing Citizen-Centric Digital Cities'. *Information, Communication & Society,* 24(6), pp. 797–812. Available at: https://doi.org/10.1080/1369118X.2021.1909095.

Courmont, A. and Le Galès, P. (2019) *Gouverner la ville numérique*. Paris: La vie des idées.

Deng, T., Zhang, K. and Shen, Z-J. (2021) 'A Systematic Review of a Digital Twin City: A New Pattern of Urban Governance toward Smart Cities', *Journal of Management Science and Engineering,* 6(2), pp. 125–34. Available at: https://doi.org/10.1016/j.jmse.2021.03.003.

European Commission, https://commission.europa.eu/eu-regional-and-urban-development/topics/cities-and-urban-development/city-initiatives/smart-cities_en.

European Commission, https://digital-strategy.ec.europa.eu/en/policies/smart-cities-and-communities (link consulted on 31.01.2024).

European Commission, https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained (link consulted on 31.01.2024).

Ferraris, M. (2014) *Documentalità. Perché è necessario lasciar trace*. Roma-Bari: Laterza.

Ferraris, M. (2021) *Documanità: filosofia del mondo nuovo*. I Robinson. Bari: GLF editori Laterza.

Floridi, L. (2017) *La quarta rivoluzione*. Milano: Raffaello Cortina Editore.

Floridi, L. (2020) *Pensare l'infosfera*. Milano: Raffaello Cortina Editore.

Forrester, J. W. (1969) *Urban Dynamics*. Cambridge, MA: MIT Press.

Giffinger, R. *et al.* (2007) *Ranking of European medium-sized cities, Final Report*, Vienna: Centre of Regional Science (SRF), available at: https://www.smart-cities.eu/.

Halegoua, G. (2020) *Smart Cities*. Cambridge, MA: The MIT Press.

O'Brien, P., Pike, A. and Tomaney, J. (2019) 'Governing the "ungovernable"? Financialisation and the governance of transport infrastructure in the London 'global city-region", *Progress in Planning*, 132, available at: https://doi.org/10.1016/j.progress.2018.02.001

Russo, F., Rindone, C. and Panuccio, P. (2014) 'The process of smart city definition at an EU level', in N. Marchettini *et al.* (eds). *Sustainable city IX: Urban Regeneration and Sustainable*. Southhampton: Wit Press, pp. 191. 979-989. Available at: 10.2495/SC140832..

Söderström, O., Paasche, T. and Klauser, F. (2014) 'Smart cities as corporate storytelling', *City*, 18(3), pp. 307-320. Available at: 10.1080/13604813.2014.906716.

Chapter XIII
Surveillance, Security, Resilience and Protection of Critical Infrastructures

by Pierluigi Perri*

Index: 1. "S" as "Smart" but also as "Surveillance". – 2. Smart cities and protection of critical infrastructures. – 3. Security and resilience of smart cities.

1. "S" as "Smart" but also as "Surveillance"

The functioning itself of a smart city is based on the collection of data to improve services or predict possible needs or outcomes of a service (e.g. transportation, healthcare, education, etc.). Surveillance is, by design, an operating criterion of smart cities that should be taken into consideration when thinking about the governance of data (both personal and non-personal) and the cybersecurity.

This should come as no surprise when one considers that under the generic phenomenon of "surveillance," a wide range of activities can be encompassed. These activities are normally traced back to different disciplines such as media sociology, communication sciences, constitutional law, labor law, criminal law, political philosophy, and legal informatics, which provide different framings aimed at hypothesizing their possible effects, especially on people's fundamental rights and freedoms.

Indeed, in this general topic, it should be immediately noted that surveillance is not necessarily a negative concept: in fact, issues related to urban security, protection of heritage, safety of the person or of the community, and prevention of criminal or terrorist phenomena may come into play.

But surveillance can also be used to invade privacy, profile users of a particular service or consumers, unfairly thwart political opponents, restrict freedom of expression or movement, and facilitate employers' abusive exercise of their powers.

However, despite the wide range of possible applications of surveillance and, consequently, the broad catalog of fields of law that can be affected by this issue (Rodotà, 1973), all scholars who have addressed this topic have

* Lawyer. Associate Professor of Information Security, Privacy and Protection of Personal Data and Data Protection, Law and Artificial Intelligence at the University of Milan and Vice Coordinator of the Information Society Law Center (ISLC).

emphasized the close link that has always existed between technology and the power derived from surveillance (Ellul, 2012).

Therefore, it becomes extraordinary and linked to a principle of necessity, at least in modern constitutional texts, to regulate this activity so that it can conform to the principles contained in the charters of fundamental rights. It therefore turns out to be of extreme interest, especially for a correct theoretical and Computer Law reconstruction, to analyze the close links between technique and theories of surveillance, which have progressively developed thanks to the expansion of the potential and use of technological tools, up to the theorization of the "surveillance society," according to the definition attributed to Gary T. Marx (Marx, 1985) and David Lyon (Lyon, 2001).

Aggregating, for simplicity's sake, the main theoretical developments on the subject, some scholars (Galič, Timan and Koops, 2017) have identified three important moments in the development of modern theories of surveillance:

1. a first phase is related to Bentham's theory on the construction of the prison and Foucault's subsequent analysis of "discipline", in which the Panopticon is used as a metaphor for institutions and society. This fundamental construction of surveillance theory continues to be a key contribution for all scholars of the subject and has aspects that are still very coherent with the current situation, although neither Bentham nor Foucault could imagine the potential offered by technology today;
2. the second phase, which could be called post-panoptic, is the one that moves from Bentham's theories to develop alternative models of surveillance. Reference is made, in particular, to Deleuze and Guattari's (Deleuze and Guattari, 2017) theories of the "society of control," related to the powers of the bureaucratic state and the underlying risks of a connected and computerized society, to Haggerty and Ericson's (Haggerty and Ericson, 2000) theories of "assembled" surveillance and Zuboff's(Zuboff, 2015) "capitalist" surveillance;
3. a third phase, lastly, is the one that aims to illustrate the more modern and current perception of surveillance: it still rests on the models just mentioned above but is much more focused on the user and his or her concrete possibilities of hiding from different forms of centralized or pervasive control, in that perspective that is called digital resistance (Ziccardi, 2013), which should be considered as a possible effect of the surveillance practices necessary for the functioning of smart cities, similarly to what happens in cities with a high presence of surveillance cameras.

The relationship between surveillance and smart cities was well analyzed by two scholars (Sadowski and Pasquale, 2015), who described the social theory that refers to Deleuze's "spectrum of control". According to this theory, in fact, Deleuze noted that "We're moving towards control societies that no longer operate by confining people but through continuous control and instant

communication" (Deleuze, 1995). The "cyborgification" of city life poses a series of questions on how we can benefit from the improvements offered by smart services and avoid that the human beings are locked into so-called "coded spaces", meaning that the physical space is infused with information that makes it impossible to exit from the pervasive presence of sensors and take back control (Howard, 2015).

It was the case recently addressed by the Italian Supervisory Authority for the Protection of Personal Data, that sanctioned the municipality of Trento for the illegal use of sensors to collect data about citizens (the Italian version of the sanction is available at http://bit.ly/3WT8Svz).

In the city of Trento the research projects named *Marvel* (Multimodal Extreme Scale Data Analytics for Smart Cities Enviroments) and *Protector* (PROTECTing places of wORship) were proposed for scientific research.

These projects, implemented thanks to European funds, aim to create technological innovations to improve urban security, adopting a model that can be identified as a "smart city" model.

On one hand, the *Marvel* program was focused on collecting footage from surveillance cameras already installed in the municipal area for urban security purposes and audio captured by microphones installed on public roads.

According to the municipality, the collected data would have been immediately anonymized. Subsequently, the data should be analyzed, using artificial intelligence technologies, to identify situations that may be classified as a risk to urban security.

The *Protector* project, on the other hand, was developed to collect not only the footage of surveillance cameras, but also to collect and analyze hate speech messages posted on social networks – in particular Twitter and YouTube – to assess if negative emotions were expressed on those comments and to identify possible information that may be of interest for the Police, especially to identify possible threats to security in places of worship.

Following an elaborate investigation, the Personal Data Protection Authority found multiple violations of the EU Regulation 679/2016 ("General Data Protection Regulation" or "GDPR") and of the national data protection regulation Legislative Decree 196/03.

The Municipality of Trento has been found to lack a proper legal basis for its research projects, particularly concerning the processing of personal data which include sensitive information, because they can harm individuals' fundamental rights and freedoms. Additionally, the sharing of data with third parties, without adequate safeguards, made the processing unlawful.

Furthermore, the anonymization techniques used were evaluated by the Authority as insufficient to protect data subjects from re-identification risks and it was noticed also a lack of transparency, because the Municipality failed to

fully disclose the processing details, such as the potential recording of conversations via the microphones placed along the streets.

Finally, the Authority emphasized that the use of these new technologies, even if intended for public surveillance, requires a prior data protection impact assessment to identify and address potential risks to data subjects' rights, including constitutional freedoms.

Therefore, due to all these non-conformities, the municipality of Trento was fined 50,000 euros and was obliged to delete unlawfully processed data.

This case is a clear example of how the topics of surveillance and smart cities are strictly interconnected due to the digital technologies that are necessary to collect and analyze the modern urban challenges, such as mobility management, public security, urban planning, waste management and optimization of essential resources (water, electricity, gas, etc.).

In fact, all the actions related to the topics enlisted above rely on a collection of large amounts of real time data captured by sensors (Internet of Things, video cameras, biometric readers) that requires some important steps, as we learnt from the case described above:

- **prior assessment:** to identify the applicable legal bases that allow a specific processing of personal data according to Art. 6 of the GDPR and manage the high risks related to the processing of personal data that can harm the fundamental rights and freedom of the natural person, according to Art. 35 of the GDPR;
- **transparency:** which should be intended as the duty to inform all the natural persons of the collection and processing of data (especially personal ones) that are necessary to achieve the intended goal of the smart city;
- **security:** which should be a fundamental trait of smart cities, to be able to deliver their services in a reliable way and to protect the data collected from inappropriate or unlawful use.

2. Smart cities and protection of critical infrastructures

Like every other digital service, the smart cities are not excluded from cyber threats, with the aggravating circumstance that the impact of cyber incidents could be far more severe, considering that different structures that are interconnected and exposed in a typical smart city configuration can be classified as "critical infrastructure".

In fact, this term is to identify the infrastructure that provides essential services for the population and where an incident would have a significant disruptive effect.

There are several definitions of "critical infrastructure", "critical entities" and "essential services" elaborated in different states, but in the EU one of the

most recent regulations can be found in the Directive (EU) 2022/2557 on the resilience of critical entities.

Looking at Article 2 of the definition, we find that:

> **critical infrastructure:** means an asset, a facility, equipment, a network or a system, or a part of an asset, a facility, equipment, a network or a system, which is necessary for the provision of an essential service;
> **essential service:** means a service which is crucial for the maintenance of vital societal functions, economic activities, public health and safety, or the environment;
> **incident:** means an event which has the potential to significantly disrupt, or that disrupts, the provision of an essential service, including when it affects the national systems that safeguard the rule of law;
> **resilience:** means a critical entity's ability to prevent, protect against, respond to, resist, mitigate, absorb, accommodate and recover from an incident.

Speaking about critical entities is fundamental, however, to focus on the effects of an incident. The criteria to determine the significance of a disruptive effect are identified by Art. 7 of the Directive 2022/2557 and are the following:

- the number of users relying on the essential service provided by the entity concerned;
- the extent to which other sectors and subsectors as set out in the Annex to the directive depend on the essential service in question;
- the impact that incidents could have, in terms of degree and duration, on economic and societal activities, the environment, public safety and security, or the health of the population;
- the entity's market share in the market for the essential service or essential services concerned;
- the geographic area that could be affected by an incident, including any cross-border impact, taking into account the vulnerability associated with the degree of isolation of certain types of geographic areas, such as insular regions, remote regions or mountainous areas;
- the importance of the entity in maintaining a sufficient level of the essential service, considering the availability of alternative means for the provision of that essential service.

The Annex to the Directive 2022/2557 identifies the sectors, subsectors and categories of entities and includes several areas that are typically interested by the action of smart cities, namely: energy, transport, banking, financial market infrastructure, health, drinking water, waste water, digital infrastructure, public administration and production, processing and distribution of food.

To foster protection of the infrastructure of a smart city, both critical and non-critical, the keyword that is adopted worldwide is "resilience". In fact, considering the significance of the services delivered via those infrastructures,

the possibility of absorbing an impact and continuing to deliver their services while protecting the people is of fundamental importance. For this reason, the EU has developed specific projects to manage the cybersecurity of smart cities in terms of resilience. One of the most interesting projects is S4AllCities, which aims to make cities' infrastructures, services, ICT systems and Internet of Things more resilient while promoting intelligence and information sharing amongst security stakeholders. The system will focus on risk-based open smart spaces security management, cybersecurity shielding, suspicious activity, behaviour tracking, the identification of unattended objects, the real-time estimation of cyber–physical risks in multiple locations and measures activation for effective crisis management.

The S4AllCities approach centers around the development of three modular digital twin sub-systems, each bringing a complementary level of intelligence to the smart city, namely:

1. Distributed Edge Computing IoT Platform: DECIoT, responsible for the intelligent edge processing of sensor observations & measurements
2. Malicious Actions Information Detection System: MAIDS, responsible for machine detection and intelligent understanding of behavior
3. Augmented Context Management System: ACMS, responsible for the augmented reality and intelligence under a Common Operational Picture.

This project was also developed with pilot scenarios in three cities:

1. the city of Trikala in Greece, where the pilot scenario was built around two possible targets: the park that hosts the Christmas Festival and the data center of the city of Trikala;
2. the city of Bilbao in Spain, where the pilot scenario was developed on the protection of crowds during massive gathering for festivities in the city center and metro stations by detection of anomaly and illicit behaviour of individuals or groups in crowded areas, of chemical precursor to explosive and suspected armed attackers and of cyber attacks to the digital infrastructure of the city;
3. the city of Pilsen in Czech Republic, where the pilot scenario was focused on a crisis situation during a mass gathering of people in a single space, including the evacuation of the football stadium of the city which accommodates 11.700 spectators.

3. Security and resilience of smart cities

Achieving resilience needs a strong effort in terms of the design, implementation and maintenance of the infrastructure, which requires specific cybersecurity practices to prevent, manage and recover from a cyber incident.

An interesting joint publication of 19th of April 2023 drafted thanks to a collaborative effort from the United States Cybersecurity and Infrastructure Security Agency (CISA), the United States National Security Agency (NSA), the United States Federal Bureau of Investigation (FBI), the United Kingdom National Cyber Security Centre (NCSC-UK), the Australian Cyber Security Centre (ACSC), the Canadian Centre for Cyber Security (CCCS), and the New Zealand National Cyber Security Centre (NCSC-NZ), aims to identify the best practices of cybersecurity for smart cities.

As with every cybersecurity process, it is important to first identify the threat model and the attack surface as a basis for the identification of the risks that should be managed. In this sense, the authorities point out that the characteristic of a smart city is the expansion of the attack surface due to the integration of a great number of previously separate infrastructure systems into one network. The interconnection of different networks also has the effect that the network administrator of a specific system or segment loses visibility of the collective system risks due to the lack of detection systems that can scan the entire network.

Another emergent threat of cybersecurity in general and in a smart city environment is connected with the security of the supply chain, because a violation by a member of the group of suppliers can facilitate access and escalation into the system of a smart city. For this reason, vulnerabilities in ICT supply chains that are the result of intentional action, or unintentional outcomes related to poor security practices can impact the operation of a smart city.

Taking into consideration the constantly increasing number of cyberattacks and the impact that they may have on many individuals, the authorities developed some recommendations to develop a safe and secure smart city:

- **Secure Planning and Design.** Communities should integrate smart city technologies with strategic foresight, ensuring cybersecurity risk management. New technologies should be securely incorporated into legacy infrastructure, necessitating potential redesigns for secure deployment. Emphasis should be on resilience, considering both physical and cyber risks, especially with IoT and IIoT systems.
  - **Apply the Principle of "Least Privilege".** Organizations should follow the principle of "least privilege", that is a security principle aimed at granting minimal necessary system resources and authorizations strictly necessary to perform the necessary tasks assigned to a natural person or a system. Administrators must review configurations, update privileges upon role changes, and employ a tiered access model, using dedicated workstations for full privileges and limiting high-risk access.
  - **Enforce Multifactor Authentication.** Multifactor authentication (MFA), that is an authentication system that requires more than one distinct authentication factor for successful authentication (e.g. a password

and a code delivered by email or SMS), should be enforced for remote access and critical actions. Configuration policies must be reviewed to protect against MFA exploitation and ensure robust authentication processes.

  - **Implement Zero Trust Architecture.** Zero trust principles (i.e the principle of i) verify explicitly every user, ii) use least privilege access and iii) assume that, because a breach may always occur, the defense against cyber-attacks must be improved routinely) should be implemented to enhance security, requiring authentication for each connection and allowing better network visibility and governance. This approach supports a layered defense and efficient security management.
  - **Manage Changes to Internal Architecture Risks.** Network administrators should manage communications between subnetworks, identifying and isolating critical systems to reduce compromise impact. Maintaining awareness of network architecture and security responsibilities is crucial.
  - **Securely Manage Smart City Assets.** Physical and logical security controls should protect smart city assets against theft and unauthorized changes of the assets used to deliver the services, including sensors and monitors.
  - **Improve Security of Vulnerable Devices.** Automatic patching processes, threat intelligence, and asset management programs should be employed to secure devices and software. This includes planning for component replacements nearing end-of-life.
  - **Review Legal, Security, and Privacy Risks.** Ongoing evaluation and management of legal and privacy risks associated with smart city deployments are essential.
- **Proactive Supply Chain Risk Management** Organizations should manage ICT supply chain risks, using trusted vendors and incorporating in the contracts minimum security requirements, together with the provisions of audit and continuous assessment. Transparency in technology supply chains is necessary for public trust.
  - **Software Supply Chain.** Security requirements for software suppliers should include secure development practices, vulnerability management, and patching. Vendors should follow security-by-design principles and maintain active product support.
  - **Hardware and IoT Device Supply Chain.** Due diligence on IoT de.vices and hardware is essential, including understanding data handling and security measures. A risk register should identify dependencies on cloud computing and external components.

  - **Managed Service Providers and Cloud Service Providers.** Security requirements for service providers should be clear, including contractual agreements addressing data security and risk management.
  - **Incident Notification and Sharing of Information.** Different laws prescribe the sharing of information about incidents as a best practice to combat cyber threats. Moreover, Article 28 of the GDPR explicitly requires the duty to cooperate between the Data Controller and the Data Processor in order to exchange the necessary information for the notification of a data breach to the authorities. These two best practices should be always implemented by specific obligations in the contract with the suppliers.
- **Operational Resilience** Contingency plans for manual operation of critical infrastructure are necessary, along with staff training. Organizations should prepare to isolate and operate autonomously in case of compromises.
  - **Backup Systems and Data.** Organizations should maintain and test backups, isolating them to prevent ransomware impact. Plans should ensure data lifecycle security and enable system restoration after attacks.
  - **Conduct Workforce Training.** Training should cover both automated and manual operations, regularly updated to include new technologies. Employees must be prepared to isolate compromised systems and restore services.
  - **Develop and Exercise Incident Response and Recovery Plans.** Incident response plans should involve all stakeholders and be regularly updated and exercised. Hard copies should be accessible in case of network inaccessibility.

By implementing at least these cybersecurity measures, strategic planning, and comprehensive training, communities can promote the secure integration of smart city technologies, fostering resilience and operational continuity in the face of evolving digital threats.

Bibliography

Cybersecurity Best Practices for Smart Cities (2023) available at: https://www.cisa.gov/resources-tools/resources/cybersecurity-best-practices-smart-cities

Deleuze, G., Guattari, F. (1980) *Mille plateaux*. Paris: Les Editions de Minuit.

Deleuze, G. (1995) *Negotiations, 1972–1990*. Translated by M. Joughin. New York: Columbia University Press.

Ellul, V.J. (2012) *Le Système technician*. Paris: Cherche Midi.

Galič, M., Timan, T., Koops, B.-J. (2017). 'Bentham, Deleuze and Beyond: An Overview of Surveillance Theories from the Panopticon to Participation', *Philosophy & Technology*, (30) 1, pp. 10-11.

