LUISS

Masters' Degree in International Relations

Course: Security Law and Constitutional Protection

Balancing Freedom of Speech and Security Imperatives in the AI Era: The EU Artificial Intelligence Act

Prof. Raffaele Bifulco

SUPERVISOR

Prof. Elena Griglio

CO-SUPERVISOR

Giulia Petroni, ID 657352

CANDIDATE

Academic Year 2024/2025

Abstract

As terrorism increasingly spreads in the digital sphere, the global response to counter it quickly follows it, raising a fundamental question: what is the right balance to be struck between security and freedom of speech when establishing counterterrorism and anti-incitement to terrorism policies online? Contemporary counterterrorism measures in the digital space, indeed, raise significant concerns about potential infringements on fundamental rights, particularly freedom of expression, fueling the ever-existing tension between human rights protection and security concerns. The growing use of artificial intelligence (AI) and new technologies in surveillance, content regulation, and predictive policing has further complicated this debate, since the employment of these technologies in counterterrorism raises numerous concerns regarding censorship and suppression of legitimate speech. Furthermore, thanks to these technologies, counterterrorism is no longer solely the domain of public authorities, due to the increasing involvement of private big tech companies in regulating content and enforcing security measures. Thus, this thesis critically analyzes the role of AI in balancing security and freedom of speech in the context of online counterterrorism in Europe. It therefore concentrates on analyzing the recent adoption at the European Union level of the Artificial Intelligence Act (AI Act), evaluating its implementation in various EU Member States and its impact on the balance between security and freedom of speech in counterterrorism. Consequently, this new European regulatory approach will be compared to the one adopted by the United States, analyzing whether the AI Act has the potential to shape the global future of AI governance, especially in the fields of online counterterrorism and digital rights protection.

Dedication

Ai miei genitori, Angela e Paolo,

a Giada,

a Gabriele,

e a tutta la mia famiglia,

che in ogni passo di questo cammino

sono stati per me guida e presenza costante.

«I' mi ristrinsi a la fida compagna:

e come sare' io sanza lui corso?

chi m'avria tratto su per la montagna?»

Dante Alighieri, Purgatorio III, vv. 4-6

Acknowledgements

I wish to express my deepest gratitude to my thesis advisor, Professor Raffaele Bifulco, whose invaluable mentorship and insight have been pivotal throughout the writing process of this thesis.

I am also immensely grateful to Dr. Federico Micari and Dr. Giorgia Valentini for their constant availability, constructive feedback, and support which have been essential to refine this work.

I extend my sincere thanks also to Professor Elena Griglio, whose teaching, together with that of Professor Bifulco and the entire team of the *Security Law and Constitutional Protection* course, was instrumental in inspiring my interest in the complex issue of the balance between security imperatives and the protection of fundamental rights.

A final thanks goes to Professor Pamela Harris for her course *Free Speech in a Comparative Perspective* which enriched the theoretical foundations of this thesis and for all her meaningful insights.

Table of Contents

List of Abbreviations

ACLU	American Civil Liberties Union
ACN	Agenzia Nazionale per la Cybersicurezza (National Cybersecurity Agency)
AESIA	Agencia Española de Supervisión de la Inteligencia Artificial (Spanish Artificial Intelligence Supervisory Agency)
AgID	Agenzia per l'Italia Digitale (Agency for Digital Italy)
AI Act	European Artificial Intelligence Act
AI	Artificial Intelligence
BKA	Bundeskriminalamt (Federal Criminal Police Office)
CCIF	Collectif Contre L'Islamophobie en France (Collective Against Islamophobia)
CCPA	California Consumer Privacy Act
CCTV	Closed-Circuit Television
CDA	Communications Decency Act
CDU	Christlich Demokratische Union Deutschlands (Christian Democratic Union of Germany)
CEP	Counter Extremism Project
CFI	Court of First Instance
CISA	Cybersecurity and Infrastructure Security Agency
CJEU	Court of Justice of the European Union
CNIL	Commission nationale de l'informatique et des libertés (French Data Protection Authority)
CoE	Council of Europe
CPS	Cyber-Physical Systems

CTC	UN Counter-Terrorism Committee
CTED	UN Counter-Terrorism Committee Executive Directorate
DDL	Disegno di Legge (Draft Law)
DFKI	Deutsches Forschungszentrum für Künstliche Intelligenz (German Research Center for Artificial Intelligence)
DHS	Department of Homeland Security
DPA	Data Protection Authority
DSA	Digital Service Act
DOJ	US Department of Justice
DPIA	Data Protection Impact Assessment
DSA	Digital Services Act
ECHR	European Convention of Fundamental Rights
ECTC	European Counter Terrorism Center
ECtHR	European Court of Human Rights
ECJ	European Court of Justice
ETA	Euskadi ta Askatasuna
EU Charter	European Charter of Fundamental Rights of the European Union
EU	European Union
EUROPOL	European Union Agency for Law Enforcement Cooperation
FBI	Federal Bureau of Investigations
FTC	Federal Trade Commission
FTFs	Foreign Terrorist Fighters
FSL	Few-Short Learner

GDPR	General Data Protection Regulation
GIFCT	Global Internet Forum to Counter Terrorism
GPAI	General-Purpose Artificial Intelligence
HMA	Hasher-Matcher-Actioner
HSPs	Hosting Service Providers
ICCPR	International Covenant on Civil and Political Rights
IEEE	Institute of Electrical and Electronics Standards Association
INTERPOL	International Criminal Police Organization
IoT	Internet of Things
IPC	Code de la propriété intellectuelle (Intellectual Property Code)
IRU	Internet Referral Unit
ISR	Intelligence, Surveillance, and Reconnaissance
LKA	Landeskriminalamt Rheinland-Pfalz (Rhineland-Palatinate State Criminal Police Office)
LLMs	Large Language Models
LTTE	Liberation Tigers of Tamil Eelam
ML	Machine Learning
NGOs	Non-Governmental Organizations
NIS	Network and Information Security Directive
OHCHR	Office of the High Commissioner for Human Rights
OSCE	Organization for Security and Co-operation in Europe
PCLOB	Privacy and Civil Liberties Oversight Board
PKK	Partiya Karkeran Kurdistan

SARI	Automatic Image Recognition System
S&T	Science and Technology Directorate
SPD	Sozialdemokratische Partei Deutschlands (Social Democratic Party of Germany)
TATE	Tech Against Terrorism Europe
TAT	Tech Against Terrorism
TCO	Terrorist Content Online Regulation
TEU	Treaty on European Union
TFEU	Treaty on the Functioning of the European Union
TFTP	Terrorist Finance Tracking Programme
TLSH	Text Locality Sensitive Hashing
UDHR	Universal Declaration of Human Rights
UK	United Kingdom
UNOCT	United Nations Office of Counter-Terrorism
UNODC	United Nations Office on Drugs and Crime
UNSC	United Nations Security Council
UN	United Nations
USSR	Union of Soviet Socialist Republics
US	United States
VLOPs	Very Large Online Platforms
VLOSEs	Very Large Online Search Engines

Introduction

If a person from the 20th century were to wake up in the present days, they might perceive the world has having experienced a deep transformation which, in less than 100 years, has allowed human beings to simultaneously live in two separate but interconnected realms: the physical world, governed by territorial boundaries, formal institutions and codified laws, and the digital one, where human interactions transcend geographical borders and where norms and regulations are often ambiguous and inconsistently enforced. The digital realm shares some aspects with a utopian reality: by allowing borderless communication, it democratizes knowledge and expands civic participation, while enhancing self-empowerment and expression in unprecedented ways (Bromell, 2021, p. 32). However, if that person from the 20th century were to be George Orwell, he would have been more struck by the dystopian aspects of this world, perhaps believing that his *1984* had, in some forms, materialized: the digital realm, indeed, enables increased mass surveillance and data collection, as well as the spread of "misinformation, disinformation … extremism" and propaganda (Bromell, 2021, p. 32). Terrorist and extremist groups, for instance, have since the 1990s, increasingly exploited digital platforms to disseminate propaganda and recruit members, with the result that even states and counterterrorism agencies have moved their fight against extremism, violence, and terrorism in the digital realm. Importantly, in this digital space, which transcends national borders and law, private corporations wield unprecedented power over the conditions and regulations of speech and of other digital rights. The problem is that such profit-driven companies are not directly bound by constitutional provisions as are public authorities for the protection of fundamental rights, hence the restrictions they impose on the users

of their platforms, for instance to counter the spread of extremist and terrorist content, are less contestable and accountable than those imposed by democratic states in the physical world.

With the spread of artificial intelligence (AI), both the positive and negative characteristics of this dual world have been augmented. In only a few years, indeed, AI has permeated so deeply every aspect of human life, from finance and infrastructure to healthcare and administration, that it is hard to imagine a digital or a physical world existing without it. Crucially, it has even been employed for terrorist and counter terrorist purposes, further cementing the digital world as a central arena in the struggle between terrorism and counterterrorism. As terrorist and extremist groups are increasingly exploiting the digital realm and AI systems to radicalize, recruit and incite to terrorism, governments and security agencies are responding with an increasing variety of online security interventions such as AI-led content moderation and surveillance. In fact, when employed for counterterrorism purposes, AI systems can serve security imperatives by increasing the predictive power of existing counterterrorism strategies, reducing their response time and enhancing their capabilities. At the same time, in the absence of adequate accountability and transparency safeguards, these systems can generate over-censorship and infringe fundamental rights, especially the right to freedom of expression. Thus, when AI systems are used in counterterrorism measures, they can easily exacerbate the ever-existing tension between security and fundamental rights.