Haggerty, K.D., Ericson, R.V. (2000) 'The surveillant assemblage', *British Journal of Sociology*, (51)4, pp. 605-622.

Howard, P.N. (2015) *Pax Technica: How the Internet of Things May Set Us Free or Lock Us Up*. New Haven: Yale University Press.

Lyon, D. (2001) *Surveillance society: Monitoring Everyday Life*. Maidenhead: Open University Press.

Marx, G.T. (1985) 'The Surveillance Society: The Threat of the 1984-Style Techniques'*, The Futurist*, June, pp. 21-26.

Rodotà, S. (1973) *Elaboratori elettronici e controllo sociale*, Bologna: Il Mulino.

Sadowski, J., Pasquale, F. (2015), 'The Spectrum of Control: A Social Theory of the Smart City', *University of Maryland Francis King Carey School of Law Legal Studies Research Paper*, 2015(26).

Ziccardi, G. (2013) *Reistance, Liberation Technology and Human Rights in the Digital Age.* Berlin: Springer Verlag, pp. 161-304.

Zuboff, S. (2015). 'Big Other: Surveillance Capitalism and the Prospects of an Information Civilization', *Journal of Information Technologyz*,, 30, pp. 75-89.

Chapter XIV
Computer Crime on the Dark Web

by Aleksandra Klimek-Lakomy*

Index: 1. Introduction. – 2. The structure of the dark web. – 3. An overview of cybercrime activities on The Onion Router. – 4. A comparison of the American and European legal approaches to dark web crimes. – 5. Selected types of dark web crimes: a few examples.

1. Introduction

The dark web has been considered one of the most (if not the most) controversial parts of the Internet. Developed since the beginning of the 21st century, it was initially considered a space ensuring the enduring anonymity of all Internet users. In time, its anonymity-oriented features have unfortunately attracted many cybercriminals who use the dark web for their illegal activities. In effect, the dark web has been perceived as a cradle of computer crime, which is unsafe to browse by ordinary individuals. It is also one of the least explored parts of the Internet due to its decentralized structure and difficult navigation (Kaczmarek, 2020).

The dark web hosts a variety of cybercrime activities that cannot be realized elsewhere due to the surveillance of authorities. Existing platforms enable illegal actors to interact with each other in various ways. Criminals use the dark web to develop, share, exchange or sell malicious software. Dedicated message boards are frequently used to distribute leaked data due to cyber-attacks, such as personal data, unreleased codes from mainstream software, and login credentials. The dark web is also used by criminals who focus on the offline world. So called "darkmarkets" are hubs for the drug trade, acquiring forged or stolen documents, credit cards or buying arms. It is a place that facilitates access to illegal services, including targeted cyber-attacks and assassinations.

State authorities struggle to fight against dark web crime effectively, as the overall scale of illicit activity in this area has not changed much in recent years. This means that the dark web continues to be a place of innovation for computer crime, in which new ways of avoiding or violating existing criminal laws are developed. This process is not always accurately realized by state authorities, which have problems

* Attorney-at-law, Ph.D. in security studies and Vice-Dean of the WSB Academy in Dąbrowa Górnicza in Poland.

keeping up with developing legal (and practical) solutions to address these changes. Difficulties in this dimension remind of the challenges in developing European laws on artificial intelligence, which do not always match the pace of the changes in this sector of technology (Riccio, Scorza, Ziccardi, 2022).

This chapter presents selected forms of computer-related crime present in the most popular part of the dark web – The Onion Router (TOR) – from a legal perspective. It shows what kinds of illicit activities can be encountered on TOR and how they can be addressed based on the criminal code. This chapter introduces a comparative perspective and investigates selected national criminal legislation applicable to the dark web. It considers attitudes adopted by legislators in the United States and selected European countries. However, it places special emphasis on comparing the criminal laws of Italy and Poland in this area.

2. The structure of the dark web

The dark web is viewed as this part of the world wide web (WWW) that is not indexed by search engines and cannot be accessed with mainstream web browsers, such as Firefox, Edge or Chrome. The dark web is also perceived as a place that ensures a certain level of anonymity for its users. The dark web cannot, however, be considered a homogenous environment, as it comprises many separate networks based on distant technical solutions. There are multiple available networks. The most prominent are: The Onion Router, I2P, Freenet, ZeroNet, GNUNet, Retroshare.

The Onion Router (TOR) is one of the most popular and the biggest part of the dark web. It can be accessed only through a dedicated but free software – the TOR Browser. The history of TOR began at the end of the 20th century, because of the research project carried out by the scientific lab of the US Navy and later by the Defense Advanced Research Project Agency (DARPA). The goal of this project was to find out ways of secret, anonymized communication. In 2003, it resulted in the launch of TOR, when its source code was published on the internet. In 2004, it was taken over by the Electronic Frontier Foundation. Currently, it is being developed by the users. The Onion Router allows hidden network activity, including website surfing, chat and sending instant messages. It hides the user's IP address based on so-called "onion routing," which is a way of encrypting and allowing communication through a network of relays. From a practical perspective, TOR Browser allows to use Clearnet content anonymously. At the same time, it also allows access to another, secret layer of Internet communication, built upon .onion domains.

Other dark web networks adopt diverse technical solutions that contrast with the approach of TOR. Freenet, one of the oldest dark web networks, is a peer-to-peer adaptive network that allows users to share and download data while maintaining anonymity. Freenet exploits peer-to-peer (P2P) Torrent-like

solutions. This means that accessing Freenet content also means sharing it by user's computer. Zeronet, one of the newest dark webs, constitutes a decentralized web platform using Bitcoin cryptography and peer-to-peer technology. Effectively, this also means that accessing dark web content forces users to share it. Both dark web networks are P2P-based, but their approaches to the anonymity of users are different. Freenet ensures high level of anonymity, while it is difficult to reach on ZeroNet, without additional measures. Other networks usually combine some of these solutions in new forms. For instance, GNUNet, first released in 2001, combines P2P with so-called F2F topology, which is an option for restricting connections only to trusted users.

In other words, the dark web does not constitute a technically unified environment, which is similar to Clearnet – this part of the internet that is accessible with normal browsers. It is composed of many different environments, adopting contrasting technical solutions to ensure the privacy of users and the resilience of posted content. It is perceived as a reaction to the pro-censorship attitudes of many governments. In each case, access to the dark web requires utilizing specialized software. It is an environment potentially beneficial for free speech and human rights, but its features have attracted masses of computer criminals.

3. An overview of cybercrime activities on The Onion Router

Most sources indicate that The Onion Router is the biggest and most popular network in this environment. In 2023, it had over two million users daily, browsing content located at over 65,000 URLs. The origin of TOR users is differentiated. The Onion Router daily traffic in 2024 was dominated by Russians (37.05% of users), Iranians (19.53%), and Americans (11.87%). In Europe, this part of the dark web is most popular among Germans (3.18%), France (2.39%) and the UK (1.96%).

As mentioned above, because of its features and growing popularity, The Onion Router has become a safe place for computer crime. According to the data gathered in 2020, around 6.7% of TOR users visited illegal .onion domains. This means that approximately 150,000 people use this network for crime-related purposes. The scale of cybercriminal activity in this environment is also demonstrated by other publicly available data. According to the study of Thomas Rid and Daniel Moore (2016), among the 2723 active websites that they discovered and classified, 1547 contained illegal content, which is around 57 percent. Of these, 423 had a connection to drugs, 327 to finances, 198 to other illicit acts, 140 to extremism, 122 to illegitimate pornography, 96 to hacking and 42 to violence (Moore and Rid, 2016). The scale of illegal activity may also be seen in actions carried out by police. In 2023, in a joint operation on three continents, they targeted illegal dark web vendors. As a result, 288

criminals were arrested. Police seized 50.8 million EUR, 117 firearms and 850 kg of drugs. Most of the arrests took place in the United States (153), United Kingdom (55), Germany (52) and Netherlands (10). Despite these efforts, cybercrime in the dark web is not getting smaller, as criminals continuously create new methods of ensuring their anonymity. As a result, the dark web remains a place for a variety of computer-related crimes.

These illicit activities on The Onion Router are concentrated around several types of domains. The most common ones are the so-called darkmarkets, offering numerous illegal goods and services that can be anonymously ordered and paid with cryptocurrencies – either BTC or Monero. The most important darkmarkets taken down in recent years were Hydra Market, UniCC and ToRReZ. First of them was a Russian market, which was unique because of the revenue it generated (around $1 billion). UniCC specialized in fewer types of illicit goods, while the third one – ToRReZ – had more than 160,000 registered users. All three trading websites were taken down in 2021 and 2022. Darkmarkets belong to the most popular types of cybercrime-related locations on TOR. They sell a broad range of merchandise in electronic or real-life form that originates from such crimes as fraud, unauthorized access to information (hacking, phishing) or the dissemination of malware. Users on darkmarkets sell, for example, leaked databases, containing stolen personal information (personal names, home addresses, phone numbers), login credentials, or even stolen documents, such as passports, IDs, and credit cards. It should be noted that while the dark web is usually not the place where these acts occur, TOR serves as an intermediary, allowing perpetrators to reach potential customers. Similar functions are realized by smaller vendor shops. In contrast to darkmarkets, they are usually specialized in only one type of illicit goods, such as drugs, malware or firearms. TOR also contains multiple message boards and "chans," which gather a variety of different communities. Some of them are directly related to computer crimes, as they are the places where individuals look for malware, exploits, stolen personal data and credentials. Hacking-related message boards frequently also contain stolen software or their source codes. One of the most famous examples in recent years was related to the leak of source codes of the most popular video games, including Grand Theft Auto 5 and Witcher 3. There are different areas on The Onion Router that act as facilitators or intermediaries in criminal actions, having effects both online and offline.

4. A comparison of the American and European legal approaches to dark web crimes

Accessing the dark web is generally considered legal in the United States and most European Union member states. This is not the case in many nations – including Russia, China and South Africa – which have blocked or banned the

dark web. While European states do not follow this approach, there are some EU member states that have blocked some of the TOR exit nodes. A complete ban on the ability to host dark web services was also debated in Europe. A brief comparison of American and selected European legislation shows different approaches to fighting against computer crimes on TOR.

In the United States, there is a complex network of regulations that address different aspects of cybersecurity and privacy online. One of the reasons for this situation is caused by its political system, as American laws are made both on the state and federal levels. This also applies to acts on cybercrime, as simultaneously with federal acts, different states adopt their own regulations. For example, in 2023, Montana, Tennessee, and Texas passed bills on personal or consumer data protection. On a federal level, there are number of relevant laws. One of the most important is Computer Fraud and Abuse Act (CFAA), which provides criminal and civil penalties related to computer crime. Its regulations cover such acts as unauthorized access to a computer, damaging a computer (up to five years in prison), trafficking in passwords (up to 12 months of imprisonment), or cyber-extortion (up to five years of imprisonment). Despite the existence of these acts, there were attempts made in the United States to introduce laws dedicated to the dark web only. This was the case with the Dark Web Interdiction Act, which was processed in Congress in 2022. The bill introduced the definition of dark web, recognized its role in facilitating cybercrimes, and provided an enhanced penalties for drug-related crimes in this environment. The bill was eventually not passed, but it shows that The Onion Router and similar networks have been prioritized by American legislators.

The national legislation of the EU member-states applicable to dark web crime is impacted by the laws adopted at the European Union level. For example, the Digital Services Act, which came into force in 2024, introduced new solutions regarding countering illegal content and protecting minors, which are also applicable to these .onion domains which are active on the territory of the EU. The EU is also a platform, which helps law enforcement coordination in fighting cybercrime in this domain, through EUROPOL.

On the state level, European countries have adopted different legislation regarding this problem. One of the most interesting cases is Germany. In 2019, bill was proposed in this country, which criminalized operating dark web platforms intended to prevent the hosting of darkmarkets. In general, the use of the dark web is legal in Germany, but this debate led to the introduction of changes in the criminal law in 2021, which are specific to dark web crimes. In 2021, Germany added a new par. 127 to the criminal code (das Strafgesetzbuch), which addresses illegal dark web exchange, including drug and firearms trade, as well as child grooming. It penalizes individuals hosting transactional platforms on the internet that facilitate illegal actions or promote them. Such acts are subject to imprisonment to up to five years. However, it should be noted that

this new dark web-oriented regulation in criminal law has been criticized by German legal experts (Zoller, 2021).

In France, the use of the dark web is legal and protected under various human rights-related acts. Some experts indicate that the French legal system includes regulations, which should directly apply to the dark web but are inapplicable due to technical reasons. For example, articles 29 to 31 of the Loi pour la confiance indicate that the use of cryptographic means is free, but requires submitting an official statement, which is difficult on the dark web. It means that most cybercrimes taking place on TOR are addressed by ordinary laws covering cybercrime. In the French criminal code (Code pénale), there are multiple relevant regulations, such as art. 323-1, which penalizes acquiring unauthorized computer access with a maximum fine of 45000 EUR and three years in prison. However, in opposition to other European states, French customs service is authorized to buy illicit goods on the dark web to uncover identities of computer criminals involved in their exchange. Such privilege is ensured under art. 67 bis of the customs law (Loi des douanes).

United Kingdom, despite not being part of the European Union, can also be perceived as a relevant example of European legislation in this area, as it has a number of specialized legislations on cybercrime. The most important one is Computer Misuse Act of 1990, which deals with several computer-related crimes. The use of the dark web in the UK is considered legal, although accessing illegal websites on TOR may be considered an offence, depending on the general context. This legal liability may refer to such cases, as accessing .onion domains with "indecent images," which is punished under Section 1 of the Protection of Children Act from 1978. Dark web activities may be also subject to the controversial Online Safety Act adopted in 2023, which aims to protect children in online spaces.

All these approaches show that in most considered states dark web-related crimes are addressed by ordinary cybercriminal and related laws. Some states attempt to introduce "special" legislation, however, their outcomes may be different. In Germany, it took years to adopt changes in the criminal code and their outcomes were criticized. In the United States, proposed bill was not even adopted. Therefore, it is important to show how exactly different types of crimes on TOR can be handled on the level of the criminal law.

5. Selected types of dark web crimes: a few examples

The definition of cybercrime has not been included in the criminal codes of Poland and Italy expressis verbis, but such a concept has been adopted in the legal regulations of these states through the ratification of the Budapest Convention on Cybercrime. Aside from the Convention and changes introduced in their criminal codes since the end of the 20thcentury, both states adopted

several special laws, dealing with the most specific types of cybercrimes, including copyright infringement or cyberstalking (Ziccardi, 2012; Gozdyra, 2017). This means that in general terms both states have a common ground regarding the perception of computer crime.

In the Polish legal system, cybercrime related laws have been developed since the 1990s, which were forced by the popularization of the internet and the rapidly growing problem of software piracy. There is no unified chapter of the Polish Criminal Code (PCC), which addresses computer-related crime. However, the Code addresses specific activities, which fall under the broad definition of cybercrime. They are mostly framed either as "computer" or "internet" fraud (art. 286 par. 1 of the PCC), sabotage (art. 269 par. 1 and 2 of the PCC) or other, more specific activities, including phishing (art. 287). Other specific cybercrimes have been addressed by, for example, art. 278 (stealing computer program or a debit card), art. 285 (telecommunication-related fraud), art. 202 (public display of pornography to those who do not wish to see it), or art. 190 (stealing and using someone's identity online) (Wróbel, 2014). Software piracy and similar copyright infringement activities are covered separately, outside of the criminal code (Ustawa o prawie autorskim i prawach pokrewnych) (Radoniewicz, 2019).

A similar approach may be spotted in the Italian Criminal Code (ICC), where the concept of cybercrime (although not expressis verbis in a form of a direct definition) was introduced in 1993 by L. 547/1993 (criminalità informatica) (G. U. n. 305 del 30 dicembre 1993). It was later modified by L. 48/2008, which ratified the Budapest Convention on Cybercrime. Some of the core articles of the ICC related to cybercrime include articles 615 ter, 615 quarter, 615 quinquies, 617 quater, 635 bis, 635 ter, 635 quarter and 640 ter. For example, art. 615 ter provides punishment for unauthorized access to computer systems, also known as hacking. Art. 635 quarter is related to blocking, damaging or destroying data, information or programs. Article 640 ter of the ICC covers computer fraud.

According to existing regulations in both countries the use of The Onion Router (or the dark web, in general) is not illegal, despite its image as a criminal underground. Using TOR to simply browse the Clearnet anonymously or carry out other legal activities, including visiting legitimate .onion websites, raises no legal concerns. The illegality of using TOR is therefore dependent on what exactly an individual is doing on the dark web. However, in the case of The Onion Router, the line between forbidden and allowed is frequently blurred and easy to cross for unexperienced users. The gray zone exists in multiple areas of the dark web, where there are no clear regulations or the technical features of this environment force individuals to face random risks. Considering these issues, it is therefore important to compare Polish and Italian perceptions of the types of computer-related crimes which are most common on TOR.

5.1. Distribution of illicit materials on the dark web

Among computer-related crimes that are facilitated or caused using the The Onion Router, a large part is directly related to hosting, sharing or even viewing illicit content. Some past studies estimated that around 46% of the accessible .onion domains could host illegal content. From the viewpoint of hosting and sharing activity, the most common problem is online illicit pornography. According to Daniel Moore and Thomas Rid (2016), among the 2723 active .onion domains detected in 2016, 122 shared such materials (4.48%). These .onion domains generally share three general types of content. The first are sexually explicit images or videos of individuals recorded without their consent. This also includes content labelled in literature as "revenge porn," shared online by former partners of the victims, or recordings of rape. The second is child pornography. The dark web has been known to host such content for a long time. Some dark web researchers note that TOR is "abundant" in such data. The third includes hate speech. These crimes manifest in statements encouraging or legitimizing acts of violence against certain groups of people, violent extremist propaganda, and even instructional material on how to kill people. It is known that some groups have used the dark web since at least 2015 (Kaczmarek, 2019).

The existence of illegal content on TOR may involve a broad array of potential situations violate criminal law. This is primarily related to its distribution or sharing, either through hosting .onion websites, posting on message boards or sharing such materials elsewhere on the dark web. What should be stressed is that, from the viewpoint of the potential "consumers" of such content, legality assessment is a bit more complicated, depending on the circumstances. This is because some types of materials may be legal to watch on the territory of European Union member states, while being illegal elsewhere. This applies to, for example, browsing terrorist content.

The distribution of sexually explicit images of a naked person has been addressed in article 191a par. 1 of the Polish Criminal Code, which states that unauthorized recording of sexual acts (by using, for example, violence or deceit) or their dissemination without consent is punishable by imprisonment from three months to five years. The article, therefore, is related to images showing even partially naked individuals, as well as scenes demonstrating all types of sexual acts. In Italy, regulations dedicated to such problems were introduced in 2019. Article 612 of the ICC covers the illegal dissemination of sexually explicit images or videos in a similar manner to Polish, however, it includes greater penalty (one to six years), as well as a fine of 5000 to 15000 EUR. The law faces some interpretational challenges, related, for example, to the scale of dissemination of sexually explicit images or motives of perpetrators that would trigger this law. What is interesting is that the article perceives an increased penalty, when the offense is committed through electronic means, including,

therefore, the dark web. In both cases, such acts are prosecuted following the complaint of the offended person, which is highly problematic due to the anonymization of the dark web. TOR websites and message boards, where such content is posted, are usually difficult to find. Even the known criminal cases related to disseminating sexually explicit images on the Clearnet have faced multiple challenges related to finding the images by the victim or proving the guilt of the defendant (case DS.136/12 PR w K., see Mozgawa and Nazar-Gutowska, 2014). This solution generally creates two potential problems. First, images or videos are shared in an anonymous environment that complicates or completely blocks the identification of the perpetrator. Second, even if the national law enforcement agencies can encounter images on TOR, it does not trigger automated prosecution. They are also usually unaware of the identities of the victims.

Interesting findings can be spotted when comparing the approaches of both legal systems to child pornography on the dark web. The Polish Criminal Code's art. 202 indicates that anyone who records pornographic content of a person under the age of 15 faces imprisonment for one to ten years. Its par. 4a clarifies that anyone who downloads, possesses or accesses such content faces imprisonment between three months and five years. ICC's art. 600 ter penalizes the production, dissemination, advertisement or incitement of underage pornography. Analysis of both regulations allows to spot certain similarities and differences. They have very similar ways of interpreting the pornographic content. However, compared to the PCC, the ICC adopted a higher age limit – eighteen years. It also has higher maximum imprisonment penalty (between six to 12 years) and fines (between 24000 to 240000 EUR). Referring to the dark web content, the most interesting difference is that the ICC does not explicitly mention watching such content online, which is explicitly punished under PCC's art. 202 par. 4a. Italian law requires a direct intent to acquire and watch such materials (dolo diretto o intenzionale). A possible intention (dolo eventuale) is not sufficient under this article. This means that visiting websites with child pornography is not enough, and to violate this law, it is necessary to download such materials on the perpetrator's computer. This creates an interesting contrast between both legal systems, when it comes to perceiving dark web dimension of this crime. Since users of TOR may never know what are inside the websites they visit, they may accidentally encounter such content, which is subject to punishment in Poland. This activity, even with direct intent, may not be considered a violation of the law in Italy. It looks different, however, in case these dark web networks that are based on P2P technology. In such cases, simply visiting a website also means that the computer used for such activity is also engaged in its dissemination. Polish jurisprudence has confirmed many times that the use of P2P to download and – simultaneously – share child pornography meets the criteria of art. 202. It also follows the ignorantia iuris nocet rule in such cases,

which means that the lack of knowledge on the technical details regarding how these dark web networks work cannot be considered a mitigating circumstance.

5.2 Exchange of illicit goods

Another type of cybercrime massively present on the TOR is the exchange of illegal merchandise. This problem manifests predominantly on the independent shops and darkmarkets, specialized in selling various types of drugs, firearms, and stolen credit and debit cards. Acquired products are usually bought with cryptocurrencies and shipped internationally in anonymized packages. From a legal viewpoint, illegal exchange of goods on the dark web can be perceived from multiple angles. Depending on the specific situation, it may involve fence activities, counterfeiting, falsification, or the trafficking of arms and drugs.

The trade of drugs, being one of the most common illegal uses of the dark web, is addressed in Poland by the Law on countering drug addiction (Dz. U. z 2023 poz. 1939). Article 54 penalizes acquiring equipment necessary in drug production. Such elements may sometimes be subject to trade on the darkmarkets. Violation of article 54 is subject to fine, restriction of freedom or imprisonment for up to two years. Article 55 of the Law refers to the illegal exchange of drugs (acquisition, delivery, transfer) both in Poland and abroad. Such act is sanctioned with a fine or imprisonment up to five years. Art. 56 stresses that anyone selling drugs or participates in its delivery on the market is subject to fine or imprisonment from six months to up to eight years (Dz. U. z 2023 poz. 1939). The law also stresses that if these actions have been committed for commercial purposes, the minimum imprisonment shall be three years.

A similar regulation can be spotted in Italy, which also adopted a special law regarding the drug-related offenses (DPR 309/90). Article 73 stipulates that everyone who grows, produces, extracts, refines, sells, offers, transfers, disseminates, trades or acquires drugs is subject to imprisonment between six to 20 years. It also includes a fine between 26000 to 260000 EUR. Article 75 stresses that anyone who acquires for personal use is subject to administrative sanctions, including the loss of driving or firearms licences. A comparison of these laws shows that Italian sanctions for drug trade, both offline and online, are much higher than Polish. Although, jurisprudence in both countries considers similar factors, including the quantity of the drugs, as well as their types. In case of the dark web, drug trade raises no particular concerns in terms of interpreting such acts. Drug vendors are customarily engaged in large-scale activity on TOR, which rules out the use of any mitigating circumstances. Moreover, most acts committed on the dark web would engage three types of activities, being trade (an act of selling psychoactive substances through an .onion website), trafficking, and acquisition. In the last case, people responsible would be also subject to administrative sanctions in Italy. Usually, there should be little consideration regarding the mitigating factors, as dark web presence indicates a continuous,

professional and most probably also large-scale engagement. In this context, functions of many darkmarkets may facilitate this assessment, as it is possible to track the transactions of all vendors.

Polish Criminal Code's art. 263 par. 1 (wyrób lub posiadanie broni bez zezwolenia) sanctions offenses related to the illegal production or trade of firearms with an imprisonment of one to ten years. Its par. 2 stresses that the illegal acquisition of firearms is sanctioned to fine or imprisonment of between six months to eight years. The ICC's art. 695 (fabbricazione o commercio non autorizzati di armi) adopts largely similar regulations. It indicates that every individual that is involved in the production, smuggling or trade of firearms without the proper licence is subject to prison up to three years, as well as a fine. Also based on ICC art. 696 (vendita ambulante di armi), every person involved in arms trade may be subject to imprisonment of between three months to three years, as well as a fine. In this case, Polish regulations are much harsher, as the illegal trade is sanction of the imprisonment of up to ten years. Similarly to drug trade, these regulations raise little place for problematic interpretation regarding the dark web. There are also no legal gaps visible between the Italian and Polish system in this regard.

Selling other types of goods having illicit origin on TOR, such as stolen credit cards or identification documents, may involve a variety of regulations, including those related to fencing. In the PCC it is regulated by art. 291 (paserstwo), which stresses that anyone who buys or helps sell an object acquired through crime, as well as everyone who keeps it or helps hide it, is subject to imprisonment between three months to five years. This article is related to both movable objects and real estate. In Polish jurisdiction, it is not necessary to physically possess the object for the fencing to be the case, which allows to consider more complicated reselling activities taking place on TOR. There must be, however, a relation between the stolen goods and the fencer, as well as his or hers awareness of the origin of the goods. The ICC's art. 648 on fencing (ricettazione) stresses that anyone acquiring, receiving or hiding money or objects originating from crime for a financial benefit is subject to imprisonment between two to eight years, as well as a fine. The article also foresees situations in which the sanction is higher, including the professional activity or the origin of the goods from the robbery. There are certain differences between both regulations. Aside from the those related to imprisonment length and fines, Italian law stresses the importance of actions directed at the financial gain of the buyer. In the case of stolen debit or credit cards, this offence is committed by the buyer at the darkmarket, as the act of acquiring such goods is caused by the intent of financial gain. The Italian Court of Cassation admitted in one of its decisions that acquiring stolen credit cards meets the conditions of Article 648.

It should be noted that the types of common offenses described in this section may also be considered from the viewpoint of regulations sanctioning

aiding or abetting crime. This is related not only to all parties of the transaction but also an individual who is responsible for running the darkmarket, where it takes place. This is confirmed by regulations adopted in both countries. The Italian Criminal Code's art. 378 (favoreggiamento personale) states that anyone helping someone avoid investigation related to the crime sanction with imprisonment, may be subject to three years in prison. The PCC art. 27 adopts a simpler regulation, indicating that aiding in crime manifests when anyone actively or vocally helps violate law. Polish regulation opens a wide range of possible interpretations. Nevertheless, even the more specific ICC clearly indicates that administrators of darkmarkets may be prosecuted based on this legislation, as they usually introduce additional measures allowing customers and vendors to avoid identification by law enforcement agencies. This includes additional encryption or safe communication measures.

Exchange of illegal goods on the dark web sometimes involves another important phenomenon, which is usually called "escrow." It is defined as "the use of a neutral third party on the dark web to ensure that the transaction payment will be made to a seller on a completion of items sent to a buyer" (Dark Web Monitoring). Escrow frequently involves the use of cryptocurrencies, which also increases the anonymity of the transaction. While escrow, by definition, is not illegal, it is usually related to trading illicit good or buying illegal services. It is frequently associated with message boards engaged in dealing with illegal content, goods or services. This activity may be prosecuted based on two sets of regulations existing in Italian and Polish legal systems. The first is aiding or abetting to crime, which has been discussed above. The second is money laundering.

In the PCC it is addressed in art. 299.

According to its par. 1:

Whoever receives, possesses, uses, transfers or exports abroad, hides, transfers or converts means of payment, financial instruments, securities, foreign exchange values, property rights or other movable or immovable property derived from benefits related to committing a prohibited act, helps to transfer their property or possession or undertakes other activities that may prevent or significantly hinder the determination of their criminal origin or place of placement, their detection, seizure or forfeiture may be subject to imprisonment between six months to up to eight years.

Italy, mostly due to its history of struggle with organized crime, operates one of the most developed anti-money laundering systems in Europe, which has been in constant development since the end of the 1970s. In ICC, this crime is handled primarily by art. 648 bis. According to this article, money laundering is subject to the penalty of between four to 12 years, and a fine between 5000 to 25000 EUR. It includes all exchange activities related to goods or money originating from crime to make it difficult to trace its origin. It also includes

transferring of the goods and money. A constitutive element of this crime must be related to the factual acquisition of the goods of money of criminal origin. In both cases, escrow services taking place on the dark web meet these criteria with ease, as they participate in transferring illicit goods and laundering cryptocurrencies, to hide the identity and origin of the original owners. This interpretation is supported by the known cases, in which organizers of escrow services, moderators on trading boards or intermediaries that have been arrested on money laundering charges.

5.3 Hacking and malware distribution on TOR

The Onion Router is an environment, where hacking-related offenses take place. For example, they manifest in actions aimed at acquiring tools and knowledge allowing unauthorized access to computer systems. This happens either on hacking message boards or through dedicated websites, where individuals having sensitive know-how or unique hacking-related programs sell them or exchange them. Due to this demand, certain types of websites and markets either sell complete malware or ransomware kits. They include banking trojans, proxy malware, rootkits, keyloggers, remote access trojans (RATs). For example, the Eternity Project .onion service in 2022 offered password stealing toolkit for $260 (annual subscription), or the cryptocurrency miner for $90. Some of the cheapest ransomware at the time started from $490. While they are not produced directly on TOR, it serves as a selling or sharing environment.

The distribution of malware is addressed in article 269b of the PCC. It stresses that everyone who creates, acquires, sells or shares equipment or software dedicated to committing enumerated crimes, as well as passwords, credentials and such, is subject to imprisonment between three months to five years.

The similar set of crimes has been addressed by the ICC art. 615 quinquies, which stresses that:

> whoever, with the aim of illicitly damaging a computer or telematic system, the information, data or programs contained therein or pertinent to it or to facilitate the total or partial interruption or alteration of its functioning, unlawfully procures, holds, produces, reproduces, imports, disseminates, communicates, delivers or otherwise makes available to others or installs equipment, devices or computer programs can be imprisoned up to two years, and subject to fine up to 10329 EUR.

Italian law separately treats illegal possession and diffusion of access codes to computer or telecommunication systems, which are covered by the article 615 quater, which is sanctioned with up to two years in prison. These regulations can be used to tackle malware shops and services on TOR, as well as prosecute activities related to sharing or selling databases containing credentials on dark

web. They are broad enough to contain together professional and incidental activities in this regard. It must be noted that according to existing commentaries, to prosecute based on these articles, it is not necessary for the actual cyber-attack to take place. Only the actual fact of distribution must take place.

While TOR is not a place where actual hacking attacks take place, it hosts websites allowing to "hire" cyber-attacks. The most common forms are related to blocking certain internet services by hiring a botnet to carry out a Distributed Denial of Service (DDoS). These services are relatively cheap to buy, as in 2023 a DDoS attack cost between 5 to 30 dollars per hour. The price of more sophisticated attacks, including hacking social media accounts, e-mail addresses or databases were much more expensive and ranged from 500 to 1600 dollars. Some services also offer gathering detailed personal information about the targets. Such acts are addressed primarily by the ICC's article 615 ter, which sanctions unauthorized access into a computer or telecommunication systems.

It states that:

> Anyone who enters unauthorized into a computer or telecommunication system protected by security measures or remains in it against the expressed or implied will of the one who has the right to exclude him, shall be sentenced to imprisonment not exceeding three years.

Moreover, art. 616 of the ICC stresses that acquires unauthorized access to closed correspondence, destroys or suppresses it, is punished with imprisonment of up to one year, or with a fine up to 516 euro. It should be stressed that based on law adopted in 1993 (L.23-121995 n. 547), correspondence is also understood as an electronic one. In case of Polish Criminal Code, hacking crimes popular on the dark web are covered by articles 267 to 269. Art. 269 states that everyone acquiring unauthorized access to undedicated information in such form as opening closed envelope, connecting to telecommunication system, or avoiding electronic or magnetic protection, is subject to fine, restriction of liberty or imprisonment to up to two years. Par. 3 of the same article also deals with eavesdropping of communication in a variety of forms, including the use of dedicated software. Art. 268a stipulates that anyone unauthorized who destroys, damages, deletes or changes computer data is subject to imprisonment to up to three years. Art. 269 deals with the damage of information important for national security.

From the perspective of these regulations, the existence of hacking-on-demand services on TOR may be therefore prosecuted in two different ways. First, computer criminals offering these services on the dark web may be held accountable for such acts based on ICC's art. 615 ter and art. 616, as well as PCC's articles 267 to 269. If the attack did not take place, it can also be potentially categorized as an action meeting the criteria of fraud on the "buyer."

Second, the "buyer" is held accountable based on abetting to hacking-related crime, which is sanctioned by art. 378 of the ICC and art. 26 of the PCC. For example, according to Polish legal system, if the hacking did not happen, the buyer can still be prosecuted for attempted crime based on art. 29 of the PCC.

To summarize, this chapter addresses only a selected group of computer crimes, which takes place on the dark web. There is no doubt that due to its unique characteristics, such acts are difficult to fight against and prosecute. It applies also to American and European legal systems, which unfortunately have not adopted a unified approach to combating these offenses. Attempts to create specialized legislation are usually not successful or raise controversies.

These problems apply only partially to the Polish and Italian criminal law. Comparative analysis carried out above demonstrates that criminal laws in both countries usually would adopt similar interpretation of illegal acts occurring on The Onion Router. Most differences between them refer to slightly harsher penalties adopted in the Italian Criminal Code. The only major legal difference is related to how both countries perceive the prosecution of crimes related to child pornography on the dark web. The fact that the Italian legal system is more liberal than Polish when it comes to merely visiting (even by accident) websites with such content demonstrates that such discrepancies may be exploited by criminals. This case therefore demonstrates how important it is to unify national legal systems on computer-related crimes.

Bibliography

Rapid7. Dark Web Monitoring Solutions. Available at: https://www.rapid7.com/solutions/dark-web-monitoring.

Gozdyra, K. (2017) 'Penalizacja zjawisk stalkingu, cyberstalkingu i mowy nienawiści – próba oceny przyjętych rozwiązań prawnych na płaszczyźnie prawnokarnej', *Studia Administracji i Bezpieczeństwa*, 3 (3), pp. 121-149.

Kaczmarek, K. (2020) 'Darknet jako przedmiot badań nauk społecznych', *Cybersecurity and Law*, 4(2), pp. 105-113.

Moore, D., Rid, T. (2016) 'Cryptopolitik and the Darknet', *Survival,* 58(1), pp. 7-38.

Mozgawa, M., Nazar-Gutowska, K. (2014) 'Utrwalanie lub rozpowszechnianie wizerunku nagiej osoby – art. 191a k.k. (analiza prawnokarna i praktyka ścigania)', *Prawo w działaniu. Sprawy karne*, 19, pp. 7-40.

Radoniewicz, F. (2019). 'Przestępstwa komputerowe w polskim Kodeksie karnym', *Cybersecurity and Law*, 1(1), pp. 193-212.

Wróbel, M. (2014) 'Cyberprzestępczość w polskim systemie prawnym', *Wiedza Obronna*, 4, pp. 72-90.

Ziccardi, G. Riccio, G.M., Scorza, G. (2022) *Intelligenza artificiale. Profili giuridici*, Padova, CLEUP,.

Ziccardi, G. (2012) 'Cyberstalking e molestie portate con strumenti elettronici: aspetti informatico-giuridici', *Rassegna Italiana di Criminologia*, VI (3), pp. 160-173.

Zoller, M. (2021) 'Die Strafbarkeit des Betreibens krimineller Handelsplattformen im Internet – Der neue § 127 StGB', *International Cybersecurity Law Review*, 2, pp. 279-296.

Chapter XV
The Regulation of Hate Speech, Antisemitism, and Terrorism Online

by Arianna Arini*

Index: 1. Online hate speech. – 2. Online antisemitism. – 3. From anti-Jewish hatred to online terrorism.

1. Online hate speech

Hate speech – "hate speech", "hateful speech" or "hateful language"- refers to all violent, denigrating, hostile and discriminatory expressions and manifestations of thought aimed at the consolidation and radicalization, through language, of prejudices, stereotypes and hostility of certain groups of people (usually in a social position of dominance within a given historical context) towards other "minority" groups, defined on the basis of certain personal characteristics or beliefs, mostly relating to political, ethnic, religious, gender or sexual orientation spheres (Ziccardi, 2016).

From the general definition just outlined, an initial problematic aspect of the subject emerges: the difficulty for scholars of the subject, European institutions and jurists in identifying a unitary definition of the phenomenon. Hate speech can be read as a linguistic act that contributes to the creation (or, in this case, to the destruction or damage) of reality and, therefore, as an example of a "performative use of language", capable of modifying and creating effects on a material level (Di Rosa, 2020).

However, given that "destructive" hate language is embedded in a given reality, the latter being conditioned by the social and cultural circumstances of a given historical moment, defining it is anything but simple, making its crystallization into something fixed and immutable increasingly complex (Pino, 2008; Citron, 2014).

There is, in fact, no legal norm that precisely and comprehensively defines what hate speech is. There are, however, a variety of international and European soft and hard law sources that attempt to fill this gap.

* Research Fellow in Legal Informatics at the University of Milan and member of the Information Society Law Center (ISLC).

A first attempt at a definition can be found in the Recommendation of the Committee of Ministers of the Council of Europe of 30 October 1997, according to which:

> The term *hate speech* shall be understood as covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, antisemitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin.

But there are numerous definitions of hate speech that emerge in the framework of European and international law (Castellaneta and De Sena, 2020). In the Universal Declaration of Human Rights of 1948 (UDHR), for example, from a systematic reading of Articles 2, 7, 29(2) and 30, it can be inferred that hate speech is an abuse, and as such is prohibited, of freedom of expression (Di Rosa, 2020). The 1966 International Covenant on Civil and Political Rights also recognizes freedom of expression and some of its limitations, which are only permitted where expressly provided for by law and necessary to respect the rights or reputation of others or to safeguard certain situations of public and general interest. Article 20 also explicitly prohibits hate speech.

At EU level, the most relevant interventions include the European Convention on Human Rights, which recognizes the right to freedom of expression (Art. 10) and the prohibition of abuse of rights (Art. 17), as well as a general prohibition of discrimination. This prohibition of discrimination is a legally binding principle, enshrined today in Article 21 of the Charter of Fundamental Rights.

A more recent definition of hate speech - not even this, however, exhaustive - can be found in Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law. In addition, in 2015, General Recommendation No. 15 of the European Commission against Racism and Intolerance (ECRI) contributed to a significant broadening of the concept of hate speech, by including any conduct aimed at fomenting, promoting or encouraging, in any form whatsoever, denigration, hatred, defamation of a person or a group, as well as at subjecting the victim to a series of prejudicial situations, such as threats or harassment, on the basis of so-called prohibited grounds or other characteristics or personal status (Di Rosa, 2020).

Given therefore that there are objective difficulties in arriving at a universal definition of the phenomenon, referring to the individual legal systems for its normative definition, it is appropriate to highlight how the common element in the construction of hate speech is the focus on the principle of non-discrimination, including in its definition all the categories envisaged by the definition of discrimination according to its evolution in international law. The most correct path therefore seems to be to adopt a broad definition of hate speech that

includes the phenomenon not as mere "incitement" but as any expression that contains within it a bias, a discriminatory prejudice and, therefore, potentially capable of implementing discrimination.