Mindful of AI's great potential and of its virtually unlimited use, the European Union has sought to establish global standards on the transparent, safe and human-centric development and use of AI, by adopting the EU Artificial Intelligence Act (AI Act), a comprehensive and risk-based legal framework which aims to safeguard Europeans' fundamental rights, including when AI is used in security operations. Importantly, the Act addresses AI application in sensitive areas such

as biometric surveillance and content moderation, even if it carves out extensive exemptions for national security and military purposes, thus creating possible gaps in domains where fundamental rights, such as the right to freedom of speech, are more at risk (Cabrera, 2024). However, the Act represents a landmark achievement in promoting a comprehensive regulation of AI, which strives to balance innovation, security and fundamental rights like no other existing legal frameworks worldwide. Indeed, when the AI Act's legal framework is compared to that of the United States, with their libertarian and innovation-driven approach to AI, it appears evident that the EU is striving to find a perfect balance between security imperatives and fundamental rights protection in the application of AI in every field of life, even in highly sensitive fields such as online counterterrorism. This signals the intent of the EU to ensure that potentially dangerous systems are regulated when being developed and used within the EU Single Market, enforcing the provisions of the Act even on foreign companies operating within the EU borders.

With a special focus on the European legal framework, this thesis examines in depth the complex relationship between artificial intelligence, counterterrorism, and the right to freedom of speech in the digital age, trying to address some core questions regarding this relationship, namely: Can freedom of speech be effectively safeguarded in a world where counterterrorism measures are increasingly employing AI systems, governed by opaque algorithms and shaped by corporate interests? Considering the new Artificial Intelligence Act, what will be the role of AI in balancing security and freedom of speech in the field of online counterterrorism in Europe? Compared to the US approach, could the new EU's regulatory framework become the reference model for AI governance in counterterrorism online or will new legal frameworks be needed to reconcile free speech and security in this platform society?

In trying to answer these questions, the first chapter introduces the foundational issue of this thesis by considering the substantial tension between security and freedom of speech which arises in the context of online counterterrorism, analyzing the growing role of online platforms in this domain which creates tensions between public and private regulatory responsibilities. The second chapter examines the impact of artificial intelligence on terrorism, counterterrorism and freedom of speech, examining whether AI can uphold both security imperatives and fundamental rights and looking at the ethical standards on AI development and use promoted by international organizations. The third chapter focuses exclusively on the EU AI Act, assessing both its implementation in four EU Member States, namely Italy, Germany, France and Spain, and its potential to strike a fair balance between security and freedom of speech in AI-led online counterterrorism measures. Chapter 4 follows with a comparative analysis between the EU and the US approaches to the protection of freedom of speech and AI governance in the field of counterterrorism, evaluating both models' strengths and weaknesses. In the conclusion, this thesis offers some final reflections on the future of AI regulation, arguing that the EU Artificial Intelligence Act represents a significant attempt to foster a safe and transparent AI which not only could provide a global model for AI governance but it could be able to reconcile security imperatives with the protection of freedom of speech and other fundamental rights even when AI is employed for counterterrorism purposes.

Methodology

Investigating the delicate balance between the protection of security and of fundamental rights, and especially of the right to freedom of speech, in the context of AI-driven online counterterrorism, and evaluating whether the European AI Act offers a viable legal framework for

reconciling these competing imperatives are the primary objectives of this thesis. In so doing, it combines a doctrinal legal research methodology with a comparative legal research methodology, merging legal analysis with elements of policy evaluation and ethical/philosophical reflection. This multidisciplinary approach ensures a detailed examination of the research questions, grounded in a normative analytical framework which consents to evaluate the existing legal approaches on AI-driven terrorism against fundamental rights and ethical standards. The principle of proportionality, discussed in Chapter 1, will serve to evaluate whether current AI-driven online counterterrorism efforts effectively balance security imperatives with the right to freedom of speech.

The doctrinal legal research methodology allows this thesis to provide a "descriptive and detailed analysis of legal rules found in primary sources", including EU legislation, such as the EU AI Act, international human rights instruments, such as the ICCPR, and relevant national constitutional provisions (Jerome Hall, 2019). This method is employed not only to "describe the law" but also to comment on the sources and identify their similarities and differences (Jerome Hall, 2019). Chapters 1 and 2 of this thesis rely primarily on this approach, combined with ethical and analytical elements: the former analyzes the legal constitutionalization of the right to freedom of speech at international and at European level, and then it examines the current regulations on online counterterrorism; the latter applies this framework to the context of artificial intelligence, assessing how AI technologies intersect with the current legal standards on freedom of speech and counterterrorism.

Then, the comparative legal research methodology is used to critically assess "different bodies of law to examine how the outcome of a legal issue could be different under each set of laws" (Jerome Hall, 2019). Chapter 3 employs this method to compare the legal and policy

responses of four EU Member States, namely Italy, Germany, France and Spain, to the enactment of the EU AI Act. This comparative method allows to understand how different national constitutional constraints interact with EU law in the fields of AI, freedom of speech, security and counterterrorism. Chapter 4 also adopts a comparative methodology, examining the approaches of the European Union and of the United States in the fields of free speech, AI, and AI-driven counterterrorism. This transatlantic legal analysis reveals not only similarities and differences between the two approaches, but also potential pathways for international harmonization of AI rules.

Concerning the sources, the data triangulation method will be employed, focusing on the use of both primary and secondary sources, from different countries, organizations, and periods of time. Primary sources will be the most employed, with a special focus on relevant national constitutional provisions, international and regional treaties, such as the ICCPR and the ECHR, and other legal instruments. Relevant case law from the European Court of Human Rights, the Court of Justice of the EU and of the US Supreme Court, parliamentary bills and official policy documents will be examined alongside relevant reports from important international organizations in the fields of freedom of speech, counterterrorism and AI. These primary sources will be fundamental to analyze the current legal approach to the balance between the right to freedom of speech and security in AI-driven online counterterrorism. Secondary sources, especially academic journals, books, and policy analyses from research institutes, will also be employed to strengthen the analysis and enrich the research with the current most relevant perspectives on these fields. These qualitative sources enable a rich interpretation of the normative structures underpinning AI governance and its implications for digital rights especially in the context of AI-driven online counterterrorism measures. This thesis adopts the APA citation style but, given their normative

primacy and enduring nature, the constitutions of the countries under examination are cited without indicating year or page number.

All in all, the implementation of both doctrinal and comparative legal methods, anchored in a normative analytical framework, provides a comprehensive and rigorous approach to understanding the evolving relationship between AI governance, freedom of speech, and security, especially in light of the new EU AI Act. However, this methodology carries with it some limitations as well. The doctrinal approach "may be too formalistic" and, with the lack of empirical data, it might underrepresent the real-world impact of AI-driven online counterterrorism on individuals (Jerome Hall, 2019). Instead, the availability and accessibility of relevant sources may affect the comparative approach, creating gaps in the analysis when information are not accessible or scarce. Such shortcomings do not however impede this thesis to effectively enter the ongoing scholarly debate on AI-driven counterterrorism, by offering a detailed legal and comparative analysis of the new EU AI Act's impact on the balance between security and freedom of speech in the field of AI-driven online counterterrorism. The thesis also bridges the EU and US perspectives on this issue, analyzing the emerging role of private tech companies in both jurisdictions in AI-driven content moderation and promoting a rights-based approach to AI-driven online counterterrorism capable of protecting both security imperatives and fundamental rights.

1. Counterterrorism in the Digital Age: The Dilemma Between Security and Freedom of Speech

The ever-existing divide between security and human rights emerges particularly prominent when considering what is the right balance to strike between ensuring security and protecting the right to freedom of speech online with regard to cases of counterterrorism. This chapter aims to explore exactly such matters. In so doing, the first paragraph gives a brief introduction to the right of freedom of speech and its legitimate restrictions, as well as its evolution in the digital era; the second then focuses on analyzing how terrorism and consequently counterterrorism have been increasingly taking advantage of the digital sphere to pursue their objectives; consequently, the third paragraph examines the difficult task of finding the right balance between security imperatives and freedom of speech protection in the field of counterterrorism online, dealing with the central issue of this thesis. The chapter goes on by exploring, in paragraphs four and five, the role of online platforms in both protecting and restricting freedom of speech and in counterterrorism, considering in the end the great tension which arises between the private and public sector in this field. Being Europe the focus of this thesis, this chapter will devote particular attention to this geopolitical area.

1.1 Introducing the Right to Freedom of Speech

A defining pillar of liberal democracy since its inception during the Enlightenment with roots as ancient as those of Athens' democracy, the right to freedom of speech has always served as a bastion against tyranny and a defensor of free public discourse and democratic governance.

What the Ancient Greeks distinctively referred to as *isegoria*, i.e. "the equal right of citizens to participate in public debate in the democratic assembly", and *parrhesia*, i.e. "the license to say what one pleased, how and when one pleased, and to whom", have, since the Enlightenment, converged into what is now recognized as the right to freedom of speech, reflecting both the democratic equality principle enshrined in *isegoria* and the principle of individual liberty embodied in *parrhesia* (Bejan, 2019, p. 95). Indeed, before the Enlightenment, the traditional or pre-modern approach to freedom of speech had triumphed, with the Catholic Church endorsing only one unquestionable truth and considering freedom of speech as the freedom to advocate that one truth. Enlightenment philosophers, however, questioned the unquestionable, endorsing a liberal view of the right to freedom of speech: Voltaire's "rallying cry was, (indeed,) *écrasez l'infâme*" (let us crush the evil thing)" going against the religious dogmas and defending people's right to disagree (Gruberg, 2009). From the early defenses of John Milton, in his 1644 *Areopagitica*, of John Locke, in his 1689 *A Letter Concerning Toleration,* and of John Stuart Mill's utilitarian analysis in his 1859 masterpiece *On Liberty*, the right to freedom of speech increasingly started to be considered essential not only for the pursuit of truth and societal progress, but also for the very concepts of human dignity and individual liberty (Mill, 1859). Mill argued that even the most unpopular or seemingly false opinions may contain elements of truth, and that silencing them robs society of the chance to achieve a fuller understanding of the issue at hand: through open debate, in fact, even false ideas may bring about alternative perspectives on an issue and allow the beholder of these ideas to scrutinize their beliefs, thus leading to personal growth and the rejection of societal dogmas (Mill, Chapter 2, 1859; Erdos, 2019, p. 20).

Freedom of speech thus appears strictly intertwined with classical liberalism's faith in rationality and individual judgment, because it allows people to think freely and express

themselves independently of state control or social conformity. This, in turn, not only improves individuals' judgment and knowledge but also contributes to the betterment of society at large, supporting the enhancement of an enlightened community. To understand how connected the conception of freedom of speech and the perception of the self during the classical liberal period are and beyond, it is sufficient to consider the three main philosophical rationales of this right. Feldman (2017, pp. 1124-1125) explains that, since the "self… exists prior to society and culture; is its own sovereign center of power; and enjoys maximum liberty so long as government is absent", freedom of speech appears not only as the best way for humans to distinguish what is true and what is false (search-for-truth rationale) and to realize their own full potential as a citizen (self-fulfillment rationale), but also as the most important principle of all democratic societies (self-governance rationale). By definition, a government can only be considered democratic if it guarantees freedom of speech, allowing all members of society to participate in the political and social life of their community, either directly or indirectly by electing representatives.

Freedom of speech was then eventually codified into legal instruments as liberal constitutionalism spread during the 18th and the 19th centuries: the *Swedish Freedom of the Press Act*, now part of the Swedish Constitution, was formulated in 1766, becoming the "first in the world to establish freedom of expression in speech and writing as a protected legal right" (Erdos, 2019, p. 20; Nordin et al., 2023, p. 4). After that, during the French Revolution, the 1789 *Declaration of the Rights of Man and of the Citizen* was enacted, establishing the right to freedom of speech in Art. 11 and describing it as 'a most precious right' for humanity (Erdos, 2019, p. 20; Declaration, 1789). The *Declaration* became one of the first constitutional recognitions of the right after the 1689 English *Bill of Right* which recognized freedom of speech in Parliament by declaring "that the freedom of speech and debates or proceedings in Parliament ought not to be impeached

or questioned in any court or place out of Parliament" (English Bill of Rights, 1689). After the American Revolution, then, the enactment of the US 1791 *Bill of Rights* constitutionalized the right to freedom of speech within the American legal system, especially thanks to the First Amendment which provided strong protection of the right against Congress' intervention and any type of censorship (US Const. Amend. I).

During the 20th century, then, the right to freedom of speech obtained a most prominent position in the construction of a liberal international order after the destruction brought about by the two World Wars and by the Cold War (Erdos, 2019, p. 20): new international human rights treaties, such as the 1966 International Covenant on Civil and Political Rights (ICCPR), constitutionalized the right under international law, while the then-still-young United Nations defined the right to freedom of speech as a universal right with the promulgation of the 1948 *Universal Declaration of Human Right* (UDHR). As liberal democracies, with the United States at the forefront, started spreading liberalism and all its values worldwide, the right to freedom of speech emerged as a fundamental cornerstone of international human rights law. During the Cold War, in a world split in two, with the US and its allies with their liberal values on the Western side and the USSR and its communist values on the Eastern one, the United States used the right to freedom of speech not only as a legal right with the codification of international post-WWII treaties, but also as a powerful ideological symbol. It became one of the central elements of Western propaganda, portraying democratic societies as havens where people could live a life with the certainty that their fundamental rights would be upheld, that they could freely express dissent, that they could engage in open political dialogue and hold democratic elections, all opportunities which were absolutely denied under the repressive communist regimes. The power of the right to freedom of speech is indescribably immense, especially because it carries with it a captivating

promise: the possibility to think independently, to demand respect for one's fundamental rights, to shape one's society, and ultimately, to shape oneself.

As of today, the right to freedom of speech is codified, under international law, in Article 19[1] of the ICCPR, which gives legal force to the same Article[2] of the UDHR. In Europe, then, the right to freedom of expression is further protected by Article 10[3] of the 1950 *European Convention on Human Rights* (ECHR) of the Council of Europe and Article 11[4] of the *Charter of Fundamental Rights of the European Union* (EU Charter), which became legally binding on December 1, 2009. Henceforth, there are different levels of protection of the right to freedom of speech, and other fundamental rights as well, within the EU Single Market and the borders of its Member States: at

---

[1] Article 19 of the ICCPR reads:

> "1. Everyone shall have the right to hold opinions without interference.
> 2. Everyone shall have the right to freedom of expression; this right shall include freedom to seek, receive and impart information and ideas of all kinds, regardless of frontiers, either orally, in writing or in print, in the form of art, or through any other media of his choice.
> 3. The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:
> (a) For respect of the rights or reputations of others;
> (b) For the protection of national security or of public order (ordre public), or of public health or morals" (ICCPR, 1966).

[2] Article 19 of the UDHR reads:

> "Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers" (UDHR, 1948).

[3] Article 10 of the ECHR reads:

> "1. Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. This Article shall not prevent States from requiring the licensing of broadcasting, television or cinema enterprises,
> 2. The exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary" (ECHR, 1950).

[4] Article 11 of the EU Charter of Fundamental Rights reads:

> "1. Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers.
> 2. The freedom and pluralism of the media shall be respected" (EU Charter, 2000).

the national level, there are the constitutions of the Member States with their laws and regulations; at the European level, there is EU law, specifically the EU Charter and the ECHR; at the international level, there are the international human rights treaties such as the ICCPR. The European Union, with its "good international citizenship", is mostly considered as 'monist' in its approach to public international law, allowing most of the "binding international norms to become part of its legal order" (Dunne, 2008, as cited in Wessel, 2011; Wessel, 2011, p. 3). Indeed, Article 21(1)[5] of the Treaty on European Union requires that:

> "The Union's action on the international scene shall be guided by the principles which have inspired its own creation, development and enlargement, and which it seeks to advance in

[5] Article 21 of the TEU states:

1. "The Union's action on the international scene shall be guided by the principles which have inspired its own creation, development and enlargement, and which it seeks to advance in the wider world: democracy, the rule of law, the universality and indivisibility of human rights and fundamental freedoms, respect for human dignity, the principles of equality and solidarity, and respect for the principles of the United Nations Charter and international law.
   The Union shall seek to develop relations and build partnerships with third countries, and international, regional or global organisations which share the principles referred to in the first subparagraph. It shall promote multilateral solutions to common problems, in particular in the framework of the United Nations.
2. The Union shall define and pursue common policies and actions, and shall work for a high degree of cooperation in all fields of international relations, in order to:
   (a) safeguard its values, fundamental interests, security, independence and integrity;
   (b) consolidate and support democracy, the rule of law, human rights and the principles of international law;
   (c) preserve peace, prevent conflicts and strengthen international security, in accordance with the purposes and principles of the United Nations Charter, with the principles of the Helsinki Final Act and with the aims of the Charter of Paris, including those relating to external borders;
   (d) foster the sustainable economic, social and environmental development of developing countries, with the primary aim of eradicating poverty;
   (e) encourage the integration of all countries into the world economy, including through the progressive abolition of restrictions on international trade;
   (f) help develop international measures to preserve and improve the quality of the environment and the sustainable management of global natural resources, in order to ensure sustainable development;
   (g) assist populations, countries and regions confronting natural or man-made disasters; and
   (h) promote an international system based on stronger multilateral cooperation and good global governance.
3. The Union shall respect the principles and pursue the objectives set out in paragraphs 1 and 2 in the development and implementation of the different areas of the Union's external action covered by this Title and by Part Five of the Treaty on the Functioning of the European Union, and of the external aspects of its other policies.
   The Union shall ensure consistency between the different areas of its external action and between these and its other policies. The Council and the Commission, assisted by the High Representative of the Union for Foreign Affairs and Security Policy, shall ensure that consistency and shall cooperate to that effect" (TEU, 2012).

the wider world: democracy, the rule of law, the universality and indivisibility of human rights and fundamental freedoms, respect for human dignity, the principles of equality and solidarity, and respect for the principles of the United Nations Charter and international law" (TEU, 2012).