Another rather controversial topic related to hate speech is that concerning the relationship - or rather, balancing, from a legal perspective - between the protection of freedom of expression and the expression of thought on the one hand and the protection of human dignity and human rights on the other. The freedom of expression of every citizen and the media plays a fundamental role in society: it is considered one of the pillars of a democratic society and an essential condition for guaranteeing the protection of other human rights (Conti, 2018). Indeed, the freedom of each citizen to freely express his or her ideas makes it possible to establish a positive dialogue not only on an individual level, but also on a collective level, enabling society to strengthen equal rights and the fight against discrimination. Although freedom of expression enjoys broad protection as a fundamental right, not all forms of expression are protected. Restrictions may apply under specific conditions and in cases of specific content, such as "forms of expression that spread, incite, promote or justify hatred based on intolerance": in practice, hate speech. At European level, freedom of expression is a fundamental right subject to balancing with others of equal rank, such as the protection of personal dignity and the principle of non-discrimination. Article 10(2) of the ECHR states that freedom of speech may be subject to restrictions necessary to safeguard the reputation and rights of others. Measures against hate speech are, therefore, fully legitimate if they are proportionate to the aim: they must, in other words, be based on a fair balance between freedom of expression and the other fundamental rights violated by hate speech. The Court of Cassation has also ruled in this sense in Judgment No. 36906/2015, stating that "in the possible conflict between freedom of expression and the equal dignity of citizens, preeminence must be given to the latter only in the presence of conduct revealing a concrete danger to the protected legal asset".

The picture becomes more complicated nowadays in the light of the social and cultural change brought about by the development of technology and the worldwide spread of platforms: whereas at one time hate speech and the use of discriminatory language took place verbally or through the use of traditional media (press, radio, television, etc.), today it is the Internet that represents the most prominent space for the dissemination of hate speech. It is unquestionable that the digital environment - and in particular that of social networks - has a far greater power to spread and publicize hatred than traditional media, as does the fact that hatred, once placed on the net, has a remarkable capacity for persistence and resistance to attempts to conceal offensive messages: the Internet therefore functions as a medium that facilitates the spread and potentiality of hatred, also due to the sense of impunity that derives, for many Internet users, from the (false) perception of being protected by anonymity.

According to a 2015 UNESCO document entitled "Countering online hate speech", the hallmarks of online hate speech would be the following:

1. the permanence over time of the manifestation of hatred: hatred and its contents circulate better online, are made more evident, are more widely shared and also more difficult to remove;
2. its extreme "viralisation" and continuous "recirculation" due to the exploitation of the same content by users of different platforms at different times. The ability to "share" and "like" can lead to a multiplicative effect, such that content can have a rapid and global dissemination without additional costs for the source (Kaesling, 2018);
3. the perception that the perpetrators of hate speech have of being protected by anonymity and the "disinhibiting" effect of technologies: the lack of an intermediation - or rather, the intermediation of a screen - can create a false sense of anonymity, of a "barrier", of distance that alters people's behaviour (Ziccardi, 2023);
4. the transnational dissemination of content and, consequently, its greater social impact than offline content;
5. the dissemination of such content through trending topics selected from the main social networks.

Another extremely dangerous feature of online hate speech concerns the effects caused by the interactions between people through the Internet in terms of the extreme nature of the opinions expressed: as Sustein pointed out (Il Mulino, 2003), groups of people who participate in online debate have a greater tendency to orient themselves ideologically towards extreme positions, a tendency that the author defines as "group polarization". This occurs through a creative process of online hatred that has characterized the network since its beginnings: normalizing the incitement to hatred under the pretext that it is irony, an apparently harmless form of expression but one that is capable of reinforcing stereotypes against weak individuals or groups (think for example of social techniques such as trolling or viral memes).

In recent years, there has been a significant increase in the presence and spread of hate speech in the form of xenophobic, nationalist, Islamophobic, racist and antisemitic content in online communication. Therefore, the spread of online hate speech is perceived as a social emergency that can lead to individual, political and social consequences. Moreover, the COVID-19 pandemic has increased feelings of hostility towards foreigners, but also towards weaker population groups, including the very young and the elderly.

It has become central, therefore, to take political and legal action to counter the phenomenon of hate speech, especially considering the impact of technology and the Internet on hate speech.

1.1. The normative approach to hate speech

Through the advent and development of technologies, hate speech has found "a new arena for its dissemination" (Casarosa, 2020). The strength of the spread of such speech in the digital space has prompted supranational and national legislators to prepare ad hoc regulatory instruments. However, the difficulty in finding a universal and unambiguous definition of hate speech has made it complex for institutions to adopt legislation to regulate hate messages.

International law offers protection against messages inciting hatred or discrimination, as mentioned above; not only that, but there is also an explicit call for action to prevent the dissemination of such messages (De Sena and Castellaneta, 2020). In fact, provisions aimed at combating incitement to hatred can be found in the Universal Declarations and Conventions dedicated to combating discrimination.

A decisive role in combating hate speech is played in the international arena by the Council of Europe, the Edu Court and the Committee of Ministers, which has adopted several recommendations to limit the spread of hate crimes and hate speech (see, for example, the Recommendation of the Committee of Ministers 97/2; Recommendation CM/Rec (2010) 5 of the Committee of Ministers to member states on measures to combat discrimination on grounds of sexual orientation or gender identity).

Hate speech is one of the forms of discrimination banned by the European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR); moreover, given the new ways in which hate messages can be spread on the Internet, in 2003 the Council of Europe signed the Additional Protocol to the Budapest Convention on Cybercrime: a document that obliges the member states to adopt criminal sanctions if racist or xenophobic content is found on computer systems.

At the level of the European Union proper, we find Article 21 of the Charter of Fundamental Rights of the European Union, which prohibits any form of discrimination; or the previously mentioned framework decision (2008/913/JHA) of the Council, which obliges member states to punish xenophobic or racist behaviour and incitement to hatred. The European Parliament has called for a revision of the framework decision, with the aim of including in the list of expressions that can be defined as hate messages also those based on antisemitism, religious intolerance, anti-Gypsyism, homophobia and transphobia.

Although several international treaties and agreements include regulation of hate speech, at European level, an agreed framework is still lacking and, as mentioned above, the only common point of reference is the 2008 Framework Decision, whose implementation by individual countries has, however, been less effective than expected, as national legislative interventions have followed different approaches.

Given the unsatisfactory results that have emerged from the adoption of a hard-law approach by the European institutions, there has been a gradual move towards a soft-law approach to hate speech. On 30 May 2016, the European Commission adopted the "Code of Conduct on Countering Illegal Hate Speech Online", signed by the major players in the online market, including Facebook, Google, Microsoft, Twitter (now X) and Instagram. The document shows a clear intention to protect freedom of expression but also the need on the part of civil society organizations to prevent the spread of hate online. It is expected that every private entity will be required to include a definition of hate speech in its regulations and to provide a comprehensive explanation of how such expressions are sanctioned. Furthermore, signatories are required to adapt their internal procedures to ensure a swift and incisive response in the event of hate speech conveyed through their online platforms: in particular, companies are asked to assess "most valid notifications for the removal of unlawful hate speech in less than 24 hours" and to remove or disable access to such content if necessary. The approach adopted by the European Commission is thus more focused on the timely removal of alleged hate speech than on the procedural safeguards that such a private enforcement mechanism should adopt in order not to unreasonably restrict users' freedom of expression. In the light of these objectives, we can observe that online platforms have progressively adapted to the expected standard, but the latest impact assessments of the Code show a slowdown in progress: the 2022 results show a decrease in the results in terms of notification and action by companies. The number of notifications examined by companies within 24 hours decreased compared to the last two monitoring exercises, from 90.4% in 2020 to 81% in 2021 and 64.4 % in 2022.

To achieve the expected results, the signatory companies have adopted several technological tools to evaluate and recognize content uploaded onto their platforms, mainly through the refinement of the algorithms used in them. However, these algorithms can only mark content based on certain keywords, which are continuously updated, but always lag behind the evolution of the language. In order to overcome this problem, one strategy used by companies has been to include a collaboration between the algorithm and human intervention, not only through reports made by individual users, but through forms of content control both internally (assigning a team of employees the task of checking sensitive cases) and externally (involvement of so-called trusted flaggers, individuals or organizations with special skills and experience in recognizing hate speech). These precautions and obligations in terms of proactive monitoring, however, bring with them a number of flaws and problems: first of all, by assessing hate content against standards specific to each platform, there is an extension of the legal definition of hate speech provided by the 2008 Framework Decision (entailing a privatization of enforcement systems in favour of private platforms with excessive interference in users' freedom of expression and the

need to implement on the platforms a control over the entire communication flow, potentially in conflict with the general prohibition of control over service providers under Art. 15 of the e-Commerce Directive). Then there is the problem of the absence of specific requirements from the point of view of procedural safeguards for users, neither in terms of judicial proceedings nor through alternative dispute resolution mechanisms, so that it is left to the IT companies to introduce a redress mechanism. Safeguards to limit the risk of removal of legal content are instead provided for in the Commission's Recommendation on the fight against illegal content online, which includes in the broader definition of illegal content also hate speech (Commission Recommendation 2018/334 on fighting illegal content online, C/2018/1177, OJ L 63, 6.3.2018): it aims at the identification and automated removal of content and emphasizes the need for a counter-notice in case of removal of legal content.

A final consideration concerns the liability regime of platforms that, with the proactive monitoring system just described, risk losing the exemption from liability provided by the e-Commerce Directive in Article 14. If the hosting provider takes technical measures to detect and eliminate hate speech, can it maintain its neutral position regarding content? However, the exemption from liability can only apply if the conditions set out in Article 14 of the Directive apply: it follows that proactive measures taken by the hosting provider may lead to the platform in question becoming aware of illegal content, which could result in the loss of the exemption from liability. However, if the hosting provider acts swiftly to remove or disable access to the content after having acquired such knowledge or awareness, it continues to benefit from the exemption from liability.

Finally, there is the Digital Services Act, presented by the European Commission on 15 December 2020 and approved as EU Regulation 2022/2065 on 19 October 2022. This document is part of the Commission's programme to create "a Europe fit for the digital age" and has the function of regulating the activities of online platforms. Thus, a new system for reporting and removing illegal content by platform users is introduced at European level. It also includes mechanisms for reporting content posted online by users and an easy redress mechanism.

At the level of individual European states, the strictest hate speech legislation is the one adopted by Germany with the enactment of the Network Enforcement Act (NetzDG) in 2018, which obliges any social media network to remove offensive content within 24 hours of publication, under penalty of a hefty fine. In Italy, there is currently no form of regulation of online hate or ad hoc criminal legislation to counter hate speech: the legal vacuum is filled by the application of the regulations on the crime of incitement to racial hatred, propaganda of ideas based on superiority and racial hatred, insult, defamation and threats, the reference legislative text of which is the so-called Mancino

Law (Law No. 205 of 25 June 1993). In the past few years, draft laws aimed at combating online hate have been presented, such as the bill "Measures for preventing and combating the dissemination of hate speech through the Internet", filed on 10 March 2021 and similar in content and method to the German NetzDG. The 2019 Agcom regulation on hate speech, on the other hand, lays down new rules of conduct for providers of audiovisual and radio information and entertainment services: in the event of serious and systemic violations, a sanctioning procedure will be initiated by Agcom.

2. Online antisemitism

One of the manifestations of hate speech that has known and still knows a sad and long history of discrimination is certainly that related to the Jewish people. Speaking of antisemitic discrimination, one cannot omit an analysis in defining terms of the phenomenon, declined in various forms and terms (such as antisemitism, Holocaust, anti-Zionism, denial and minimization of the Shoah), as well as the relationship between antisemitism and the Shoah, the absolute exacerbation in terms of antisemitic hatred. And it is by delving into these different forms of antisemitism that the phenomenon of denial, minimization and trivialization of the Shoah then emerges. Lastly, we investigate how antisemitism is developing with the new forms of communication, through the media, social networks, and messaging apps, where the language of antisemitic hatred is amplified and reproduced at an exponential level, often concealed and "justified" by the use of social humour (*e.g.*, the use of memes) capable of de-sensitizing the user towards such issues and normalizing strongly discriminatory content.

First, it is important to define what antisemitism is: to understand the phenomenon, it is worth noting the difficulties encountered over the years in its conceptualization. The first definition is given by the International Holocaust Remembrance Alliance (IHRA) in 2016:

> Antisemitism is a certain perception of Jews, which may be expressed as hatred towards Jews. Rhetorical and physical manifestations of antisemitism are directed towards Jewish or non-Jewish individuals and/or their property, towards Jewish community institutions and religious facilities.

Non-binding operational definition that supports national governments and anyone interested in the issue of antisemitism to take informed and effective action to prevent and combat it (NOA, 2022). It represents the first internationally agreed conceptualization of antisemitism in history and has, therefore, become the main definition of antisemitism used worldwide. The International Holocaust Remembrance Alliance (IHRA) would like to intervene on the

spelling of the term "antisemitism", often rendered as "anti-Semitism": this is because the IHRA believes that the hyphenated spelling leads to the creation of a so-called "Semitism" that not only legitimizes a pseudo-scientific form of racial classification that has been completely discredited by its association with Nazi ideology, but also divides the term, depriving it of its meaning of opposition and hatred towards Jews.

Since some historians considered the IHRA definition of antisemitism perfectible, a second type was provided: the document Jerusalem Declaration on Antisemitism proposes "a definition of antisemitism in relation to racism and other forms of discrimination, offering a series of reflections and suggestions on the historical and contemporary analysis of these phenomena" (Sullam, 2021).

The third definition published in March 2021 is the Nexus Document, a project of the University of Southern California's Annenberg School that recognizes that:

> antisemitism consists of systemic anti-Jewish beliefs, attitudes, actions, or conditions. It includes negative beliefs and feelings about Jews, hostile behaviour directed against Jews (as Jews), and conditions that discriminate against Jews and significantly impede their ability to participate as equals in political, religious, cultural, economic, or social life. As the embodiment of Jewish collective organization and action, Israel can be a target of antisemitism and antisemitic behaviour. Therefore, it is important for Jews and their allies to understand what is and what is not antisemitic in relation to Israel.

According to Kenneth. L. Marcus:

> Antisemitism is a set of negative attitudes, ideologies, and practices directed at Jews as Jews, individually or collectively, based on and sustained by a repetitive and potentially self-fulfilling latent structure of hostile mistaken beliefs and assumptions that stem from the application of double standards to Jews as a collectivity, manifested culturally in myth, ideology, folklore, and imagery, and urging various forms of restriction, exclusion, and suppression (Kenneth and Marcus, 2015).

It is necessary to make a further terminological clarification: when the discourse focuses on the Jew as a person and creates prejudices based on his way of being, behaving or thinking, one speaks of Jewishophobic antisemitism. When the hatred focuses on the events of the Holocaust and the extermination justified, denied or belittled, it is called denialist antisemitism. Finally, if the discourse concerns the state of Israel and criticizes Israel for its actions, denying its existence or not, then it is called anti-Zionist antisemitism.

Antisemitism is inevitably connected, in its crudest imagery, with the extermination of the Jews (as well as Roma, Sinti, homosexuals, people with disabilities,

Slavic peoples, political opponents and Jehovah's Witnesses) during the Second World War (Levi and Rothberg, 2003). The term "Holocaust" is used to refer to the extermination of the Jewish people and means "sacrifice by burning", but can also be defined as Shoah, which means "catastrophe". From a Jewish point of view, the correct term to refer to the extermination of the Jews during Nazism would be Shoah, but the term generally used is Holocaust. Another definition of the Holocaust is as follows:

> The Holocaust was the systematic persecution and murder of Jews by Nazi Germany and its collaborators between 1933 and 1945. It was a continent-wide genocide, which destroyed not only individuals and families, but also communities and cultures that had developed over centuries (IHRA, 2019).

The UN (UN General Assembly Resolution on Holocaust Denial) defines Holocaust denial or distortion as:

> (a) Intentional efforts to justify or minimize the impact of the Holocaust or its main elements, including collaborators and allies of Nazi Germany, (b) Gross minimization of the number of Holocaust victims in contradiction to reliable sources, (c) Attempts to blame the Jews for causing their genocide, (d) Statements presenting the Holocaust as a positive historical event, (e) Attempts to obfuscate responsibility for the creation of concentration and extermination camps designed and operated by Nazi Germany by blaming other nations or ethnic groups" (Resolution A/RES/76/250, 20 January 2022).

According to the IHRA operational definition of Holocaust denial and distortion:

> Holocaust denial is discourse and propaganda that denies the historical reality and extent of the extermination of Jews by the Nazis and their accomplices during World War II, known as the Holocaust or Shoah. Holocaust denial specifically refers to any attempt to claim that the Holocaust / Shoah did not take place.

In addition to the explicit repudiation of the Holocaust, denialism includes the minimization, trivialization and relativization of facts and events that occurred during the Holocaust (Wistrich, 2012). The theses of the deniers are as follows: the non-existence of the gas chambers; the minimization of the number of Jews killed by the Nazis; the impossibility of blaming the Holocaust on the Nazis alone; and the belief that the Holocaust was an invention of allied propaganda (Pisanty, 2014).

Finally, assessing how antisemitism has taken root in modern digital society, it should be noted that today, antisemitic hatred is still very much present and proliferates due to the mutual feeding of misinformation, hatred towards Israel, social platforms and antisemitic stereotypes. Monitoring work carried out by

the European Agency for Fundamental Rights (FRA) has found that expressions of antisemitism in online forums are on the rise in the European Union, mainly including antisemitism in media and communication, antisemitism in social media (FRA, 2018). In 2017, the annual "Digital Terrorism and Hate" online report published by the Simon Wiesenthal Centre identified 24 antisemitic games accessible on the Internet, as well as antisemitism and hate speech on social networking platforms, blogs, messaging apps, discussion forums, video channels and other online media (Simon Wiesenthal Center, 2017). Sometimes, antisemitism can be expressed through symbols on social media posts, handles, blogs and Twitter threads, often targeting high-profile personalities and celebrities (e.g. Liliana Segre). Furthermore, according to an FRA survey on discrimination and hate crimes against Jews in the EU (2018), in twelve countries Internet antisemitism is among the most perceived problematic forms of antisemitic discrimination.

The virtual sphere of the Internet constitutes an environment where attitudes of prejudice and hatred against Jews can proliferate, not only in online platforms that are less known to the general public, where content moderation and removal activities are limited, but also, with partially different forms, in large mainstream platforms, such as Facebook, Instagram, YouTube, Reddit and Twitter. It seems interesting to point out that the level of antisemitism present in a piece of content depends on the guidelines of the platform's community. It is therefore not surprising that platforms such as Discord, 4chan, 8kun, Gab and Parler, which are less known and controlled than Facebook or Telegram, are a "home" for antisemitic content. In general, on social networks, humour is used - by the authors of antisemitic content - as a justification for antisemitic positions; moreover, humour can "desensitize the use of violence" (cit. Marone, 2023). Memes are used on the Internet because they "help to confer social acceptability and normalize antisemitism and racism within society".

Today, antisemitic online groups feed each other's extremism: the essence of their ideals finds its greatest outlet in conspiracy theories, which see Jews as puppeteers of dark powers. Online antisemitism has risen sharply especially during the pandemic period, where conspiracists have found fertile ground to sow their ideas, holding Jews responsible. Global Conspiracy Theory, in the sense of the myth of a Jewish conspiracy, which finds its roots in the New Testament's "imputation of responsibility to the Sanhedrin for calling for the arrest of Jesus" is an important aspect of contemporary antisemitism (Veca, 2019): one need only think of the alleged Jewish control of the media, the economy, the government or institutions.

The unlimited power of the Jews is another profile that is linked to conspiracy theory; antisemitic stereotypes are dangerous because they can generate discrimination and hate crimes.

Moreover, the COVID-19 pandemic "demonstrated how old antisemitic prejudices can resurrect and fuel new conspiracy myths and hatred online and offline" (European Commission. Towards an EU Free from Antisemitism. EU Strategy on Combating Antisemitism and Fostering Jewish Life (2021-2030), p. 2).

See, for example, the use of the term Holocaust during the dissemination of COVID-19, which is "rooted in the historical practice of linking Jews to disease, providing antisemites with familiar stereotypes" (American Jewish Committee, 2021). Translate Hate. Stopping Antisemitism starts with understanding it, p. 12. As has been written, conspiracy theories about COVID-19 "are helping to spread antisemitic beliefs to a wider audience" (Jee, 2021). According to a paper that was published by CST (Protecting our Jewish Community), conspiracy theorists believe that the virus is a Jewish conspiracy (fake or real) and Jews are believed to be the main spreaders of the virus. In addition to classic conspiracy theories, there are now anti-Jewish motifs in "contemporary conspiracy fantasies" such as QAnon (Marone, 2023), which develops classic antisemitic fantasies, such as the accusation of blood against children (Molle, 2021), or "white genocide" and the "Great Replacement", according to which Jews intend to destroy Christian and white civilization (Marone, 2023).