While there have been cases in which the European Court of Justice (CJEU) has emphasized the autonomy of European Union law with respect to international law, the binding nature of international law, but also of customary international law and secondary international law "deriving from international agreements such as Association Council decisions", is generally inferred from Art. 216(2)[6] of the Treaty on the Functioning of the European Union (TFEU) which states that international agreements concluded by the Union are binding both on its institutions and Member States (Wessel, 2011, p. 4). Furthermore, the Union law's interaction with international law is framed by the principles of conferral (Artt. 5(1), (2)), subsidiarity (Art. 5(3)), and proportionality (Art. 5(4)) set out in Article 5[7] of the TEU which defines "the limits of Union

[6] Article 216 of the TFEU recites:

"1. The Union may conclude an agreement with one or more third countries or international organisations where the Treaties so provide or where the conclusion of an agreement is necessary in order to achieve, within the framework of the Union's policies, one of the objectives referred to in the Treaties, or is provided for in a legally binding Union act or is likely to affect common rules or alter their scope.
2. Agreements concluded by the Union are binding upon the institutions of the Union and on its Member States" (TFEU, 2012).

[7] Article 5 of the TEU reads:

1. "The limits of Union competences are governed by the principle of conferral. The use of Union competences is governed by the principles of subsidiarity and proportionality. 2
2. Under the principle of conferral, the Union shall act only within the limits of the competences conferred upon it by the Member States in the Treaties to attain the objectives set out therein. Competences not conferred upon the Union in the Treaties remain with the Member States.
3. Under the principle of subsidiarity, in areas which do not fall within its exclusive competence, the Union shall act only if and in so far as the objectives of the proposed action cannot be sufficiently achieved by the Member States, either at central level or at regional and local level, but can rather, by reason of the scale or effects of the proposed action, be better achieved at Union level. The institutions of the Union shall apply the principle of subsidiarity as laid down in the Protocol on the application of the principles of subsidiarity and proportionality. National Parliaments ensure compliance with the principle of subsidiarity in accordance with the procedure set out in that Protocol.
4. Under the principle of proportionality, the content and form of Union action shall not exceed what is necessary to achieve the objectives of the Treaties. The institutions of the Union shall apply the principle

competences" (TEU, 2012). According to the principle of conferral, "the Union shall act only within the limits of the competences conferred upon it by the Member States" (Art 5(2)), while the principle of subsidiarity ensures that

> "in areas which do not fall within its exclusive competence, the Union shall act only if and in so far as the objectives of the proposed action cannot be sufficiently achieved by the Member States, either at central level or at regional and local level, but can rather, by reason of the scale or effects of the proposed action, be better achieved at Union level" (Art. 5(3) TEU, 2012).

Lastly, the principle of proportionality requires that "the content and form of Union action shall not exceed what is necessary to achieve the objectives of the Treaties" (Art. 5(4) TEU, 2012). All these principles are pivotal not only in guiding the Union's interpretation and implementation of international law within its borders but also in safeguarding the constitutional balance between the Union and its Member States. The principle of subsidiarity[8], in particular, has been conceived as an instrument to prevent an "over-centralization of powers in Europe" and, since its creation in 1992, during which it was acclaimed as the "word that saved Maastricht", it has "remained a cornerstone of the EU legal architecture" (Fabbrini, 2016, p. 6).

---

of proportionality as laid down in the Protocol on the application of the principles of subsidiarity and proportionality (TEU, 2012).

[8] The principle of subsidiarity has both a material and a procedural dimensions: "Material subsidiarity involves a number of conditions of a substantive character which assist […] the [EU] institutions to answer the question of whether the [EU] should act or not, that is: whether the [EU] intervention is appropriate. Procedural subsidiarity involves the establishment of a number of procedural conditions which the [EU] must fulfil before implementing subsidiarity" (Estella, 1997 as cited in Fabbrini, 2016, pp. 7-8). To fulfill the material dimension of the principle of subsidiarity, both the "national insufficiency test" and the "comparative efficiency test" delineated in Art. 5(3) TEU must be satisfied cumulatively, when the Union acts in fields such as the environment, public health, transport, education, and social policy, and especially where transnational issues or market distortions are involved (Fabbrini, 2016, p. 8). Additionally, the procedural dimension, developed in the 1997 Amsterdam Protocol, obliges EU institutions to justify compliance with the principle through stakeholder consultation, explanatory memoranda, and, where possible, qualitative or quantitative indicators "at an early stage of the legislative process" (Fabbrini, 2016, p. 9).

For what concerns the relationship between the ECHR and the EU Charter, it is important to underline that EU Member States are bound by both legal instruments: all are indeed contracting parties not only of the EU Charter and of the European Union, but also to the ECHR and of the Council of Europe (CoE), a broader international organization founded in 1949 and composed of 46 Member States aimed at promoting human rights, democracy, and rule of law through the implementation of the ECHR, its only binding treaty enforced by the European Court of Human Rights (ECtHR) and by the CJEU when its application comes within the framework of Union Law (Lenaerts & Smijter, 2001, p. 91). However, when EU law is applied, the EU Charter prevails over the ECHR; when, instead, EU law is not applied, for instance when an EU Member States acts in an area outside EU competence, then the ECHR prevails in protecting fundamental rights; finally, when both the EU Charter and the ECHR apply, Art. 52(3)[9] of the EU Charter states that the "meaning and scope of (the Charter) … rights shall be the same as those laid down by the (corresponding provisions of the) Convention." as also established by the ECtHR, even if the CJEU is not bound by the ECtHR case law, in an effort to maintain a coherent legal framework. Henceforth, when it comes to the right to freedom of speech, the EU citizens benefit from the protection not only of their national constitutions but also of the overlapping legal frameworks of the ECHR, the EU Charter, and of the relevant international treaties.

1.1.1 A Qualified Right and its Justifiable Restrictions

Enshrined in various international and regional legal instruments, the right to freedom of speech is essential to protect public discourse and to guarantee democratic participation and

[9] Article 52(3) of the EU Charter recites:

> "In so far as this Charter contains rights which correspond to rights guaranteed by the Convention for the Protection of Human Rights and Fundamental Freedoms, the meaning and scope of those rights shall be the same as those laid down by the said Convention. This provision shall not prevent Union law providing more extensive protection" (EU Charter, 2000).

individual autonomy. However, this right is not an absolute one, but rather a qualified one. An absolute right is indeed one which "cannot be limited or infringed under any circumstances, not even during a declared state of emergency", while a qualified one "permit(s) interferences subjects" to some limitations (European Commission, n.d.a). Common absolute rights under international and European laws are the right to life (Art. 6 ICCPR; Art. 2 ECHR), the prohibition of torture or inhuman/degrading treatment (Art. 7 ICCPR; Art. 3 ECHR), the prohibition of slavery and servitude (Art. 8 ICCPR; Art. 4 ECHR) and the principle of *nullum crimes sine lege* (Art. 15 ICCPR; Art. 7 ECHR). These absolute rights are listed under Art. 4(2) of the ICCPR, which also considers the prohibition of imprisonment for debt (Art. 11 ICCPR), the right of recognition of a person before the law (Art. 16 ICCPR), and the right to freedom of thought, conscience and religion (Art. 18 ICCPR) as absolute rights, and under Art. 15(2) of the ECHR. Such rights are formulated, both under international and European law as negative injunctions ("No one shall be …"; "depriving…"), clearly "identifying some action that is prohibited, an injunction against a deed, a duty not to perform a given act" (Webber, 2016, p. 2). Instead, qualified or relative rights are typically phrased as entitlements to abstract goods ("Everyone has the right to…"), and "no action or inaction by anyone or any groups of persons is highlighted as giving content to the rights that each and everyone has" (Webber, 2016, p. 2). These rights thus require interpretation and balancing. This means that, under some circumstances and within specified limits, qualified rights, among which the right to freedom of expression, may be restricted (Bromell, 2021, p. 135). Generally, restrictive measures aim to defend either public welfare and interest, or the rights of others, provided that such restrictions have a legal basis, a legitimate justification, and constitute the least necessary restrictive limitation of the right: every restriction must henceforth be "lawful, necessary and proportionate" (Bromell, 2021, p.135).

Art. 19(3) of the ICCPR outlines all the possible restrictions of the right to freedom of speech, reciting:

> "3. The exercise of the rights provided for in paragraph 2 of this article carries with it special duties and responsibilities. It may therefore be subject to certain restrictions, but these shall only be such as are provided by law and are necessary:
>
> (a) For respect of the rights or reputations of others;
>
> (b) For the protection of national security or of public order (ordre public), or of public health or morals" (ICCPR, 1966).

In Europe, then, there are fundamental legal instruments which recognize and protect the right, underlying also the cases in which restrictions of the right might be allowed. The first is the ECHR of the CoE which is legally binding on all the 46 CoE Member States; the second is the EU Charter which applies only to EU institutions and Member States. The ECHR dedicates Art. 10 to the right to freedom of expression, specifying in paragraph 2 that the right

> "may be subject to such formalities, conditions, restrictions or penalties as are prescribed by law and are necessary in a democratic society, in the interests of national security, territorial integrity or public safety, for the prevention of disorder or crime, for the protection of health or morals, for the protection of the reputation or rights of others, for preventing the disclosure of information received in confidence, or for maintaining the authority and impartiality of the judiciary" (ECHR, 1950).

Furthermore, Art. 15[10] of the ECHR allows states the possibility to temporarily derogate from certain rights, among which the right to freedom of speech, in times of emergency "to the extent

[10] In its entirety, Article 15 of the ECHR reads:

> "1. In time of war or other public emergency threatening the life of the nation any High Contracting Party may take measures derogating from its obligations under this Convention to the extent strictly required by

strictly required by the exigencies of the situation" (ECHR, 1950). However, this provision set a high threshold for such derogations, requiring that (1) a public emergency threatening the life of the nation exist, (2) the measures adopted are strictly necessary and proportionate to deal with the emergency, and that (3) every taken measure must be reported and justified to the Secretary General of the Council of Europe (Article 15, ECHR, 1950).

The EU Charter instead recognizes the right to freedom of expression in Art. 11, specifying in Art. 52 that

> "1. Any limitation on the exercise of the rights and freedoms recognized by this Charter must be provided for by law and respect the essence of those rights and freedoms. Subject to the principle of proportionality, limitations may be made only if they are necessary and genuinely meet objectives of general interest recognized by the Union or the need to protect the rights and freedoms of others" (EU Charter, 2000).

Clearly, the harm principle first introduced in *On Liberty* by Mills (1859) has developed overtime, but the basic tenet remains alive: speech may be restricted only when it causes direct and significant harm to others.

1.1.2 The Right to Freedom of Speech in the Digital Age

The 21st century ushered in a new revolution which, much like the Enlightenment in the 17th and 18th centuries, deeply transformed the entire society, revolutionizing not only the way of

---

the exigencies of the situation, provided that such measures are not inconsistent with its other obligations under international law.
2. No derogation from Article 2, except in respect of deaths resulting from lawful acts of war, or from Articles 3, 4 (paragraph 1) and 7 shall be made under this provision.
3. Any High Contracting Party availing itself of this right of derogation shall keep the Secretary General of the Council of Europe fully informed of the measures which it has taken and the reasons therefor. It shall also inform the Secretary General of the Council of Europe when such measures have ceased to operate and the provisions of the Convention are again being fully executed" (ECHR, 1950, pp. 13-14).

living of most but also the very nature of the right to freedom of speech. What was once confined to printed newspapers and public manifestations now unfolds into vast, and most importantly privately owned, new spaces: the digital platforms and the internet. The digital revolution has amplified voices in an unprecedented manner, facilitating mass communication and allowing for the rapid spread of ideas in all the places connected to the network. At the same time, however, it has enabled unparalleled surveillance, manipulation and the commodification of speech, for the digital realm is controlled less by democratic deliberation than by corporate interests and algorithmic structures. The emergence of the internet and of the digital world, the *opium gentium* of the 21st century, has indeed radically transformed how speech is produced, disseminated, and regulated.

Moreover, as the world embraced the digital revolution, the modern conception of the self shifted, challenging the classical philosophical foundations of the concepts of the self and of the right to freedom of speech. Feldman (2017, pp. 1163-1171) argues that, since the self can now be understood as "an emergent self" which is "socially constructed … (and) fundamentally relational", freedom of expression has transformed as well, becoming central to the emergence of the modern self: it can, in fact, foster and protect the creativity and the dynamic process behind the self's emergence in both the physical and the digital spaces. However, while the physical world is dominated by laws and customs which are mostly clear and defined, the digital one remains a realm of its own. Since traditional legal frameworks are structured on the concepts of state power and public spaces, they seldom easily adapt to the evolving landscape of the digital world, which is a hybrid domain where private corporations wield unprecedented influence over the conditions and regulations of speech. In this context which transcends national borders and law, it becomes more difficult to understand what are and who imposes the limitations on the right to freedom of

speech, but it appears evident that such restrictions are less visible, less contestable and less accountable than those imposed by democratic states in the physical world. In light of this, a new philosophical debate has arisen about the role and responsibility of internet platforms: in particular, scholars question whether online platforms "have moral duties to respect the free speech of their users" or to moderate "harmful speech", "and finally, (whether) … those duties should be legally enforced" (Howard, 2024).

While some argue that platforms, as private entities, enjoy editorial freedom akin to publishers, others contend that they function as public forums and that their "duties to respect speech are (roughly) identical to the duties of states" (Kramer, 2021 as cited in Howard, 2024). Instead, with regard to the removal of harmful content, most scholars and legislators recognize the need for private platforms to actuate some forms of online content regulation, deeming them "morally responsible for removing extremist content" (Howard, 2024).

The digitalization of the right to freedom of speech has thus changed the traditional approaches to the right, obliging philosophers, legal scholars, and policy- and law- makers to reinvent the normative foundations of freedom of expression in a world where speech is both enabled and constrained by digital online platforms (Howard, 2024).

1.2 Counterterrorism and Anti-Incitement to Terrorism Online: Why is it Needed?

In the online realm, where the laws are difficult to apply and which was initially strongly underregulated, terrorist organizations have often found a fertile ground to disseminate their propaganda, to spread terror by posting videos or messages about their actions, and to accrue their ranks by including new members attracted by their propaganda.

The Christchurch attacks on March 15, 2019 is an example of the use of the internet to spread terrorist content: 36 minutes livestreamed on Facebook allowed an Australian far-right extremist, Brenton Tarrant, to shoot 51 people in 2 mosques in Christchurch, New Zealand, filming its act with a GoPro to simulate a "first-person shooter" experience, effectively gamifying mass murder (Macklin, 2019, p. 18). While less than 200 people followed the attack live, the video was not initially reported to Facebook and before being removed it was watched circa 4000 times; furthermore, the Christchurch video went viral on YouTube and was heavily commented on platforms such as 8chan and Reddit before the moderation systems of these platforms could react, thus revealing severe shortcomings in their automated content control systems (Macklin, 2019, p. 20). Facebook took down "about 1.5 million videos of the attack globally within the first 24 hours, blocking 1.2 million of these attempts automatically at the point of upload and thereby preventing its viewing"; YouTube, on its part, disabled many search functions for a while as the platform was initially overpowered by users repackaging "footage of the killings in … more than 800 visually distinct versions of the video, (with) … tens of thousands of such videos uploaded … at a rate of one per second, in the hours immediately after the shootings" (Macklin, 2019, p. 20). Tarrant's digital performance of his attack effectively adapted a communication strategy previously associated with jihadi terrorism to far-right terrorism, succeeding where Anders Behring Breivik had failed during his 2011 Utøya attack because he could not afford an Iphone with which the attack could be livestreamed on YouTube (Macklin, 2019, p. 19).

This tactic of livestreaming terrorist attacks aims at reinforcing the spectacle and the shareability of violence to amplify the message of the act beyond the attack itself and its victims. By transforming acts of terror into horrific global spectacles, such attacks not only instill fear in

those directly affected and those who share key traits with the victims (such as nationality, religion, political opinion, etc.) but also serve as a source of inspiration for potential attackers worldwide, and especially for Lone-Wolf terrorists, encouraging further acts of violence. Jihadist groups were the first to understand the potential of live streaming attacks online, and indeed this tactic came to be known as "selfie jihad": during the 2016 Magnanville terrorist attack in France, a jihadi killed two police officers in their home, livestreaming the "aftermath of the attacks", "while several other jihadi attackers, notably Mohamed Merah, Mehdi Nemmouche, and Amedy Coulibaly, also sought to film their crimes" (Macklin, 2019, p. 19).

Furthermore, the Islamic State has been using social media such as X (previously Twitter), Youtube, Telegram, or Tik Tok to disseminate propaganda, radicalize users, and coordinate attacks. The Counter Extremism Project (CEP) found a "a pro-ISIS account on TikTok that posted instructions for making explosives and advice for committing attacks with knives" whose videos received more than 10,000 view in only one week; it also discovered "a pro-ISIS Telegram bot that shared posts encouraging acts of lone-actor terrorism and shared bomb-making information" with guides made by the pro-ISIS online group, Al-Saqri Foundation (CEP, 2025). The online dissemination of an Islamist utopia has also been used to target women by the Islamic State, in an effort to indoctrinate them in pushing their husbands to embrace the Jihadi ideology and join the cause. The CEP has also identified a Spanish Neo-Nazi accelerationist group, called The Base, on Telegram, which encouraged "the creation of militant cells to create all white separatist mountainous strongholds" and invoked "the great replacement theory" (CEP, 2025).