2.1. Legal framework and strategies to counter antisemitism

Considering the legal profiles in countering the spread of antisemitic hatred, offline and online, and starting from the international perspective, it is worth mentioning the contribution made by both the United Nations and the Council of Europe.

As far as the United Nations is concerned, its involvement in the fight against antisemitism is mainly limited to acts of *soft law,* as no treaty explicitly addresses the subject. Among the main interventions: the General Assembly designated 27 January as the International Day of Remembrance in commemoration of the victims of the Holocaust in 2005 (A/RES/60/7); adopted resolution 61/255 in 2007 condemning Holocaust denial (A/RES/61/255); held its first informal meeting in New York on 22 January 2015 on the rise of antisemitic violence in the world; on 26 June 2019, it held an informal meeting entitled "Combating antisemitism and other forms of racism and hatred - The challenges of teaching tolerance and respect in the digital age", while three years later, on 13 January 2022, it doctored a new Resolution on Holocaust denial (A/RES/76/250).

The Council of Europe, which is considered the first organization to recognize "the relationship between antisemitism and racism" (Roubache, 2000), like the United Nations, does not yet have binding legislation explicitly sanctioning antisemitism; however, the European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR) can be considered the key treaty

to combat it within the broader framework of the European system of human rights protection. This is due to the jurisprudence of the European Court of Human Rights, which has ruled on several cases concerning Holocaust denial and antisemitism, based on Article 10 ("Freedom of expression") and Article 17 ("Prohibition of abuse of rights") of the ECHR. In its 30-year evolution, there has been the progressive and frequent use of Article 17 to deny Article 10 protection in cases of antisemitism and, in particular, Holocaust denial.

Soft law acts include: the Action Plan on Combating Racism, Xenophobia, Antisemitism and Intolerance, adopted as part of the Vienna Declaration on 8-9 October 1993 and continued since then; a large number of Recommendations of the CoE Committee of Ministers and Recommendations and Resolutions of the CoE Parliamentary Assembly, covering, among other areas of concern, cyberspace, including cybercrimes related to antisemitism; the 1997 Hate Speech Recommendation, which explicitly includes antisemitism among hate speech; the 2007 Resolution on Combating Antisemitism in Europe, which contains an exhaustive list of recommendations to governments; and the 2016 Resolution 2106 for a renewed commitment to combat antisemitism in Europe.

As far as online forms of antisemitism are concerned, the First Additional Protocol to the 2003 Council of Europe Convention on Cybercrime (also known as the Budapest Convention), concerning the criminalization of acts of a racist and xenophobic nature committed through computer systems, can play a key role, especially in light of the clarification by the European Commission against Racism and Intolerance (ECRI) that the Additional Protocol "requires States Parties to prosecute Holocaust denial if the offence is committed with the intent to incite hatred, discrimination or violence online".

As an important tool for combating antisemitism, and in particular denialism on social media, the 2022 Report of UNESCO and the United Nations (in collaboration with the World Jewish Congress, WJC), in which the first findings on Holocaust distortion and denial on the major social media platforms (Facebook, Instagram, Telegram, TikTok and Twitter) were published, should be considered. The report shows that Holocaust denial and distortion is massive on Telegram: on this platform, almost half (49%) of public content related to the Holocaust denies or distorts the facts. This percentage rises to over 80% for German-language messages (about 50% for English and French). Less relevant is the impact on other platforms. Denial and distortion affect 19% of Holocaust-related content on Twitter, 17% on TikTok, 8% on Facebook and 3% on Instagram. The researchers pointed out that Holocaust denial has changed form: people often try to avoid fact-checking content, using humorous memes and parodies.

The report shows how "a fundamental part of the United Nations' mission" is to combat antisemitism online. Antonio Guterres, UN Secretary-General, said:

> Holocaust denial and distortion take many forms online. The report outlines this worrying phenomenon and makes it clear that it can no longer be ignored. What is clear is that when platforms make a concerted effort to address this unique form of hate speech, it leads to results. But more needs to be done to eliminate it. As Holocaust deniers become more sophisticated, so must all those who are working to combat this evil.

Among the practical recommendations to avoid this misinformation, UNESCO and UN have suggested several measures. Starting with monitoring online platforms and, if necessary, intervening in content that denies or distorts the Holocaust. Platforms should redirect and give visibility to verified information about the Holocaust story, as Facebook and TikTok do in their partnership with UNESCO and the WJC, with the website aboutholocaust.org. Furthermore, platforms should actively collaborate with teachers and education systems to develop teaching and learning resources and support education (including digital) in schools, universities and non-formal education.

The European Union, on the other hand, has an extensive body of legislation to combat antisemitism, both at the level of primary legislation (see Articles 2, 10 and 19 of the Treaty on the Functioning of the European Union and Articles 1 and 21 of the Charter of Fundamental Rights of the European Union) and at the level of secondary legislation, covering a wide range of relevant acts. The latter include Directive 2000/43/EC implementing the principle of equal treatment between persons irrespective of racial or ethnic origin and Directive 2000/78/EC establishing a general framework for equal treatment in employment and occupation; in 2008, the EU took a strong stance against racism and harmful content with the Framework Decision on combating certain forms and expressions of racism and xenophobia by means of criminal law ("Framework Decision"). This text provides a solid legal framework to combat antisemitic hate crimes and hate speech, including the denial or gross trivialization of the Holocaust in a way that incites violence or hatred. The Preamble explicitly refers to antisemitism.

Two years later, the EU adopted Directive 2010/13/EC on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive - AVMSD), which was revised in 2018 (Revised AVMSD), where, in Article 6, Member States are required to ensure that audiovisual media services provided by media service providers "do not contain any incitement to hatred based on race, sex, religion or nationality". The revised directive contains many innovations that can have an impact on the fight against antisemitism, especially with regard to minors. Indeed, this law strengthens the protection of minors from "harmful content" in the online world, including greater protection for video-on-demand services. The period

in which the revision of the AVMS Directive took place coincides, among other things, with the EU's development of a legal framework for the digital space, culminating in the adoption of the aforementioned Digital Services Act (DSA), a text that contributes to protection against antisemitism in the online sphere, given the notion of "illegal content" explained in recital 12.

In terms of soft law, there are several texts that deal with antisemitism: 2021 can be considered a watershed year with the adoption of the first EU Strategy on combating antisemitism and fostering Jewish life (2021-2030), which states that "antisemitism is incompatible with Europe's fundamental values".

It should be emphasized that measures taken by the EU to counter antisemitism include the establishment of the Antisemitism Coordinator in December 2015; the establishment of an EU High Level Group on Combating Racism Xenophobia and Other Forms of Intolerance in 2016; and the operational definition of antisemitism by the IHRA in the same year. However, the Strategy represents a significant step forward in the fight against antisemitism in the EU. It is a comprehensive strategy adopted by the EC based on three pillars: preventing and combating all forms of antisemitism; protecting and promoting Jewish life in the EU; and education, research and remembrance of the Holocaust.

Lastly, it is worth mentioning that on 16 November 2023, UNESCO and the Permanent Representation of the European Union convened an urgent discussion on the role of education in addressing the recent global rise of antisemitism, following the very recent events in the Arab-Palestinian conflict. Indeed, since 7 October 2023, monitoring bodies in Europe and North America have recorded an exceptional spike in antisemitic hate speech, online and offline, and violent incidents. In particular, the online dimension of the ongoing crisis, which has become impossible to handle, has been addressed (suffice it to say that, according to the Anti-Defamation League, a Jewish advocacy group, since 7 October, antisemitic content on Twitter has increased by over 900%): as part of a response, UNESCO published guidelines for the governance of digital platforms, which aims to safeguard the right to freedom of expression and other human rights in the governance of digital platforms, while dealing with content that can be lawfully restricted under international human rights laws and standards.

Stefania Giannini, Deputy Director-General for Education at UNESCO, outlined the global educational response to antisemitism that UNESCO has been supporting since 2018: mobilizing against Holocaust denial and distortion through education and public campaigns, teaching about the Holocaust and genocide, incitement to hatred, and online misinformation, as well as developing tailored programmes to support educators at all levels to prevent and mitigate the impact of antisemitism.

A final look at some national legislations addressing the issue of countering anti-Jewish hatred is worthwhile. Some Member States have provided for the punishment of denialism in their penal codes; others have introduced special legislation.

In 1995, Belgium passed the law on denialism, which provides for "a prison sentence of eight days to one year and a fine of 26 to 5,000 Belgian francs (approximately EUR 124) for anyone who denies, grossly trivializes, attempts to justify or approves of the genocide committed by the National Socialist German regime during the Second World War". The first conviction for an antisemitic demonstration on the Internet took place in 2021.

In Spain, in order to punish genocide denial, the latter must be accompanied by the conduct of inciting hatred against certain groups of people.

In Italy, Holocaust denial is an aggravating circumstance, as provided for in Article 604-bis of the Criminal Code. According to this provision, the penalty of imprisonment is from two to six years "if the propaganda or incitement, based in whole or in part on the denial, gross minimization or apologia of the Shoah or of crimes of genocide, crimes against humanity and war crimes, as defined by the Statute of the International Criminal Court, represent a concrete danger or wider dissemination". In Italy, the Italian Extraordinary Commission for Combating the Phenomena of Intolerance, Racism, Antisemitism and Incitement to Hatred and Violence, chaired by Senator for Life Liliana Segre, approved an important document in the previous Legislature: "Comparative analysis on the phenomenon of incitement to online hatred", which investigated and defined areas such as hate speech, its online dissemination, particularly on social media, and its impact on individual and collective rights. During the summary report no. 2 of the Commission's work held on 20 September, the chairwoman recalled the importance of the research and analysis work carried out previously and which will have to continue in the current legislature, with particular attention to the impact of Artificial Intelligence on the spread of hate speech and the phenomena of intolerance and antisemitism.

In Romania, the first legal document explicitly focused on Holocaust denial was Emergency Government Ordinance No. 31 of March 2002, on the "banning of fascist, racist or xenophobic organizations and symbols and the promotion of the worship of persons guilty of crimes against peace and humanity". According to Article 6 of Law 217 of 23 July 2015, which updates EGO No. 31 of 2002, "Denial, disputing, endorsement, justification or trivialization in a blatant manner by any means in public of the Holocaust or its effects shall be punishable by imprisonment from six months to three years and a fine". Another important step was taken with the adoption in 2018 of the Law on Preventing and Combating Antisemitism (Law 157/2018), which specified which types of organizations are prohibited. In addition, since 2016, the National Institute for the Study of the Holocaust "Elie Wiesel" has been

involved in the systematic monitoring of antisemitism in the country, first with projects focused on social media from 2016 to 2018 and, since 2019, with the publication of annual monitoring reports.

3. From anti-Jewish hatred to online terrorism

Anti-Jewish hatred occupies a prominent position in various forms of violent extremism, especially in the extreme violent right and in jihadism, going so far as to induce the execution of serious terrorist attacks.

Since 7 October 2023, the day of Hamas' terrorist action, antisemitic and Islamophobic hate speech has soared on social media platforms (from Twitter to Telegram to 4chan).

As highlighted in the previous section, since that date, monitoring bodies in Europe and North America have recorded an exceptional spike in antisemitic hate speech, online and offline, and violent incidents.

The examination cannot but begin with X, formerly Twitter, and the anti-Jewish hate campaign, for several reasons: because of the huge scale of the phenomenon; because of the direct involvement of X's father-in-law Elon Musk; because of the battle, even judicial, waged by Musk against certain organizations guilty, in his view, of falsifying data.

According to the Anti-Defamation League, almost two million posts with the hashtag #IsraeliNewNazism and another 40,000 with #ZionistsAreEvil or #ZionistsAreNazis appeared on X during that period. Memetica, a digital investigation company, reported that the hashtag #Hitlerwasright had 46,000 uses since 7 October. The Center for Countering Digital Hate (CCDH) identified 200 posts that promoted antisemitism and other forms of hate speech in the conflict, and 196 of these remained on the platform. Seventy-six accumulated 141 million views in 24 hours after the explosion at the Gaza City hospital on 17 October.

TikTok, on the other hand, has been criticized for videos against Israel or in support of Palestinians bearing the hashtag #freepalestine. Dozens of high-profile Jewish content creators sent an open letter complaining that the platform has not done enough to counter hatred and abuse against the Jewish community. According to the Chinese-owned company, more than 730,000 videos containing hate speech and antisemitism have been removed since 7 October, and it has set up a virtual "command centre" with Arab and Jewish moderators in Singapore, Ireland and the United States, monitoring war-related content 24 hours a day.

Turning to other web realities and anti-Islamic content, the Global Project Against Hate and Extremism found that antisemitic and anti-Muslim posts on 4chan, Gab, Odysee and Bitchute increased by 461% from 6 to 8 October.

Messaging apps such as Telegram were also used to spread hate speech in the conflict. A Hamas-linked Telegram channel shared an image of a paraglider descending with a Palestinian flag and the words "I stand with Palestine", referring to the Hamas gunmen who used the very paragliders to enter the Nova music festival, killing more than 260 people. Within 24 hours, the image was shared thousands of times on X, Instagram, Facebook and TikTok, according to ActiveFence, a cybersecurity company that advises platforms.

These forms and manifestations of online hatred, which have been on the rise in recent months due to the Arab-Palestinian conflict, also lay the foundations for reflection on an issue that is becoming both central and worrying at the same time: the phenomenon of so-called "cyber-terrorism". This is the complex of criminal activities carried out using the Internet and which aim to spread fear and destabilize security, motivated mostly by religious, political and ideological foundations.

An early presence of terrorists online has been documented as dating back to 1998 by the U.S. Antiterrorism and Effective Death Penalty Act of 1996, for which more than half of the organizations defined as "foreign terrorist organizations" had their own publicly accessible website online. However, the actual emergence of online terrorism is connected to the proclamation in June 2014 of ISIS, which managed to aggregate various factions belonging to Islamic radicalism to pursue a totalitarian regime and declare war on opponents. All the groups that are part of ISIS are perfectly aware that to carry out their project, *i.e.*, the recruitment of the military and the spread of the programme, they must maintain an active presence within cyberspace. This explains why the Internet has assumed a decisive role in terrorist strategies.

Online terrorism is mainly based on two actions: propaganda and direct activity. In the first case, the objective is to disseminate a certain thought capable of winning over the public: for this, the web is used, which makes it possible to reach so many people all over the world in a very short time. Direct activities, on the other hand, consist in exploiting cyberspace as a tool to strike and give a clear demonstration of strength (think of demonstration attacks, such as piracy actions aimed at attacking institutional websites or sophisticated intrusions into computer systems).

The Internet has thus become both a tool for the visibility of terrorist activities and a new means of carrying out cyberattacks for terrorist purposes (Strano *et al.*, 2002).

Terrorist attacks in Europe in recent years have been carried out by citizens born and raised on the continent, radicalized without leaving their country of origin, but who have exploited social media and the dark web as factors in their radicalization.

The Internet and social media are proving to be powerful tools in the hands of terrorist groups, allowing them to communicate, spread messages, raise funds,

recruit supporters, inspire and coordinate attacks and target vulnerable people: consider that on 2020 Referral Action Day, Europol and 17 Member States identified and assessed for removal as many as 1,906 URLs linking to terrorist content on 180 platforms and websites in one day. Facebook indicated how, over the course of two years, it had to remove more than 26 million pieces of content from groups such as the Islamic State of Iraq and the Levant (ISIL) and Al-Qaeda.

G. Weimann, in his book Terror on the Internet. The New Arena, the New Challenges (Weimann, 2006), describes how terrorists exploit the Internet daily to carry out their deadly plans and identifies seven purposes for which it is used:

1. Data mining: terrorists use the web to research key details on targets such as nuclear power plants and airports, and even on anti-terrorism measures. They can extract 80% of the information they need for an attack from the Internet using sources that are legally accessible to the public;
2. Network: the Internet enables different terrorist groups to communicate and coordinate their activities effectively, reducing communication costs and increasing the variety and complexity of information that can be shared;
3. Recruitment and mobilization;
4. Online instructions and manuals: the Internet is full of sites offering manuals that teach readers how to build chemical and explosive weapons;
5. Planning and coordination of specific attacks;
6. Fundraising: using demographic data gathered from personal information entered in online questionnaires and order forms, terrorists use the Internet to identify likely sympathizers, who are then solicited to make donations through e-mails sent by groups representing terrorists;
7. Attacking other terrorists: the network is also used as a virtual battleground between (and within) terrorist organizations. Debates between groups like Hamas, as well as conflicts within the groups themselves, take place on their websites.

In the new digital scenario of modern terrorism, countering this phenomenon becomes as important as it is difficult: the Internet has always been seen, by supranational and country legislators, as a technology that is as powerful and transparent as it is difficult to control (think of the problem of anonymity).

Such concerns related to the world of the Internet had first led the international legislator to adopt a rather cautious and respectful approach to technology and freedom of expression with the Madrid Charter of 2005.

In recent years, however, with the increasing development of cyber-terrorism, the major online platforms have implemented a whole series of counter-terrorism measures and several European policies have been initiated to significantly limit web-based terrorist attacks and all terrorism risks in general.

In the UN Global Counter-Terrorism Strategy (A/RES/60/288), Member States agreed to work with the UN to explore ways to coordinate international

and regional efforts to counter terrorism in all its forms and manifestations on the Internet, and to use the Internet as a tool to counter the spread of terrorism.

On 25 March 2017, the European Union adopted the Directive on Combating Terrorism (EU Directive 2017/541) under which all countries must ensure that they criminalize conduct such as training and travel for terrorism, as well as the financing of terrorism itself, through harmonized definitions of terrorist offences, which serve as a benchmark for cooperation and information exchange between national authorities. Recital 6 expressly calls for the punishment of terrorist conduct also where "carried out through the Internet, including social networks". Furthermore, the removal at source of online content that constitutes public provocation to commit a terrorist offence or, where this is not possible, the blocking of access to such content, is considered an effective means of combating terrorism on the Internet; furthermore, voluntary action taken by the Internet industry to prevent abuse of its services, or any support by Member States for such action, such as the detection and reporting of terrorist content, is not prejudiced; no general obligation should be imposed on service providers to monitor the information they transmit or store, nor to actively seek out facts or circumstances indicating illegal activity; finally, hosting service providers should not be held liable as long as they do not have actual knowledge that the activity or information is illegal and are not aware of facts or circumstances that make the illegality of the activity or information apparent.

On 24 July 2020, the European Commission adopted a new EU Security Strategy covering the period from 2020 to 2025, which sets out tools and measures to be developed in the coming years to ensure security in our physical and digital environment. On 9 December 2020, on the other hand, it adopted a new Counter-Terrorism Agenda for the Union, which sets out a way forward for counter-terrorism actions, seeking to better anticipate, prevent, protect against and respond to terrorist threats.

On 29 April 2021, the EU Parliament and Council adopted the "Regulation on addressing the dissemination of terrorist content online" (EU Regulation 2021/784), which establishes a clear and harmonized legal framework and defines the responsibilities of Member States and the obligations of hosting service providers to effectively and expeditiously detect and remove terrorist content online from their platforms in order to counter the misuse of hosting services for the dissemination of terrorist content online. Thus, the competent authority of each Member State will have the power to issue a removal order requiring platforms to remove or block access to terrorist content throughout the EU. Terrorist content will be removed within a maximum period of one hour after receipt of the removal order.

Platforms will therefore have to take specific, reasonable and proportionate measures to protect their services from the dissemination of terrorist content online, but the choice of such measures is left to each platform, including the

adoption of automated measures, adapted according to the capabilities of the service provider and the nature of the services offered. In addition, if the competent authority considers that the specific measures put in place are not sufficient to address the risks, it may require the adoption of additional appropriate, effective and proportionate specific measures. Finally, the types of sanctions applicable in the event of violation of the Regulation are left to the individual Member States: in Italy, for example, Legislative Decree No. 107 of 24 July 2023, implementing the Regulation, provides for both administrative and criminal sanctions.

However, one has to consider how cyber-terrorism is constantly changing and therefore employs new strategies to instill fear and mistrust in people. Despite the precautions taken at the international level, the phenomenon has still not been stopped completely and continues to cause concern.

Bibliography

Casarosa, F. (2020) 'L'approccio normativo europeo verso il discorso dell'odio online: l'equilibrio fra un sistema di "enforcement" efficiente ed efficace e la tutela della libertà di espressione', *Questione Giustizia,* 8 luglio. Available at: https://www.questionegiustizia.it/articolo/l-approccio-normativo-europeo-verso-il-discorso-dell-odio-online-l-equilibrio-fra-un-sistema-di-enforcement-efficiente-ed-efficace-e-la-tutela-della-liberta-di-espressione.

Castellaneta, M. e De Sena, P. (2020) La libertà di espressione e le norme internazionali, ed europee, prese sul serio: sempre su Casapound c. Facebook. [Blog]: *SIDIblog*. Available at: http://www.sidiblog.org/2020/01/20/la-liberta-di-espressione-e-le-norme-internazionali-ed-europee-prese-sul-serio-sempre-su-casapound-c-facebook/.

Citron, D. (2014) *Hate Crimes in Cyberspace*. Cambridge: Harvard University Press.

Conti, G. (2018) 'Manifestazione del pensiero attraverso la rete e trasformazione della libertà di espressione: c'è ancora da ballare per strada?', *Rivista AIC*, 2018(4). Available at: https://www.rivistaaic.it/it/rivista/ultimi-contributi-pubblicati/gian-luca-conti/manifestazione-del-pensiero-attraverso-la-rete-e-trasformazione-della-liberta-di-espressione-c-e-ancora-da-ballare-per-strada.

DeKoven Ezrahi, S. (2003) 'Representing Auschwitz'. In: Levi, N. and Rothberg, M. (eds). *The Holocaust*. Berlin: De Gruyter, pp. 318-322.

Di Rosa, A. (2020). *Hate speech e discriminazione. Un'analisi performativa tra diritti umani e teorie della libertà.* Modena: Mucchi Editore.

Jee, C. (2021) 'Covid conspiracy theories are driving people to anti-Semitism online', *MIT Technology Review*, 13 October. Available at: https://www.technologyreview.com/2021/10/13/1037035/covid-conspiracy-theories-are-driving-people-to-antisemitism-online/.

Kaesling, K. (2018) 'Privatizing Law Enforcement in Social Networks: A Comparative Model Analysis', *Erasmus Law Review*, 3(2018), pp. 151-164. Available at doi: 10.5553/ELR.000115.

Marcus, K. L. (2015) 'Conclusion'. In: Marcus, K. L. *The Definition of Antisemitism*. Oxford: Oxford University Press.

Marone, F. (2023) 'Antisemitismo ed estremismo violento: una piaga che l'occidente non ha ancora sconfitto', *www.ispionline.it*, 27 January 2023. Available at: https://www.ispionline.it/it/pubblicazione/antisemitismo-ed-estremismo-violento-una-piaga-che-loccidente-non-ha-ancora-sconfitto-37335.

Molle, A. (2021) 'Analisi del rischio delle teorie del complotto militante', *Start Insight. Strategic Analyst and Research Team*. Available at: https://www.startinsight.eu/wp-content/uploads/2021/03/2021_molle_cospirazionismo-analisi-rischio_start-insight.pdf.

Pino, G. (2008) 'Discorso razzista e libertà di manifestazione del pensiero', *Politica del diritto*, 39(2), pp. 287-305. Available at: https://www.giorgiopino.net/uploads/1/3/1/5/131521883/pino_discorso_razzista.pdf

Pisanty, V. (2014) *L'irritante questione delle camere a gas. Logica del negazionismo*. Milano: Bompiani.

Roubache, J. (2000) 'The Council of Europe was the first to recognize the relationship between racism and antisemitism', *Justice,* 2, pp. 8-9.

Strano, M., Negre, B., Galdieri, P. (2002) *Cyberterrorismo – L'impiego delle reti telematiche da parte del terrorismo internazionale.* Milano: Jackson Libri.

Sullam, S. L. (2021) 'Antisemitismo e usi politici nel dibattito internazionale', *il Mulino*, 30.03.2021 (trad.).

Veca, I. (2019) 'Sovrani del mondo. Alle origini dell'accusa di complotto ebraico', *il Mulino*, 2.

Weimann, G., (2006) *Terror on the Internet. The New Arena, the New Challenges.* Washington: USIP Press.

Wistrich, R. S. and Vidal Sasson International Center For The Study of Antisemitism (2012) *Holocaust Denial: The Politics of Perfidy*. Berlin: De Gruyter; Jerusalem: The Hebrew University Magnes Press. Available at: https://www.loc.gov/item/2019666770/.

Ziccardi, G. (2016) *L'odio online, violenza verbale e ossessione in rete*. Milano: Raffaello Cortina Editore.

Ziccardi, G. (2023). Disciplinare le parole d'odio nella società tecnologica. In: *Parole pericolose. Conflitto e bilanciamento tra libertà e limiti. Una prospettiva trasversale. Atti del workshop tenutosi il 25 marzo 2021 presso l'Università degli Studi di Milano.* Pubblicazioni del Dipartimento di Diritto pubblico italiano e sovranazionale n. 106. Torino: Giappichelli Editore.

Chapter XVI
Fake News, Conspiracy Theories, Misinformation and Disinformation

by Paulina Kowalicka*

Index: 1. A preliminary introduction. – 2. Historical overview. – 3. Legal perspectives. – 4. Sociological and political factors. – 5. Psychological issues.

1. A preliminary introduction

In today's digitised society, we are increasingly witnessing the spread of fake news, misinformation and conspiracy theories that find widespread using new media.

The mass media have always been a functional place for rational and critical discussion, as they are able to insert themselves into the social structure and exerting an important, even manipulative, influence on the population, to be defined as the "fourth power". The media have always represented the clock of a society's history. The major mass media, in the past decades, have punctuated the flow of human history with their headlines, accompanying the daily succession of news, political or sports events. In fact, before the advent of the information society era, the communication systems used were defined as one-to-many, precisely because journalistic information, although already varied and partly biased, could rely on a consolidated infrastructure, a loyal readership, and a mutual control over the veracity of the information. The interactivity offered by the advent of the Internet transformed the communication system into the so-called many-to-many. Indeed, the supply of information has expanded not only in the audience, which is certainly broader than a few decades ago, but also in the plurality of sources that stand as guarantors of news.

The rise of digital platforms marked the beginning of a new era for the public sphere. The postmodern reality is characterised by social, economic, political and existential uncertainties and hybrid, often immaterial or invisible, threats. The fragility of the hierarchical structures of information and knowledge transmission has bequeathed a sense of mistrust, a culture of relativism and what is described as the "post-truth era": the culture of knowledge has been

* Research Fellow in Legal Informatics at the University of Milan and Managing Director of the Information Society Law Center (ISLC).

replaced by a culture of risk. Even though the digital evolution has involved users previously underrepresented by the traditional media, the system, in fact, currently supports the interests of an elite that looks primarily – and perhaps exclusively – at business.

The impact of misinformation, extremism and conspiracy theories affects data protection, privacy and human dignity on the one hand, and freedom of expression and to be informed on the other. The basis of democracy is the open public debate that requires a harmony between media freedom, the right to be properly informed and freedom of expression. The risk, which becomes more and more concrete in such an open and digitised space, is that the sheer volume of information often complicates access to reliable and verified content. The ability of each user to be both recipient and possible author of information sometimes does not allow for the juxtaposition of an effective filter and control, thus risking undermining the stability of the democratic system. This left space for phenomena that have taken over the web such as misinformation, fake news, denialism, extremism and hate speech. These phenomena rode the wave through various digital platforms that took over the new market generated by digitalization.

1.1. Definitions

Fake news is defined as "false stories that appear to be news, spread on the Internet or using other media, usually created to influence political views or as a joke".

The terms misinformation and disinformation are used to refer respectively "dissemination of false or wrong content that does not presuppose any malicious intent" and "purposely created false/misleading information that is meant to manipulate or harm people, places and institutions, typically in pursuit of political, social or financial agendas". Thus, a significant difference is evident, emphasising the intentionality of the action and the awareness of the falsity of the information disseminated.

Conspiracy theories represent a particular form of disinformation, aimed at conveying a reconstruction of reality in which complex political and social events revolve around a secret organisation, alternative to the institutions, that would control, or aim to control, the world order. "A belief that an event or situation is the result of a secret plan made by powerful people".

2. Historical overview

Misinformation and conspiracy theories have very ancient origins. The earliest evidence of conspiracy theories dates to the Peloponnesian War; Thucydides reports how the Athenians attributed the spread of the plague, which struck

the city to the invaders, who were accused of polluting the cisterns of Piraeus with poisons. But this is not the only ancient evidence of conspiracy theories. One of the best known is the one that accompanied the great fire of Rome in 64 AD. At that time, in Rome, a feeling of mistrust was erupting into outright religious hatred, towards Christians. It was precisely this climate of mistrust and suspicion of a religious minority, that premised the theory that it was Christians who had set the city on fire, a theory that historians attribute to the emperor Nero himself, who blamed the devastation of the city on the Christians.

Misinformation campaigns can also be traced back to ancient times. Consider again of the Peloponnesian War and the letter Pausanias allegedly sent to Xerxes, King of the Persians, offering to betray his part. As Thucydides reported, the letter turned out to be false, but the misinformation campaign was successful: Pausanias was sentenced to death, and in fact he starved to death in the temple of Athena where he took refuge to escape execution of the sentence.

The phenomena of misinformation and conspiracy theories crossed the centuries, from the French Revolution to the Covid-19 health emergency, through the early 20th century with the Protocol of the Elders of Zion, to aliens and the Kennedy assassination.

Conspiracy theories seem to have in common a narrative that places at the centre of its reconstruction of reality the presence of a secret and powerful elite, which through to an intricate project aimed at harming humanity, with the purpose of maintaining or attaining power, would be able to control, or even generate, harmful phenomena (such as, for instance, the coronavirus) and by deception make the world population believe that such events have a natural origin.

3. Legal perspectives

The relationship between law, misinformation and conspiracy theories can be complex and multifaceted. While law is a normative system that seeks to regulate society and ensure order, conspiracy theories are concepts that claim that certain events are the result of secret or malevolent conspiracies. Scientific doctrine, in the context of law, is based on rational, evidence-based analysis. Whereas conspiracy theories often lack a solid empirical basis and are supported by speculation, distorted interpretations of facts or unverifiable testimony, scientific doctrine focuses on the objective evaluation of evidence and the search for truth.

The relationship between law, misinformation, extreme content and conspiracy theories is complex and requires a balance between the right to freedom of expression and the protection of the public from misleading or harmful behaviour. The law can intervene when such content leads to unlawful behaviour, but, in general, countering such phenomena requires a multifactorial approach

involving education, accurate information and open dialogue. Scientific doctrine plays a crucial role in the evaluation of evidence in the legal context and can help counter misinformation and conspiracy theories by providing a rational, evidence-centred basis for evaluating claims.

3.1. Freedom of expression

The legal framework of Western democracies is particularly sensitive to the phenomenon of misinformation, propaganda, extremism and conspiracy theories. The European constitutional and regulatory framework was structured in a context that was totally different from today's landscape, at a time when technological development and available digital tools were still a long way off and requirements were far removed from the difficulties that citizens must face today. The European approach, based on the American matrix, protects the free formation of personal opinions, which are confronted in a "free market" of ideas (Ziccardi, 2019), giving ample space for the free circulation of thoughts and content, with the only limitation being respect for the fundamental rights of the individual, to guarantee a healthy information environment free of discrimination. It is precisely this freedom that makes Western democracies an attractive target and incites instances of misinformation and propaganda interfering with democracy by altering public debate.

European law provides for a rather limited and deficient strategy regarding the approach to misinformation and fake news due to the temporal aspect of the phenomenon, which often has a reasonably short life cycle. In individual legal systems, it is possible to find rules that protect misinformation across the board, in the case, for example, of the principle of direct damage, in the case of false, overstated or tendentious news that is liable to disturb public order or cause alarm. Or in the case of the offence of defamation, specifically in the form of direct damage to reputation, where the injury caused is of immediate perception. This is conduct that can be carried out by any person and in different ways, with many European legal systems providing for an aggravated form if the offence is conveyed online.

Sometimes the phenomenon of fake news concerns a broader and more complex project, also contributing in the long term to a substantial change in the perception and opinion of the victim-subjects. One only must think of conspiracy theories, cases in which fake news is used as a strategy to manipulate groups of people with the aim of polarising an ideology, which often has no foundation. From this perspective, the phenomenon could perhaps be framed as a true hybrid threat and as part of the "information war", thus providing a more consolidated frame of reference and filling the gaps (Suffia, 2022). The relationship between freedom of expression, misinformation and conspiracy theories is a complex issue that raises several legal and ethical questions. Freedom of expression in the European Union is a fundamental principle enshrined in

the Charter of Fundamental Rights of the European Union. Article 11 of the Charter states that "everyone has the right to freedom of expression", which means that everyone has the right to express their opinions, ideas and beliefs without unjustified interference by public authorities.

The right of expression in the EU covers a wide range of forms of communication, including freedom of speech, freedom of information, freedom of the press and freedom of the media. This right is one of the main pillars of democracy and European values and is enshrined in the most evolved mindsets as the combination of "freedom of expression" and "democracy", as it enshrines the right of citizens to express their opinions and ideas without government interference or undue restrictions. It is considered essential to foster open and pluralistic public debate, allow democratic participation and promote diversity of opinion. The main challenge in the context of the right of expression is to strike a balance between protecting this fundamental right and dealing with the potential negative consequences of misinformation and conspiracy theories: while it is important to guarantee freedom of expression, there are situations where conspiracy theories may cause tangible harm or raise concerns for public safety and the well-being of society.

Freedom of expression is not an absolute right: the Court of Justice of the European Union and the European Court of Human Rights played a significant role in defining the boundaries of the right of expression in the EU, developing jurisprudence that balances the right of expression with other legitimate needs, ensuring that restrictions are applied in a manner consistent with democratic principles and human rights. Striking a balance requires a considered approach that considers fundamental rights, scientific evidence, education and the public interest.

The development of a new digital dimension and the spread of new channels of communication led to the rise of a true technological identity. Users, the holders of what is known as digital citizenship, are the bearers of new rights, ranging from participation, information and interaction to new duties, designed to guarantee the security of users themselves, the protection of their data and the set of principles and values underpinning the main human rights Charters. In 2015, the new text of the Declaration of Rights on the Internet was drafted by the Commission on Internet Rights and Duties. The Charter represents the tool for building citizenship in the age of the Internet of Things and it is a fundamental instrument since without citizenship there is no democracy. It was the object of numerous consultations, and its main goal is to rebalance rights by laying the foundations of an ever-changing reality, recognising the freedom, equality, dignity and diversity of each person, promoting innovation, growth and fair competition within a global context. Article 1 of the Declaration guarantees the fundamental rights of every person and one of the passages, that created not a few controversies within the digital space, is precisely that relating to

freedom of expression, provided, as we have seen, in Article 11 of the Charter of Fundamental Rights of the European Union and Article 10 of the ECHR.

The Internet is seen as a borderless reality that allows for the existence of an ideally more liberal and free society, and it is from this perspective that the greatest effort must be made, so that the full exercise of all rights can be guaranteed. Freedom of speech stems from the need to allow everyone to express their ideas without any pre-conceptual veto. This fundamental claim, which goes well with democratic ideals, is declined, on the web, in an almost exponential way since all social profiles, by definition, are created at the same level of notoriety and potentially have the same media resonance. The problems generated by a society with these premises are many: starting from the consideration of a new world in which the weight of words loses its meaning, since the statements of a leader of a State can potentially have the same audience and resonance as those of an ordinary person. This starting point begins to blur the contours of a far more multifaceted right of expression in the digital sphere, which inevitably takes on negative implications, impossible to foresee initially.

Another problem that arises and on which it is inevitable to dwell is the role of users, who hold a dual position: on the one hand, they are the recipients of an excessive amount of information, on the other hand, they may represent the source itself. It is inevitable that this gives rise to an unstoppable flow of content and information accessible to all, which, while representing one of the secrets of the success of these platforms, is also one of its greatest flaws.

However, the Internet is not a paradise for freedom of expression worldwide.

In Europe, the digital platforms are not subject to government control, therefore, it is the companies, owners of the content and data flows, that hold the power. Here, the freedom of speech, the right of access and the protection of anonymity provided for by the Declaration are fully guaranteed, allowing subjects to express their personalities within society. The social phenomenon had the merit of connecting the entire world and shortening distances by making them just a click away, but this sudden change has brought different cultures and traditions of peoples and States that have found themselves living together within the same platform. The difficulty lies, precisely, in accommodating these diversities by allowing a more gradual adaptation of different cultures to a world without borders, which undoubtedly represents the future of humanity.

In a world where the makers of news are also their own users and vice versa, the users are at the mercy of billions of pieces of information at any given moment. This enormous mass of news, that invades and pervades everyday life, leaves no space for careful evaluation or contradiction and, on the one hand, this is all instrumental in providing information that is often hasty and incomplete information, while on the other, it generates a feeling of rejection and apathy in the web population. Although the welter of content makes one think of more varied and complete information, it often has the opposite

effect, generating misinformation and distrust in the information media. This mechanism has created important consequences in people's daily lives, in the political scenario and in national security. Although at first glance they seem very distant, in a world in which we are all connected, the consciousness of an entire community – perhaps – resides on the web.

The problem of the right of expression shows the various governments approach the new difficulties, often trudging along and leaving space for the Tech companies, which do not struggle to fill the void left by legislative norms, thus becoming the true guarantors of the rules of the new world. Humanity, therefore, finds itself in a situation where two opposing forces try to take control of a new reality in which people are looking for references and new rules to follow and rely on. The key word is "balance". The balance that needs to be found between institutions and companies that allow for regulation that protects users and encourages, at the same time, the economy, offering everyone certainties and eliminating pretexts.

3.2. The relationship between misinformation and Artificial Intelligence

The rapid development and wide deployment of artificial intelligence systems, in particular Large Generative AI Models (LGAIM), are revolutionising several areas of society.

Such systems may offer enormous potential, contributing to a more efficient and potentially fairer allocation of resources, but they also carry significant risks. Due to their design and capabilities, LGAIMs (*e.g.*, ChatGPT, Stable Diffusion, Synthesia, MusicaLM) (Hacker, Engel & Mauer, 2023) and artificial intelligence chatbots can engage in the large-scale dissemination of false narratives, leading unsuspecting individuals to believe and share inaccurate information. In recent years, end-to-end neural conversational agents have made remarkable progress in interacting with humans, but being trained on huge datasets from the Internet, there is a tangible risk that they will learn and be deployed based on toxic behaviour or extreme and harmful stances. Misinformation has the potential to undermine trust, exacerbate social and political divisions and – even – incite acts of violence. From a legal and regulatory perspective, it is of considerable interest to study possible policy proposals for content moderation in LGAIM, given their versatility and wide range of applications. It is pragmatically accepted that the models used for regulation should be as flexible as possible to allow greater freedom and innovation for users. Some believe that AI can solve the problems it creates by using automatic content filtering systems (through a combination of AI tools, developer and user interventions and a set of rules) that can identify harmful content and provide a means for effective self-regulation and co-regulation by platforms. However, AI algorithms may not be the only way to regulate consent in the future, as their accuracy is limited, especially for expressions where cultural or contextual input is required. Others argue

that the answer to misinformation problems on LGAIM can be found through the implementation of transparency obligations and specific mechanisms that could be made mandatory by regulators, such as trust flags indicating potentially problematic content according to different categories (hate, violence, gender, etc.) and/or external audits and evaluations. The debate focuses on identifying the most appropriate legislation to directly address the risks posed by these technologies. Some (Hacker, 2023) believe that the regulation of LGAIM risks should generally focus on the applications of these technologies, rather than the preformed model, through technology-neutral laws designed to address critical issues more effectively and focused on three levels of obligations. Firstly, the analysis of minimum standards for all LGAIMs is noted, secondly, high-risk obligations are defined for high-risk use cases and, finally, collaboration along the AI value chain is envisaged, including obligations on transparency, risk management and content moderation rules (Hacker, Engel & Mauer, 2023). According to this guideline, specific legislation risks becoming obsolete even before, or soon after, its implementation, given the changing nature of these technologies (Hacker, Engel & Mauer, 2023). Regulators and legislators need to act quickly to keep up with the unconstrained dynamics of AI models, updating regulation and striking the right balance between promoting innovation and safeguarding against potential harm, thus creating a level playing field for the development and deployment of future AI models within European borders and overseas. Given these premises, it is evident how the limitations of these technologies in recognising the digital pollution of the infosphere are emerging and, through an important renewal and adaptation activity, it is necessary to define a solid and updated regulatory framework.