These events represent just a few instances of extremist groups exploiting the digital realm to spread their message and accrue their ranks, in an effort to amplify the resonance of their cause. Naturally, such incidents have sparked a pressing debate at the international level about tech

companies' responsibilities in monitoring and removing harmful content online, prompting states and international organizations to try and regulate terrorism online.

1.2.1 International and European Responses to Online Terrorism and Incitement to Terrorism

The proliferation of terrorist content online has led to an increasing focus on counterterrorism policies in the digital sphere. At the international level, the UN Security Council (UNSC) has been incredibly vociferous in this regard, particularly after the 2005 terrorist attacks in London: after only 2 months, in fact, it adopted Resolution 1624 (2005)[11], calling on states "to prohibit by law incitement to commit a terrorist act or acts" (Tao, 2023, p. 41). Since 1999 with the enactment of Resolution 1267 (1999)[12], the Council has adopted different resolutions on

[11] Resolution 1624 (2005) builds on previous Resolutions enacted by the Security Council, such as Resolutions 1267 (1999), 1373 (2001), 1535 (2004), 1540 (2004), 1566(2004) and 1617 (2004), to fight against "threats to international peace and security caused by acts of terrorism", in particular: it "1. calls upon all states to adopt such measures as may be necessary and appropriate … to … (a) Prohibit by law incitement to commit a terrorist act or acts; (b) Prevent such conduct; (c) Deny safe haven" to alleged terrorists; it pushes them to "2. cooperate ... to strengthen the security of their international borders" and "3. to enhance dialogue and broaden understanding among civilizations, in an effort to prevent the indiscriminate targeting of different religions and cultures"; and it "6. directs the Counter-Terrorism Committee to: (a) Include in its dialogue with Member States their efforts to implement this resolution; (b) Work with Member States to help build capacity, including through spreading best legal practice and promoting exchange of information in this regard; (c) Report back to the Council in twelve months on the implementation of this resolution" (United Nations Security Council, 2005, pp. 1-3).

[12] Resolution 1267 (1999) was enacted to fight the threats posed by the Taliban which were providing "a safe haven to Usama Bin Laden (and were allowing) him and others associated with him to operate a network of terrorist training camps from Taliban-controlled territory and to use Afghanistan as a base from which to sponsor international terrorist operations" (United Nations Security Council, 1999, p. 1). It obliges the Taliban to comply with international law, stopping the training of international terrorists and turning Usama Bin Laden (1. – 2.); and then it pushes states to "(a) Deny permission for any aircraft to take off from or land in their territory if it is owned, leased or operated by or on behalf of the Taliban …; (and) (b) Freeze funds and other financial resources, including funds derived or generated from property owned or controlled directly or indirectly by the Taliban" (4.) (United Nations Security Council, 1999, p. 2). Furthermore, it establishes a "committee of the Security Council consisting of (all its) members:

> "(a) To seek from all States further information regarding the action taken by them with a view to effectively implementing the measures imposed by paragraph 4 above;
> (b) To consider information brought to its attention by States concerning violations of the measures imposed by paragraph 4 above and to recommend appropriate measures in response thereto;
> (c) To make periodic reports to the Council on the impact, including the humanitarian implications, of the measures imposed by paragraph 4 above;

counterterrorism, with Res. 2129 (2013)[13] and Res. 2617 (2021)[14] addressing specifically the use of the internet and "other emerging technologies for terrorist purposes" (United Nations, n.d.). Moreover, the UN Counter-Terrorism Committee (CTC) has been active in "monitor(ing), facilitat(ing) and prompt(ing) Member States' implementation" of the UNSC Resolutions working alongside many international organizations active in the field of counterterrorism, namely the Tech Against Terrorism (TAT), the Global Internet Forum to Counter Terrorism (GIFCT), and the Global Initiative on Handling Electronic Evidence (United Nations, n.d.). All these international bodies have been extensively working to protect critical infrastructures, enhance data protection

---

(d) To make periodic reports to the Council on information submitted to it regarding alleged violations of the measures imposed by paragraph 4 above, identifying where possible persons or entities reported to be engaged in such violations;
(e) To designate the aircraft and funds or other financial resources referred to in paragraph 4 above in order to facilitate the implementation of the measures imposed by that paragraph;
(f) To consider requests for exemptions from the measures imposed by paragraph 4 above as provided in that paragraph, and to decide on the granting of an exemption to these measures in respect of the payment by the International Air Transport Association (IATA) to the aeronautical authority of Afghanistan on behalf of international airlines for air traffic control services" (United Nations Security Council, 1999, p. 3).

[13] Resolution 2129 (2013) underscores the importance of the work carried out by the CTC and the CTED which shall continue (1. -5.), encouraging them "to cooperate with member states and regional and subregional organizations, upon request, to assess and advise them on formulating national and regional counterterrorism strategies to further the implementation of resolutions 1373 (2001) and 1624 (2005) (7.) (United Nations Security Council, 2013, pp. 5-6). Furthermore, it "14. Notes the evolving nexus between terrorism and information and communications technologies, in particular the Internet, and the use of such technologies to commit terrorist acts, and to facilitate such acts through their use to incite, recruit, fund, or plan terrorist acts, and directs CTED to continue to address this issue, in consultation with Member States, international, regional and subregional organizations, the private sector and civil society and to advise the CTC on further approaches" (United Nations Security Council, 2013, p. 7).

[14] Resolution 2617 (2021), recognizing the need to strengthen the CTED, the CTC, the INTERPOL, the UNODC and all the relevant UN bodies' mandate to fight against terrorism and incitement to terrorism, it recalls the Christchurch Call to Action on preventing the use of the internet for terrorism-related purposes (United Nations Security Council, 2021, p. 6). It thus "34. Recognizes CTED's work on countering use of the internet, other information and communications technology (ICTs), and other emerging technologies for terrorist purposes, while respecting human rights and fundamental freedoms, and taking into account Member State compliance with applicable obligations under international law, and taking note of the need to preserve global connectivity and the free and secure flow of information facilitating economic development, communication, participation and access to information, and stresses the importance of cooperation with civil society and the private sector in this endeavor" (United Nations Security Council, 2021. p. 12). Furthermore, it encourages continued cooperation among the CTED, IACO all the other relevant Un organizations to work "together on identifying gaps and vulnerabilities relevant to counterterrorism and aviation security, promoting the work and tools of each agency, and coordinating closely on CTED assessments and the development of recommendations" (38. – 39.); hence, it encourages such organizations to assist member states in the development of their counterterrorism strategies (40.) (United Nations Security Council, 2021, pp. 12-13).

and privacy, and augment online investigative capabilities in the field of counterterrorism: Resolution 2341 (2017)[15] of the UNSC allows the CTC, in conjunction with the UN Counter-Terrorism Committee Executive Directorate (CTED), "to examine Member States' efforts to protect critical infrastructure from terrorist attack … with the aim of identifying good practices, gaps and vulnerabilities in this field"; concurrently, the Working Group on Criminal Justice, Legal Responses and Countering the Financing of Terrorism of the Global Counter-Terrorism Coordination Compact launched a data protection initiative to support international cooperation through model legislation and best practices on privacy in counterterrorism which came to be known as the "UN CTC Programme on Data Protection"; furthermore, to enhance investigative capabilities, the CTED, the United Nations Office of Counter-Terrorism (UNOCT), the International Criminal Police Organization (INTERPOL), and the United Nations Office on Drugs and Crime (UNODC) have introduced "capacity-building programs" to strengthen states' technical skills in conducting "open-source and dark web investigations" (United Nations, n.d.). All these

[15] Resolution 2341 (2017) pushes all member states to: (1.) "make concerted and coordinated efforts, including through international cooperation, to raise awareness , to expand knowledge and understanding of the challenges posed by terrorist attacks; (2.) develop or improve "their strategies for reducing risks to critical infrastructure from terrorist attacks"; and (5.) "strengthen national, regional and international partnership with stakeholders, both public and private, as appropriate, to share information and experience in order to prevent, protect, mitigate, investigate, respond to and recover from damage from terrorist attacks" (United Nations Security Council, 2017, pp. 3-4). Furthermore, it:

> 10. Directs the CTC, with the support of the Counter-Terrorism Executive Directorate (CTED) to continue as appropriate, within their respective mandates, to examine Member States efforts to protect critical infrastructure from terrorist attacks as relevant to the implementation of resolution 1373 (2001) with the aim of identifying good practices, gaps and vulnerabilities in this field;
> 11. Encourages in this regard the CTC, with the support of CTED, as well as the CTITF to continue working together to facilitate technical assistance and capacity building and to raise awareness in the field of protection of critical infrastructure from terrorist attacks, in particular by strengthening its dialogue with States and relevant international, regional and subregional organizations and S/RES/2341 (2017) 17-02174 5/5 working closely, including by sharing information, with relevant bilateral and multilateral technical assistance providers; (and)
> 12. Encourages the CTITF Working Group on the Protection of Critical Infrastructure including Vulnerable Targets, Internet and Tourism Security to continue its facilitation, and in cooperation with other specialized United Nations agencies, assistance on capacity-building for enhancing implementation of the measures upon request by Member States" (United Nations Security Council, 2017, pp. 4-5).

efforts signal the great attention that the international community is posing on matters of combating terrorism, considered a major, if not the greatest, public security threat worldwide.