The year 2023 is the year in which legislators were able to agree on a vision and set a strategy, 2024 will be, probably, the year in which policies will start to turn into concrete actions. It is fundamental for law to establish definitions, procedures and results that can be comprehensible and acceptable for the integrity of the system. It is the task of law to provide the principles based on which laws and regulations can be adopted. The United States, China and Europe, albeit with different approaches and strategies, are the main players on the global scene dominating the race for technological supremacy. The United States was the first country to implement a strategic plan for research and development of AI technologies, with the intention of minimising government interference. China has set itself the goal of becoming a world leader by 2030, while keeping political control and decision-making power in the hands of the State. Finally, the European Union, although from a research and development point of view it is at a distinct disadvantage, represents a model of guarantee aimed at ensuring that the development of AI technologies does not harm fundamental rights, democracy, the rule of law and environmental sustainability.

The United States is attempting to promote and develop regulation that demands greater transparency and new standards in each area in which AI plays an important role. The executive order issued by President Biden envisages a series of initiatives and procedures aimed at building a system that, on the one hand, allows for the safe and controlled development of such technologies and, on the other, protects citizens by assessing and mitigating risks in the use of AI. The next few months will be a great challenge for the American landscape: in fact, the presidential elections are expected to influence much of the political debate on the use of generative artificial intelligence, the management of misinformation and the dissemination of polarised content on social media platforms, and the regulation and prevention of the harms and risks that such technologies can generate, globally.

Chinese regulation was initially fragmented, having preferred ad hoc legislative acts for algorithmic recommendation services, deepfake and generative artificial intelligence. The Chinese effort, in recent months, is turning to the definition of a unified regulatory framework aimed at guaranteeing an efficient algorithm training phase, a careful selection of data sets and respect for privacy and intellectual property. Whilst it seems to be in line with the European approach, the drive to control information and a surrender of civil rights to social control by the authority, following the model of Chinese socialism, is strongly evident. China has been very responsive to the development of new technologies; it was, in fact, the first Country in the world to introduce legislation on generative artificial intelligence a few months after the ChatGPT breakthrough.

The European Union is working on several fronts to try to effectively regulate artificial intelligence systems with specific instruments (Artificial Intelligence Act, Artificial Intelligence Directive, Product Liability Directive, Digital Services Act and Digital Markets Act).

4. Sociological and political factors

The spread of fake news and misinformation in the political context could be the harbinger of particularly damaging consequences for democracy, it is therefore possible to understand the great danger that misinformation campaigns and conspiracy theories can represent for liberal and democratic systems and, consequently, on the social level.

We are therefore witnessing the use of fake news and conspiracy theories, within political dynamics, as a weapon capable of polarising discourse and negatively influencing the electorate, often by appealing to emotional states and attacking the foundations of pluralism and democracy. For these reasons, addressing conspiracy theories also requires an understanding of the social and psychological factors that support them, as well as educational and communication efforts to promote accurate information and rationality.

Although the factual basis of such narratives is often quite questionable and irrational, it is often not possible to directly refute such beliefs, as it is complicated to find evidence that something did not happen.

It is, therefore, crucial to develop a sociological understanding of the structural conditions associated with conspiracy narratives, focusing on the socio-psychological processes and social conditions that underlie such beliefs. Indeed, social conditions associated with insecurity or threat in the sociological and political environment, unemployment, political changes and demographic shifts are associated with an increase in conspiracy ideation (DiGrazia, 2017).

From a sociological perspective, the potential social causes of conspiratorial beliefs are examined by considering the role of perceptions and interpretations individuals have of their social environment leading to polarised, conspiratorial beliefs. In this way, complex issues and concerns are projected as the fault of a single individual or group to whom they attribute immense power. This, in turn, allows them to believe they can understand the source of the threat (DiGrazia, 2017) and focus the blame on a specific group or individual.

Another key aspect of believing in conspiracy theories and misinformation is the level of education. For example, people with a university degree were less likely to believe COVID-19 misinformation and more likely to trust preventive measures than those without a university degree and may lead individuals with inadequate levels of scientific and digital literacy to proliferate and spread scientific misinformation.

This indicates that education, and in particular digital literacy, plays a decisive role in the proliferation of misinformation and conspiracy theories. The ability to access and critically evaluate information is essential for democracy and active citizenship and, in a digitised society, there is a strong need for education.

The use of technology, levels of education and perceptions of personal identity are therefore key to understanding the likelihood and vulnerability of believing in conspiracy theories, which can have potentially deleterious social and political consequences. Digitisation and globalisation have enabled an unprecedented spread of narratives and counter-narratives on a wide range of topics. The Internet, and social media in particular, play a decisive role in gathering momentum and facilitating the dissemination and sharing of unfounded and conspiratorial content.

In these terms, it seems that future events may highlight the same dynamics revealed during the Covid-19 pandemic: the event or crisis erupts, unfounded and uninformed content begins to circulate, new conspiratorial narratives are constructed and spread, and social stratification, conflict and polarisation of opinions become mainstream.

The work of governments and institutions is therefore crucial to overcome the proliferation and reasons why conspiracy theories and misinformation gain so much space online and within communities. Using empirical evidence,

governments and institutions can develop effective policies and operate in a prudent and transparent manner. They can stimulate more educational policies for all demographics and provide accurate and verified information through official channels to encourage the dissemination of scientifically based and demonstrable knowledge and counter-narratives.

5. Psychological issues

The relationship between psychology, disinformation, extremism and conspiracy theories encompasses a wide range of aspects and complexities, and several psychological factors may contribute to the adherence to these ideologies. Indeed, misinformation content and conspiracy theories often offer a simple narrative that can generate an apparent sense of control.

The socio-cognitive factors associated with conspiracy thinking are important, and the use of infodemics, religious fatalism, and political distrust combined with distrust in mainstream media can negatively affect an individual's psychological well-being by amplifying the likelihood of emotional and non-rational reasoning, and a diminished perception of socio-political control by fueling social polarisation and disorder and undermining trust in democratic institutions.

Generally, the explanations from which conspiracy theories are drawn are not supported and accompanied by scientific evidence. Such theories often emerge in times of crisis, when people seek simple and conclusive explanations for complex events, offering an alternative narrative that seems to make sense of events and identify those responsible. Conspiracy thinking can be based on extreme skepticism, leading to the rejection of notions not in line with established beliefs, forcing an interpretation of facts that supports this worldview, without verifying or questioning the choices made (Lewandowsky *et al.*, 2022).

Psychology attempts to explain conspiracy theories and the reasons for their widespread prevalence in society by investigating possible combinations of motivations and personality traits. Indeed, belief in conspiracy theories may be related to certain psychological traits or conditions that may influence the tendency to believe in these narratives (Van Prooijen, 2022). The profile of the conspiracy theorist reflects a logic of thought, defined by Popper, as self-immunising: "I believe in something, I cannot find the proof, here is the proof that it is true, and they want to hide it" (Popper, 1945). These people may present dysfunctional or problematic personalities and delusional-like modes of reasoning, in which a belief, once ingrained, is never reasoned out even in the face of logical and objective evidence of facts. But belief in conspiracy can also be a defence mechanism. When an individual is faced with a feeling of fear, he may look for ways to control the unknown, attributing an identity to it or attempting to identify and neutralise it. He may try to protect himself from the effects of

chance, natural events, epidemics and so on, to manage and negate the feeling of fear. From this analysis, it can be assumed that the belief in conspiracy is largely a dysfunctional way of being, on the one hand accompanied by a distorted lifestyle and, on the other, by a feeling of superiority and omnipotence, oriented primarily towards deceiving oneself and others, which makes it a risk to the community as it may be able to influence those who share the same narrative or vision, to make dangerous choices. People may also be inclined to believe in conspiracy theories due to a combination of their personality traits and their motivations which may include financial or political gain, or a desire to feel valued or special to satisfy their motivational needs.

All people have potentials and limitations when it comes to reasoning, and Suzuki and Yamamoto, in 2020, found that individuals often tend to reach conclusions without thorough analysis, perform many everyday actions without thinking, create beliefs based on stereotypes, and make decisions according to cognitive biases, which are constructs derived from misperceptions, from which judgements, prejudices and ideologies are inferred; moreover, individuals tend to make their decisions using a limited number of heuristics (an unconscious, intuitive and hasty mental process that leads to a quick and simple conclusion with minimal effort) rather than sophisticated rational processes. Whilst this approach may work well in many areas of human life, there is a risk of producing systematic biases in judgement leading to flawed decision-making. Biases, in fact, are forms of ineffective heuristics that alter the perception of events and, as they are not generated on real data, are acquired a priori without criticism or foundational support.

During the Covid-19 pandemic, scholars focused on social media exposure, limited supervision, control and regulation and correlations between the use of social media or particular media sources and the tendency towards certain types of behavior or beliefs, including support for political violence and criminal behavior, established a causal relationship between exposure, which is likely to promote beliefs, and promoting and incentivizing deleterious behaviors (Uscinski *et al.*, 2022).

Algorithms have less influence than one might a priori assume, assuming, rather, a more relevant role in the subsequent phase of the polarization of ideology, rather than in the initial reception of content. People, in fact, tend to select information that coincides with their own identity and ideologies and reject counter-narratives that may be contrary or that may allow for critical thinking.

In general, conspiracy theories seem to provide broad and coherent explanations, allowing individuals to preserve beliefs in the face of uncertainty and contradictions. They have an impact on individual's psychophysical health, social behaviour, socio-economic and political factors and, in general, the well-being of society. The ways in which human beings make decisions and act are

strongly influenced by their view of reality and their epistemic, existential and social needs.

Many conspiracy theories provide alternative explanations for high-impact events in the world, such as pandemics, wars, environmental disasters; in this perspective, relying on unfounded but less worrisome content and such theories, therefore, can help the individual feel gratified by such "discovery". Studies show that people are more likely to become radicalised because of situations of injustice or humiliation (Kruglanski *et al.*, 2014), associating such beliefs, at first glance, with a positive perception, capable, in reality, of defending only relatively fragile forms of self-perception.

However, it is important to note that such benefits are temporary in nature, suggesting that conspiracy theories provide people with a form of instant gratification (Van Prooijen, 2022).

The picture that emerges is, therefore, very complex. There are countless factors that lead individuals to believe and rely on conspiracy theories. They include personality, social factors related to behavior, or purely random factors linked to information exposure, which is often chaotic and overly broad. These are cognitive factors and factors that point to the functional role of such beliefs, which are probably self-protective in nature.

Therefore, to reduce the influence of such theories and foster an informed and cohesive society, the promotion of media literacy and critical education, the use of evidence-based communication strategies, greater community involvement, strengthening a sense of belonging and social support, and investment in research and scientific dissemination are essential.

It is only when uninformative, polarising, misleading arguments manage to be first identified and then rejected, that the "Holy Grail of deliberative, inclusive, transparent and accountable decision-making" can be achieved (Lewandowsky *et al.*, 2022).

Bibliography

DiGrazia, J. (2017) 'The Social Determinants of Conspiratorial Ideation', *Socius: Sociological Research for a Dynamic World.* Available at: https://doi.org/10.1177/2378023116689791.

Hacker, P. (2023) 'The European AI Liability Directive – Critique of a Half-Hearted Approach and Lessons for the Future', *Computer Law & Security Review*, 51. Available at: https://doi.org/10.1016/j.clsr.2023.105871.

Hacker, P., Engel, A. and Mauer, M. (2023) Regulating ChatGPT and Other Large Generative AI Models. *Oxford Business Law Blog.* Available at: https://doi.org/10.1145/3593013.3594067.

Kruglanski, A., et. al. (2014) 'The psychology of radicalization and deradicalization: how significance quest impacts violent extremism', *Political Psychology*, 35. Available at: https://onlinelibrary.wiley.com/doi/full/10.1111/pops.12163

Lewandowsky, S., Armaos, K., and Cook J. (2022) 'When Science Becomes Embroiled in Conflict: Recognizing the Public's Need for Debate while Cobating Cospiracies and Misinformation', *American Academy of Political and Social Science*, 700(1). Available at: https://doi.org/10.1177/00027162221084663

Popper, K. (1945) *The Open Society and its Enemies*. London: Routledge.

Suffia, G. (2023) *Pulire l'infosfera. Intelligenza artificiale e contrasto alla disinformazione*. Milano: Giuffrè Francis Lefebvre.

Uscinski, J., et. al.-(2022) 'Cause and effect: On the antecedents and consequences of conspiracy theory beliefs', *Current Opinion in Psychology*, 47(101364). Available at: 10.1016/j.copsyc.2022.101364.

Van Prooijen, J.-W. (2022) 'Psychological benefits of believing conspiracy theories', *Current Opinion in Psychology*, 47(101352). Available at: https://doi.org/10.1016/j.copsyc.2022.101352

Ziccardi, G. (2019). *Tecnologie per il potere. Come usare i social network in politica*. Milano: Raffaello Cortina.

Chapter XVII
Protection of Minors Online

by Samanta Stanco*

Index: 1. Introduction. – 2. Minors' data protection: the provisions of the GDPR. – 3. The online exposure of minors. – 4. The debate on access barriers to online platforms. – 5. Some of the main risks for children in the digital society. – 6. An extra-EU perspective. – 7. Final considerations.

1. Introduction

The digital world is constantly evolving and expanding: from a simple recreational and information space, it has become a place where the individual develops and expresses his or her personality, interfacing with many users and experimenting with new forms of social interaction.

Such a transformation has also brought a series of risks associated with the misuse, or careless use, of technological devices: on the one hand, in fact, traditional crimes have found, in the virtual world, the ideal terrain in which to proliferate, overcoming the classic temporal and spatial limits and counting on digital inexperience on the part of many users; on the other hand, new forms of persecution and aggression, facilitated by the pervasiveness of technologies, have also emerged and new addictions have developed, due to the obsessive and intensive use of the electronic tool.

If this is true with reference to the world of adults, it is even more so when the focus is shifted to minors: in fact, these are subjects who, precisely by virtue of the age group to which they belong, are deserving of greater protection, since – often – they are less aware of the risks they could face in their digital experience and the rights that can be enforced there, to protect their interests and their data.

On several occasions, the importance of providing suitable guarantees to safeguard the proper development of the identity and personality of minors online has been reiterated, giving rise to an integrated system of protection that includes acting on three distinct levels:

1. technological, through the design of tools (including control) that are increasingly evolved and efficient;

* Lawyer. Research fellow in Legal Informatics at the Faculty of Law at the University of Milan and Principal investigator of the Information Society Law Center (ISLC).

2. legal, with the provision of a regulatory framework aimed at preventing the commission of certain conducts, on the one hand, and able to ensure a timely intervention, on the other;
3. educational, through the collaboration of schools, families and all the stakeholders in various capacities involved in the process of growth of the child, called upon to play a role of care and teaching towards the same.

2. Minors' data protection: the provisions of the GDPR

The amount of data shared online by individuals of minor age, consciously or not, is such that it requires the presence of a specific discipline to protect their rights and ensure their effective protection. Regulation (EU) 2016/679 (GDPR) intervenes in this regard with specific provisions.

Article 6, dedicated to the principle of lawfulness, qualifies as legitimate a processing that does not prejudice the fundamental rights and freedoms of the data subject, "in particular if the data subject is a child". In fact, minors are deemed deserving of special protection regarding the processing of their personal data, as they are less aware of the risks and rights associated with it.

Taking a central role in the protection of children's data, however, is Article 8 GDPR, headed "Conditions applicable to child's consent in relation to information society services".

According to this provision, the direct offer of digital services (such as signing up for a social network or messaging services) is permitted only where the minor is at least 16 years old; otherwise, the processing will be considered lawful only where consent is given or authorized by the holder of parental responsibility.

In the Italian legal system, therefore, there is an exception to the general rule in Article 2 of the Civil Code, according to which capacity to act is acquired upon reaching the age of majority, *i.e.*, at the age of 18.

Art. 8 GDPR, provides, then, that the age limit may be further decreased by member States, but without falling below 13 years of age. The Italian Legislature, specifically, set this threshold at the age of 14.

Finally, the provision requires the Data Controller to make concrete efforts, based on available technologies, to verify that consent is actually given or authorized by the person exercising parental authority over the child.

Corollary to the provision just reviewed is, then, the "right to be forgotten" under Article 17 GDPR, which allows the data subject to obtain from the Data Controller the erasure of personal data concerning him or her without undue delay. Again, as also specified by Recital No. 65, this protection must be understood in a strengthened manner in the case where consent has been given by a minor, especially in view of the repercussions that certain content, shared at a young age, may have on the future reputation of the same.

3. The online exposure of minors

An important debate has concerned the publication of images of children and young people online, including in relation to the activity of parents and relatives on social networks.

Although the right to image was expressly recognized well before the advent of the Internet and social media, the changing technological panorama has drawn attention to the point, underscoring how any infringement of this right is likely to create harm to the honour, reputation or decorum of the individual, especially if he or she is a minor.

In recent times, the phenomenon of so-called "sharenting" has found popularity: the term, which originates from the union of the words "share" and "parenting", denotes the constant online sharing, by parents, of content concerning their children (information, photos, videos and, even, ultrasound scans).

This practice, in addition to representing an obvious interference in the privacy of minors, also exposes them to additional risks related to the presence of ill-intentioned people in the digital world: think of a teenager embarrassed in front of friends by private information revealed on social networks by parents, suitable to make him a victim of bullying episodes; or, again, think of photos of children intercepted by pedophiles and groomers and used to feed the market of child pornography material.

This is why the Italian Data Protection Authority has dictated, on this point, a series of "best practices" aimed at stimulating a greater level of caution in the online sharing of data about minors, to ensure the proper and healthy formation of their personalities.

4. The debate on access barriers to online platforms

A growing issue when it comes to the presence of minors online is that of barriers to accessing platforms.

Bringing attention to the point was a dramatic incident that occurred in January 2021, when a girl only 10 years old lost her life while participating in a challenge that went viral on the social networking site TikTok.

Online challenges are situations that require participants to engage in highly risky behaviors that, however, are not perceived as such by the individuals to whom they are directed.

In fact, to be involved are children who feel somehow "invincible" and have no real perception of the possible consequences arising from such behaviors.

Some of them involve risks from a strictly medical point of view, requiring the use of soap, medicine or commonly used substances such as cinnamon, salt and baking soda. A prime example is the challenge known as "Eyeballing",

which consists of pouring vodka directly into the eyeball as if it were eye drops, so as to accelerate and amplify the narcotic effects of alcohol.

Other challenges, on the other hand, induce actions that are likely to cause serious injury, either to the person performing them or to others. This is the case, for example, of the so-called "extreme selfie", which induces to take a selfie in highly risky situations or places (train tracks, the roof of a skyscraper, etc.) and, then, share the image on social networks; or, again, of the "Blackout challenge", consisting of a competition in which young people measure their resistance to the lack of oxygen, tying a rope, belt or noose around their necks and stopping only at the moment they arrive at unconsciousness.

The latter was precisely the challenge that proved fatal for the 10-year-old girl.

As a result of the affair, the Italian Data Protection Authority decided to intervene as a matter of urgency and ordered against TikTok the immediate blocking of the use of data of all those users whose age had not been ascertained with certainty.

In addition, the dramatic episode has opened the debate on the issue of barriers to platform access.

Precisely with reference to the issue of the enrolment of underage individuals in socials, several proposals were made, some of them related to the use of SPID, the authentication system that, in Italy, allows citizens to access online services of the Public Administration entities with a single digital identity. However, none of the proposed solutions were actually implemented.

Also, AGCOM, in Resolution 9/23, adopted specific Guidelines, aimed at guiding operators regarding the implementation and configuration of child protection systems and the provision of clear and transparent information on how electronic communication service contract holders should use them.

5. Some of the main risks for children in the digital society

As anticipated, indiscriminate sharing of data about minors and unconscious surfing in the digital world can lead to the occurrence of dangerous phenomena.

These include, first, cyberbullying, which is an aggressive and intentional behaviour realized by an individual or a group of individuals, through the electronic medium, against a victim who has difficulty defending himself or herself. The forms in which cyberbullying can manifest itself are the most varied: harassment, threats, insults, pressure, blackmail, insults, denigration, identity theft, unlawful processing of personal data, and so on.