In Europe, then, the European Union has been at the forefront in trying to regulate this important issue, especially by adopting Regulation (EU) 2021/784 of the European Parliament and the Council on 29 April 2021 (on addressing the dissemination of terrorist content online) (TCO), with the aim of "ensur(ing) the smooth functioning of the digital single market in an open and democratic society, by addressing the misuse of hosting services for terrorist purposes and contributing to public security across the Union" (Regulation, 2021). The TCO Regulation, in force since June 7, 2022, obliges all hosting service providers (HSPs) working within the European single market to remove terrorist content within one hour of receiving a removal order from a competent national authority, including "judicial, administrative, or law enforcement bodies appointed by EU Member States", with penalties reaching up to 4% of global turnover for noncompliance (TATE, n.d.). Obviously, the Regulation provides a clear definition of what are deemed to be 'terrorist offences[16]' and 'terrorist content,[17]' the latter of which is broadly defined

[16] Regulation (EU) 2021/784 recalls Art. 3 of Directive (EU) 2017/541 to define what are 'terrorist offences' under European Law. Art. 3(1) lists the following 'terrorist offences': "(a) attacks upon a person's life which may cause death; (b) attacks upon the physical integrity of a person; © kidnapping or hostage-taking;(d) causing extensive destruction to a government or public facility, a transport system, an infrastructure facility, including an information system, a fixed platform located on the continental shelf, a public place or private property likely to endanger human life or result in major economic loss; (e) seizure of aircraft, ships or other means of public or goods transport; (f) manufacture, possession, acquisition, transport, supply or use of explosives or weapons, including chemical, biological, radiological or nuclear weapons, as well as research into, and development of, chemical, biological, radiological or nuclear weapons; (g) release of dangerous substances, or causing fires, floods or explosions, the effect of which is to endanger human life; (h) interfering with or disrupting the supply of water, power or any other fundamental natural resource, the effect of which is to endanger human life; … (j) threatening to commit any of the acts listed in points (a) to (i)" (Directive, 2017).

[17] To define 'terrorist content,' Art. 3(7) of Regulation (EU) 2021/784 lists all the material that: (a) incites the commission of one of the offences referred to in points (a) to (i) of Article 3(1) of Directive (EU) 2017/541, where such material, directly or indirectly, such as by the glorification of terrorist acts, advocates the commission of terrorist offences, thereby causing a danger that one or more such offences may be committed; (b) solicits a person or a group of persons to commit or contribute to the commission of one of the offences referred to in points (a) to (i) of Article 3(1) of Directive (EU) 2017/541; © solicits a person or a group of persons to participate in the activities of a terrorist group, within the meaning of point (b) of Article 4 of Directive (EU) 2017/541; (d) provides instruction on the making or use of explosives, firearms or other weapons or noxious or hazardous substances, or on other specific methods or

to include incitement, glorification, instructions for committing attacks, and solicitation to terrorist groups (Regulation (EU) 2021/784, 2021). Of course, while setting uniform rules to detect and remove terrorist content online, the TCO strives to prioritize security imperatives without forgetting to uphold the protection of fundamental rights, particularly the right to freedom of speech. Indeed, with regard to removal orders, the competent national authorities must provide a statement of reasons (Art. 3(4)(a)(b)(c)(d)(e)(f)(g)) which the interested platforms may even legally challenge "before the courts of the Member State of the competent authority" (Art. 9(1)) (Regulation, 2021). Furthermore, Articles 7 and 8 impose transparency obligations on both the competent authorities and the HSPs, which must deliver annual transparency reports especially where they had to implement measures for taking down terrorist content. Obviously, all these safeguards have been imposed to ensure respect for fundamental rights of users, and particularly of their right to freedom of expression, when both analyzing and removing content dealing with terrorism (Art. 5(1); Art. 5(3)(c)) (Regulation, 2021). Other previous legislative measures that the European Union has enacted in the field of counterterrorism online are: the Parliament and Council's 2002 *Directive on Combating Terrorism* aimed at "harmonizing (Member States' national) criminal laws", "allowing for prosecution of cyber-terrorism"; while the *Directive on Countering Money Laundering By Criminal Law* alongside the *Regulation on Controls on Cash Entering or Leaving the Union* and the *Regulation on the Mutual Recognition of Freezing and Confiscation Orders* strive to combat terrorism financing and money laundering, even in the virtual space (Bąkowski, 2023, pp. 4-6). The 2018 *Network and Information Security Directive* (NIS), then, has served to accrue the EU's cybersecurity, obliging server providers to report any

techniques for the purpose of committing or contributing to the commission of one of the terrorist offences referred to in points (a) to (i) of Article 3(1) of Directive (EU) 2017/541; (e) constitutes a threat to commit one of the offences referred to in points (a) to (i) of Article 3(1) of Directive (EU) 2017/541 (Regulation (EU) 2021/784, 2021).

kind of cyber incident to competent national authorities (Bąkowski, 2023, p. 6). Moreover, the EU has given more responsibilities in this field to the European Union Agency for Law Enforcement Cooperation (EUROPOL), allowing it to create the European Counter Terrorism Center (ECTC) which "provides strategic and operational support and is also in charge of the Internet Referral Unit (IRU), tackling online terrorist propaganda, and the Terrorist Finance Tracking Programme (TFTP)" (Bąkowski, 2023, p.6).

Evidently, both at the international and at the European level, the attention that is being given to combating terrorism and incitement to terrorism online is only growing in scope, in an effort to ensure the security of the digital space and of all its users. These efforts, however, have sometimes met some resistance by civil society, because they risk infringing fundamental rights and especially the right to freedom of speech. The latter, which represents one of the digital world's core imperatives, is indeed very hard to reconcile with security imperatives when the content to be analyzed and possibly removed deals with terrorism and incitement to terrorism.

1.3 How to Balance Security and Freedom of Speech in Online Counterterrorism?

The greatest dilemma of online counterterrorism is certainly striving to enforce measures which strike a fair balance between security imperatives and the protection of fundamental rights. While it is fundamental to prevent incitement to terrorism and to ensure security, in fact, it is also important that the right to freedom of speech is protected: most often than not, there is only a thin line between removing harmful content and stifling legitimate dissent or minority opinions. The problem lies mainly in the overreach of moderation systems, which are often automatic and lack the contextual understanding to differentiate between incitement to terrorism, satire, or critical commentary: thus, they may mistakenly remove and censor legitimate content. Another problem

arises when automatic moderation systems have to deal with content "which may not go so far as to incite or promote the commission of terrorist acts, but might nevertheless applaud past acts", or apologia (High Commissioner, n.d., p. 43). The glorification or apologia of terrorism must not, indeed, be used to justify limitations on the right to freedom of speech in so far as, some experts on freedom of expression clarify, "incitement should be understood as a direct call to engage in terrorism, with the intention that this should promote terrorism, and in a context in which the call is directly causally responsible for increasing the actual likelihood of a terrorist act occurring" (High Commissioner, n.d., p. 43). The line between apologia and incitement to terrorism is very blurred, most of the time, and it can be safe to say that automated systems are not always capable of distinguishing between the two.

The right balance, as the saying goes, should be in the middle with automated systems capable of contextual understanding and of removing only illicit and harmful content supervised by a human moderator who strives to protect security without impinging upon fundamental rights. Unfortunately, however, the "existing international law has not drawn a clear balancing line" between security and freedom of speech in the field of counterterrorism online, leaving the issue open for legal debate under national or regional jurisdiction (Tao, 2023, p. 84). Since 9/11, many countries have preferred to take a securitized approach to the issue, clearly bending the line in favor of security imperatives and creating, as UN Special Rapporteur Fionnuala Ní Aoláin underlined, an "increasinging tendency to neglect fundamental rights and freedoms when addressing terrorism issues and "a severe impact on the integrity of the rule of law and governance" (OHCHR, 2021). To ensure a better balance, she explains:

> "We are in dire need of an overhaul of capacity-building and technical assistance so that efforts to counter terrorism and violent extremism are not only effective but also respect

human rights and are transparent and accountable. … (Henceforth,) the growth in counter-terrorism institutions, frameworks and programming must be matched by meticulous attention to human rights and rule of law, and the implementation of appropriate monitoring and oversight strategies" (OHCHR, 2021).

1.3.1 An European Perspective on this Balance

In Europe, the Council of Europe's Convention on the Prevention of Terrorism obliges its States parties "to criminalize the unlawful and intentional public provocation to commit a terrorist offence", defining this as "…the distribution, or otherwise making available, of a message to the public, with the intent to incite the commission of a terrorist offence, where such conduct, whether or not directly advocating terrorist offences, causes a danger that one or more such offences may be committed" (art. 5 (1))" (High Commissioner, n.d., p. 42). In order to balance security imperatives with the protection of fundamental rights in this field, the ECtHR leads by example, "draw(ing) the balancing line (based on) … a three-part test under Art. 10(2) of the ECHR" claiming that restrictions on the right to freedom of speech should be: (1) "prescribed by law", (2) for the achievement of … legitimate aims" and (3) "necessary in a democratic society" (Tao, 2023, p. 85).