In Italy, the phenomenon was first regulated by Law No. 71 of May 29, 2017, which, for the first time, introduced specific protection on the point, providing a regulatory intervention with a distinctly re-educational, rather than punitive, character.

This provision, in Article 2, provides for a special procedure that allows the child over the age of 14 or the exercisers of parental responsibility to apply to the operator of the website or social network to obscure, remove or block the child's personal data illegally disseminated online and perceived as harmful. If removal does not occur within 48 hours, it is possible to file a petition directly with the Data Protection Authority.

Also increasingly common are episodes of online grooming. The term indicates methods of psychological manipulation adopted to select, involve and maintain in a situation of sexual abuse and exploitation victims of minor age (Salter, 1995).

The main characteristic of grooming is, therefore, the construction of a conditioning communicative relationship through which the adult seduces and manipulates the minor and extracts his or her consent to induce him or her to take part in sexual abuse or exploitation.

In general, the groomer uses a standardized procedure: the process begins with establishing contact with the victim through social networks, chat-rooms or gaming platforms, posing as a peer and engaging in conversations on topics of common interest; second, the groomer aims to gain the trust of the minor and, to this end, employs manipulative techniques to identify the minor's weaknesses (such as school problems, family tensions or love disappointments) and, progressively, assumes the role of confidant and point of reference; finally, the groomer takes the conversations to an increasingly personal level, gradually leading them to sexual content.

The Italian Legislature, by ratifying the Lanzarote Convention, introduced Article 609-undecies of the Criminal Code. The provision punishes anyone who, for the purpose of committing one or more crimes related to the sexual exploitation or abuse of minors, solicits a minor under the age of 16, unless the act constitutes a more serious crime.

And again, a phenomenon that is reaching increasingly important dimensions is sexting, which is the practice consisting of creating and sharing texts, images or videos with sexually explicit content.

It should be made clear that sexting is not an illegal practice if it is done consensually and between individuals who have reached sexual maturity. Problems arise when the material thus produced is used for illicit purposes, as happens in the case of revenge porn.

In revenge porn, in fact, there is the dissemination of intimate, or sexually explicit, material for the purpose of revenge at the end of a relationship, to subjects who were not the recipients of that material and without the consent of the person portrayed. The crime fits into the broader framework of non-consensual pornography and is a source of great concern since, although there are specific laws on the point, it is always a matter of damage mitigation

maneuvers, which, however, do not completely eliminate the consequences of such insidious conduct.

In Italy, revenge porn has been regulated and framed as a crime by Article 612-*ter* of the Criminal Code, which punishes the conduct of anyone who, after making, taking, acquiring or receiving, "sends, delivers, gives, publishes or disseminates" images or videos with sexually explicit content, intended to remain private, without the consent of the person portrayed. The rule, therefore, is aimed at punishing not only the conduct of the person who, after making the image with sexually explicit content, first disseminates it without the consent of the person portrayed, but also that of the so-called "second distributors", *i.e.*, those who received such images from others or downloaded them from the network.

The world of video games has also undergone radical change in recent years, with reference to both the technologies used and the ways in which games are played (Ziccardi, 2023).

This transformation has, unfortunately, proved capable of generating new criminal possibilities, suitable for endangering the activities of minors.

Gaming experiences, in fact, are no longer relegated to the use of a device disconnected from the network or participation on closed platforms, but have shifted to large digital spaces in which it is possible to meet new people, talk to unknown players and create bonds with other users.

We find ourselves, thus, in the presence of an environment that fosters the bad intentions of criminals and groomers in their goal of finding young victims, who end up being targets of bullying conduct or sexual abuse.

6. An extra-EU perspective

In 1998, in the United States of America, a law passed with the express purpose of protecting minors from contact with material of a sexual nature found on the Internet: This was the Child Online Protection Act (COPA), which targeted the Internet transmission of harmful material to minors distributed for commercial purposes.

"Harmful material", according to the law, was defined as pictures, images, graphic image files, articles, recordings, writings, or other communications that, according to contemporary community standards, were judged to have indecent content, prurient interest or showing sexual acts or naked bodies (including female breasts).

Unfortunately, the law was blocked in later years, between 1998 and 1999, as it was found unconstitutional, since it limited the First Amendment rights of adults in the process of protecting minors and the reference to "contemporary community standards" was deemed too vague.

Since 2000, however, another federal law, known as the "Children's Online Privacy Protection Act" (COPPA), has been in effect. The law was passed in response to online marketing techniques that found widespread use in the 1990s to strengthen the protection of minors' privacy.

COPPA applies to every website that collects data from children under the age of 13, including all social media platforms, and imposes specific requirements on operators of websites and online services to protect the privacy of these minors.

Specifically, the law requires that the collection of personal data be preceded by the publication of a privacy policy, also specifying what the minimum content of the policy should be.

It also requires website operators to seek verifiable parental consent for the collection or use of any personal information from young website users, specifying their responsibilities from a legal point of view.

Although COPPA does not explicitly define how parental consent should be gained, the Federal Trade Commission has established guidelines to help website operators. In general, parents must have full access to all user records, profiles and login information and may delete their child's personal information.

7. Final considerations

As seen, while access to the digital world represents a source of great opportunities for children, it can also pose significant risks.

To minimize these risks as much as possible and enable the creation of a healthy and safe digital environment, certain good practices must therefore be followed, both by the children and by those exercising parental responsibility.

Parents have a right-duty of control over their minor children, which requires them to give appropriate instruction and education, verifying the correct reception of the educational message, including in relation to the use of technological tools.

Of course, a proper balance must be struck between the parents' right-duty to control and the minors' right to privacy and confidentiality.

In fact, the control of the children's lives cannot be considered unlimited and, therefore, the duty of supervision cannot justify any intrusion into the private sphere of the child itself, but must be limited to those interferences that are determined by an actual need, to be evaluated according to the concrete circumstances of the case and, in any case, always with a view to protecting the interests of the child.

To this end, the Data Protection Authority has also identified, some best practices that include, among others, confronting children about the use of the Internet, explaining the possible risks and listening to what they say; imposing

rules to control and limit online interactions; avoiding oversharing of data and being cautious and prudent in the case of online acquaintances.

Only in this way, the construction of a healthy, aware and safe virtual environment be possible will be possible.

Bibliography

Pennetta, A. L. (2019) *Bullismo, cyberbullismo e nuove forme di devianza*. Torino: Giappichelli.

Salter, A. C. (1995) *Transforming Trauma: A Guide to Understanding and Treating Adult Survivors of Child Sexual Abuse*. Thousand Oaks: SAGE Publications.

Sorgato, A. (2019) *Revenge porn. Aspetti giuridici, informatici e psicologici*. Milano: Giuffrè Francis Lefebvre.

Ziccardi, G. (2023) 'I minori online tra videogiochi e metaverso', *Ciberspazio e Diritto*, 3(2023), pp. 325 – 336.

Chapter XVIII
Online Reputation and Right to Be Forgotten

by Simone Bonavita*

Index: 1. The right to be forgotten: legal foundations and implications. – 2. Data collection and management in smart cities. – 3. The right to be forgotten in smart cities: open issues.

1. The right to be forgotten: legal foundations and implications

The concept of the right to be forgotten has its roots in the desire of individuals to "be forgotten" online, especially in relation to information that is no longer relevant or that could damage their reputation after some time. This right is now also affirmed in the General Data Protection Regulation 679/2016 (GDPR) in Article 17, headed "Right to erasure (right to be forgotten)":

1. 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:

1. 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).
2. 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;
3. 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);
4. the personal data have been unlawfully processed;
5. 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;
6. the personal data have been collected in relation to the offer of information society services referred to in Article 8(1).

---

* Professor AC of sensitive data treatment at the University of Milan, attorney and Executive Director of the Information Society Law Center (ISLC).

2. Where the controller has made the personal data public and is obliged pursuant to paragraph 1 to erase the personal data, the controller, taking account of available technology and the cost of implementation, shall take reasonable steps, including technical measures, to inform controllers which are processing the personal data that the data subject has requested the erasure by such controllers of any links to, or copy or replication of, those personal data.

3. Paragraphs 1 and 2 shall not apply to the extent that processing is necessary for:

1. exercising the right of freedom of expression and information;
2. compliance with a legal obligation which requires processing by Union or Member State law to which the controller is subject or for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller;
3. reasons of public interest in the area of public health in accordance with points (h) and (i) of Article 9(2) as well as Article 9(3);
4. archiving purposes in the public interest, scientific or historical research purposes or statistical purposes in accordance with Article 89(1) in so far as the right referred to in paragraph 1 is likely to render impossible or seriously impair the achievement of the objectives of that processing; or
5. the establishment, exercise or defence of legal claims.

In the GDPR, the legislature wanted to expressly provide this option for the data subject, precisely to give due weight to this important right.

The implications of the right to be forgotten are, therefore, profound and manifold, especially in the digital age, where information spreads rapidly and can remain accessible indefinitely. This right puts an obligation on data controllers, in particular, not only to delete data within their own systems, but also to take reasonable steps to inform others who may have made that data public, including links or copies, of the request for deletion.

2. Data collection and management in smart cities

The technologies behind smart cities collect, store and use the personal data of the various individuals who populate them to provide innovative services, but this cannot help but have, conversely, numerous potential implications for the privacy of those citizens.

Data collection in smart cities takes place through an extensive network of smart devices distributed throughout the city. These devices may be owned by companies that provide smart city-related services or by citizens who wish to use specific services.

These devices collect real-time data on various aspects of urban life, such as traffic, air quality, energy consumption, and business activity. This data is then

sent subject to processing and analysis, to you not only to provide service to citizens, but also to optimize the service itself.

The main types of data processed include location data, essential for traffic and public transport management; demographic data, used for service and infrastructure planning; data related to the use of utilities such as water, electricity and waste, which help optimize urban resource management; and security data, collected through cameras and sensors to ensure a safer urban environment, geolocation data, data related to payment instruments and many others.

2.1. Focus: payment data

In the context of smart cities, data on service payments, including those made through credit cards, assume a key role within the digitized urban ecosystem. This information makes it possible to simplify and make more efficient economic transactions related to a wide range of services from urban mobility through the payment of municipal taxes to the use of public and cultural services. Therefore, smart cities' digital infrastructures typically implement advanced security protocols, such as data encryption and the adoption of multifactor authentication systems, to ensure the confidentiality and integrity of transactions. In addition, adherence to international payment data security standards, such as the Payment Card Industry Data Security Standard (PCI DSS), is essential to ensure that citizens' data is handled with the highest security guarantees.

2.2. Focus: location data

Location data is among the most sensitive information handled by smart cities applications, as it reveals the daily movements and habits of individuals. Protecting this data is crucial to preserving citizens' privacy by preventing risks related to excessive surveillance and unwanted profiling. The collection and analysis of location data, if not properly protected, can lead to the creation of detailed profiles of individuals, exposing them to risks of discrimination, behavioral manipulation, and violations of their privacy. For this reason, it is critical that smart cities adopt privacy by design policies in compliance with the GDPR as well, limiting data collection to only what is necessary and ensuring transparency about how it is used. It is also important to implement advanced technical measures, such as data anonymization and minimization, to protect the identity of individuals and reduce the risks associated with profiling. In this context, compliance with data protection regulations, such as the GDPR, ensures that citizens' rights are safeguarded through, among others, informed consent, the ability to access their data, and the right to have their data deleted.

Once collected, data must be managed efficiently to ensure that they can be used to improve urban services. This includes data cleansing, integration, storage, and analysis. Effective data management requires the use of advanced

technologies, such as distributed databases, cloud computing, and machine learning algorithms, which can process large volumes of data in real time.

Important privacy and security issues must also be considered in such processing cases. It is critical that smart cities adopt robust privacy policies and use data protection technologies, such as encryption and anonymization, to protect citizens' sensitive information and comply with data protection regulations such as GDPR for the processing of personal data.

2.3. Focus: so-called open government data or "open data"

The sharing of personal data through open data platforms, while intended to promote transparency and innovation in smart cities contexts, raises significant privacy and data security issues. These data, if not properly anonymized or if shared without strict safeguards, can expose individuals to risks of direct or indirect identification, compromising their privacy. They also pose real issues in relation to the concrete application of the right to be forgotten.

The subject has recently been innovated by the provisions of EU Directive 2019/1024, which organically regulates the process of open data and the reuse of public sector information.

This Directive, provides that four main types of open data can be identified, namely:

- **dynamic data** in digital format, subject to frequent or real-time updates, particularly because of their volatility or rapid obsolescence; sensor-generated data are usually considered dynamic data;
- **research data** in digital format, other than information contained in scientific publications, collected or produced in the course of scientific research and used as evidence in the research process, or commonly accepted in the research community as necessary to validate research conclusions and findings;
- **high-value datasets** whose reuse is associated with important benefits to society, the environment, and the economy, sin particular in view of their suitability for the creation of services, value-added applications, and new decent, high-quality jobs, as well as the number of potential beneficiaries of value-added services and applications based on these datasets;
- **raw data** government-derived data that has not undergone any modification or reuse.

As will be well understood, applying the right to be forgotten to these different types of data is quite complex, as it will be strictly necessary to consider the individual implications of each case.

3. The right to be forgotten in smart cities: open issues

Effectively managing the right to be forgotten in smart cities is a complex but critical challenge to ensuring privacy in the digital age.

3.1. Right to be forgotten and interconnection

Determining which data should be deleted can be complex, especially when data is spread across multiple, interconnected platforms or systems owned by different companies. Identifying the specific data to be deleted in response to a citizen request to exercise the right to be forgotten is one of the most significant challenges for smart cities operators. The interconnected and complex nature of urban digital infrastructures, which collect, store and process significant volumes of personal data across a variety of applications and services, makes it challenging to track and isolate information about an individual. This operation requires advanced technological systems capable of effectively analyzing and managing the data, ensuring that every piece of information that can be traced back to a specific person is accurately identified and, if necessary, eliminated. The effectiveness of this process depends heavily on the accuracy of the cataloguing tools and the capability of the recognition algorithms, which must be designed to deal with the complexity and dynamism of urban data systems.

3.2. Reconciliation with historical data

Reconciliation with historical data is a critical challenge in exercising the right to be forgotten in smart cities, especially considering the large number of data controllers involved. Smart cities are complex ecosystems where personal and historical data are collected and processed by a multitude of public and private entities, each with its own data management system. This fragmented landscape greatly complicates the process of identifying and deleting specific information about an individual, as each data controller may have different policies, procedures and technologies for handling such data. In addition, historical data are often stored in formats that make it difficult to distinguish and isolate personal information without compromising the integrity of the overall dataset. Interoperability between different information systems therefore becomes a fundamental requirement, but at the same time a significant challenge, to ensure that a citizen's request for a forgetting application is fully and effectively respected across all levels and actors involved. Addressing this issue requires a coordinated approach and the establishment of common interoperability and data management standards, as well as a concerted effort to continuously update privacy technologies and policies in response to evolving regulations and citizen needs.

3.3. Secure destruction

Security risks in the destruction of personal data pose a significant challenge in the context of smart cities, compounded by the large number of data controllers involved. Each entity involved in data management, whether public agencies, private companies or other third parties, very often operates as an autonomous data controller, following its own security procedures and policies. This diversity of approaches can complicate the assurance of secure and permanent destruction of personal data, as the procedure must be executed consistently and reliably across all systems and platforms. The challenge lies not only in deleting data from active databases, but also in ensuring that they are removed from any backups, storage copies, and, most importantly, that no recoverable traces remain that could expose personal information to the risk of unauthorized access or data breaches. This task is made even more daunting by the interconnectedness of digital services in smart cities, which requires flawless synchronization and communication between data controllers and data controllers (under the GDPR) to ensure that the request to be forgotten is complied with in full. To effectively address these risks, it is essential to establish shared security protocols and cross-checking mechanisms that ensure the integrity of the data deletion process, thereby ensuring the protection of individuals' privacy.

3.4. Public administration data and personal data of private individuals

The handling of right to be forgotten requests involving the transmission of personal data from public administration to private companies raises complex legal issues, arising from the application of potentially divergent regulations. In this context, the processing of personal data lies at the intersection between the public regulatory regime, governed by principles of transparency, access to information, and data protection specific to the public sector, and the private regulatory regime, governed primarily by the GDPR in the European Union or similar regulations in other jurisdictions. The discrepancy between these regulatory regimes can generate ambiguity in the interpretation of data subjects' rights, particularly regarding the legal basis for processing, the terms for data retention, and the conditions under which the right to be forgotten can be exercised.

In addition, responsibility for implementing requests for deletion becomes a critical node: while government may be required to comply with strict standards of data protection and data subjects' rights, private companies may not be subject to the same obligations or may interpret their responsibilities differently, especially if they operate in multiple jurisdictions with disparate compliance requirements. This highlights the need to establish clear protocols and data-processing agreements that explicitly outline the roles and responsibilities

of each party involved, including procedures for responding to right-to-forget requests, to ensure that all parties act in accordance with applicable regulations and the principles of proportionality, data minimization, and respect for the fundamental rights of data subjects. Resolving these issues requires a thorough legal assessment to ensure that right to be forgotten requests are handled in a manner that respects both individual privacy needs as well as regulatory obligations and public interests, thereby avoiding legal conflicts and ensuring the effective protection of personal data in a complex and increasingly interconnected digital ecosystem.

3.5. Right to be forgotten and historical memory of the smart city

The tension between the preservation of the historical memory of cities, particularly those using advanced technologies to become "smart cities", and the right of individuals to be forgotten represents a complex legal dilemma. On the one hand, the collection and storage of personal data through advanced digital systems enable detailed documentation of urban events, interactions and evolutions, contributing significantly to the historical understanding and analysis of the city fabric. On the other hand, the right to be forgotten, as mentioned above, under certain conditions grants individuals the ability to request the deletion of their personal data from records and databases when such information is no longer necessary for the purposes for which it was collected or when data subjects withdraw their consent to its processing. The legal challenge emerges in harmonizing these two seemingly opposing interests: on the one hand, the need to preserve an accurate and accessible record of urban dynamics for historical, research and planning purposes; on the other hand, the need to protect individuals' rights to privacy and self-determination. All this brings back an additional problem: much of the data is held by private operators. The heart of the matter lies in identifying clear and operational criteria that determine when and how personal data can be processed for archiving purposes in the public interest, scientific or historical research, without prejudice to the rights of data subjects. A balanced legal solution requires a careful assessment of the specific circumstances of each case, considering factors such as the importance of the data for historical understanding, the possibility of anonymization of personal data to reduce the impact on privacy, and the presence of an overriding public interest.

Bibliography

Bonavita, S. (2017) 'Le ragioni dell'oblio', *Ciberspazio e diritto: rivista internazionale di informatica giuridica*, 1, pp. 85-111.

Kitchin, R. (2014) *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences*. Los Angeles: Sage.

Townsend, A. (2013) *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*. New York: Norton & Company.

References

European Data Protection Supervisor (2020) *The EDPS Strategy 2020-2024: Shaping a Safer Digital Future*. Available at: https://www.edps.europa.eu/data-protection/our-work/publications/strategy/edps-strategy-2020-2024-shaping-safer-digital-future_en.

SMART CITIES, ARTIFICIAL INTELLIGENCE AND DIGITAL TRANSFORMATION LAW

A Handbook for Students and Professionals

Edited by E. E. Akin, S. Klimbacher, G. Ziccardi

This open access handbook is the course book adopted by the faculty of law at the University of Milan for the course entitled: "Smart Cities, Artificial Intelligence and Digital Transformation Law".

The subject matters analysed in the book are divided in three parts. The first part focuses on the digital transformation law relevant to the advent of legal informatics, the need for digital skills and literacy, the European digital strategy, and on the American and Chinese approach to the digital society. The second part focuses on the relevant issues connected to artificial intelligence with a specific focus on data protection, healthcare, ethics and copyright. Finally, the third part provides a framework on smart cities, computer crimes and data protection.

Students, professionals and scholars around the world are encouraged to read this handbook to study the legal and technical aspects of the digital environment. Teachers are welcome to adopt it for their digital and non-digital courses.

ISBN 979-12-5510-164-2 (print)
ISBN 979-12-5510-166-6 (PDF)
ISBN 979-12-5510-167-3 (EPUB)
DOI 10.54103/infolawsoc.186