The ECtHR has defined what "prescribed by law" means in its 2005 case *Leyla Şahin v. Turkey*[18], explaining that this first requirement concerns the "substantive sense, (and) not (the)

[18] The case of *Leyla Şahin v. Turkey* (2005) originated from an application made by Leyla Şahin, on 21 July 1998, to the European Commission of Human Rights against the Republic of Turkey, alleging violations of her rights under Articles 8, 9, 10, and 14 of the ECHR and Article 2 of the Protocol No. 1, "by regulations on wearing the Islamic headscarf in institutions of higher education (Leyla Şahin v. Turkey, 2005). A 1998 directive from Istanbul University had indeed prohibited students with headscarves or long beards from participating in lectures and exams (Şahin v. Turkey, 2005). The Court found no violations of any rights, as the restrictions made by the Turkish government were "foreseeable to hose concerned and pursued the legitimate aims of protecting the rights and freedoms of others and maintaining public order" (158.) (Leyla Şahin v. Turkey, 2005). Sections 98. And 99. Indeed read:

formal one" of any law, which includes "both written law, encompassing enactments of lower ranking statutes and regulatory measures taken by professional regulatory bodies under independent rule-making powers delegated to them by Parliament, and unwritten law; furthermore, in *Sanoma Uitgevers B.V. v. the Netherlands* (2010)[19], the Court has declared that "the law should be both adequately accessible and foreseeable" for individuals to regulate their conduct (Tao, 2023, p. 85). In cases dealing with terrorism and incitement to terrorism, determining whether national restrictions on freedom of speech "are "prescribed by law" is uncontroversial and does not raise any particular problems" (Tao, 2023, p. 85).

The second requirement, in accordance with Article 10(2) of the ECHR, considers as legitimate aims the "protection of national security, territorial integrity or public safety, the prevention of disorder or crime, the protection of health or morals, the protection of the reputation or rights of others, preventing the disclosure of information received in confidence, or maintaining

---

"98. In these circumstances, the Court finds that there was a legal basis for the interference in Turkish law, namely transitional section 17 of Law no. 2547 read in the light of the relevant case-law of the domestic courts. The law was also accessible and can be considered sufficiently precise in its terms to satisfy the requirement of foreseeability. It would have been clear to the applicant, from the moment she entered Istanbul University, that there were restrictions on wearing the Islamic headscarf on the university premises and, from 23 February 1998, that she was liable to be refused access to lectures and examinations if she continued to do so.

3. Legitimate aim

99. Having regard to the circumstances of the case and the terms of the domestic courts' decisions, the Court is able to accept that the impugned interference primarily pursued the legitimate aims of protecting the rights and freedoms of others and of protecting public order, a point which is not in issue between the parties" (Leyla Şahin v. Turkey, 2005).

[19] The case *Sanoma Uitgevers B.V. v. the Netherlands* (2010) was lodged under Article 34 of the ECHR "by a limited liability company (*besloten vennootschap met beperkte aansprakelijkheid*) incorporated under Netherlands law, Sanoma Uitgevers B.V. ("the applicant company"), on 1 December 2003" (Sanoma Uitgevers B.V. v. the Netherlands, 2010). The company alleged violations of Article 10 of the ECHR "as a result of their having been compelled to give up information that would allow sources of journalistic information to be identified" (Sanoma Uitgevers B.V. v. the Netherlands, 2010). The Court found a violation of Article 10, since national law was "insufficiently accessible" (87.) and the fact that "orders to disclose sources potentially have a detrimental impact, not only on the source, whose identity may be revealed, but also on the newspaper or other publication against which the order is directed, whose reputation may be negatively affected in the eyes of future potential sources by the disclosure, and on members of the public, who have an interest in receiving information imparted through anonymous sources" (89.) (Sanoma Uitgevers B.V. v. the Netherlands, 2010). Henceforth, the Court reiterated that "the national law should be both adequately accessible and foreseeable, that is, formulated with sufficient precision to enable the individual – if need be with appropriate advice – to regulate his conduct" (81.) (Sanoma Uitgevers B.V. v. the Netherlands, 2010).

the authority and impartiality of the judiciary" (Tao, 2023, pp. 85-86). The criminalization of acts of terrorism and/or incitement to terrorism clearly meets the criteria of Article 10(2), in so far as such acts represent "one of the most serious threats to international peace and security", posing "a serious and growing danger to the enjoyment of human rights, threaten(ing) the social and economic development of all States, (and) undermin(ing) global stability and prosperity" (UN Security Council Resolutions 2178 (2014)[20] & 1624 (2005)[21], as cited in Tao, 2023, p. 86).

The third requirement, "the test of necessity and proportionality", implies the existence of a "pressing social need", which must be faced with proportionate measures, as the Court has explained in *Hogefeld v. Germany* (2000)[22] (Tao, 2023, p. 85). Such requirement must also be used

[20] Building on previous Resolutions to address the evolving challenges of global terrorism, Resolution 2178 (2014) of the UN Security Council addressed the threat posed by foreign terrorist fighters (FTFs), demanding them to "(1.) disarm and cease all terrorist acts and participation in armed conflict" (United Nations Security Council, 2014, p. 4). It mandates that all UN Member States "(2.) cooperate in efforts to address the threat posed by foreign terrorist fighters, including by preventing the radicalization to terrorism and recruitment of foreign terrorist fighters, including children, preventing foreign terrorist fighters from crossing their borders, disrupting and preventing financial support to foreign terrorist fighters, and developing and implementing prosecution, rehabilitation and reintegration strategies for returning foreign terrorist fighters" (United Nations Security Council, 2014, p. 4).
Furthermore, it "(11.) calls upon Member States to improve international, regional, and subregional cooperation, if appropriate through bilateral agreements, to prevent the travel of foreign terrorist fighters from or through their territories" and "(16.) Encourages Member States to engage relevant local communities and non-governmental actors in developing strategies to counter the violent extremist narrative that can incite terrorist acts, address the conditions conducive to the spread of violent extremism, which can be conducive to terrorism, including by empowering youth, families, women, religious, cultural and education leaders, and all other concerned groups of civil society and adopt tailored approaches to countering recruitment to this kind of violent extremism and promoting social inclusion and cohesion" (" (United Nations Security Council, 2014, p. 6).

[21] See footnote 11.

[22] The case *Hogefeld v. Germany* (2000) was dismissed by the ECtHR (Hogefeld v. Germany, 2000). The applicant, a former member of the left-wing terrorist group Red Army Faction (RAF), was arrested in June 1993 and "taken in detention on remand", while in 1994 a warrant of arrest was issued against her based on concerns she might continue terrorist activities (Hogefeld v. Germany, 2000). Between 1995 and 1996, multiple courts in Germany, among which the Frankfurt Court of Appeal and the Federal Constitutional Court, denied requests by journalists and filmmakers to interview or film the applicant in prison, citing risks of promotion of terrorism and of renewed support for the RAF in light of the words that the applicant might have voiced during those interviews (Hogefeld v. Germany, 2000). During her trial in front of the Frankfurt Court of Appeal (1994–1996), the applicant made several public statements reflecting on the RAF's history, her role in the organizations and her crimes: while she admitted responsibility for her role in the group and condemned past violence she reaffirmed her belief in revolutionary ideals and called for critical engagement with the past to inform future struggles (Hogefeld v. Germany, 2000). The Frankfurt Court thus sentenced her to life, a decision reaffirmed by the Federal Constitutional Court of Justice (Hogefeld v. Germany, 2000). The European Court of Human Rights, in front of the reasonings of the German Courts, considered "that the restrictions imposed on the applicant's freedom of expression could reasonably be regarded as answering a "pressing social need" and that the

to assess “measures taken by domestic authorities to maintain national security and public safety in the fight against terrorism” (Tao, 2023, p. 86).

The Court strives to apply these principles to “the specific circumstances of each case” since there are “no clear rules as to where the line should be drawn in a general sense”, thus guaranteeing a higher degree of understanding of the different cases (Tao, 2023, p. 87). In cases involving incitement to terrorism and terrorism, the ECtHR's analysis concentrates mainly on the third requirement, carefully weighing state security measures against the right to freedom of speech, especially when the speech involves controversial or politically sensitive content: on its part, the Court pushes its Member States to apply its same approach and rigor, so as to prevent terrorism without unjustifiably suppressing dissent or critical opinions (Tao, 2023, pp. 86-87).

International and regional organizations are therefore increasingly striving to find the right balance between the two imperatives, even if the growing role of online platforms in this field may impinge upon their responsibility and authority, creating great tensions between the public and the private sphere in the field of counterterrorism online.

1.4 The Role of Online Platforms in Protecting and Restricting Freedom of Speech

With most of society living in two separate but interconnected realms, the real and the digital ones, the role of online platforms and of the private companies governing the latter in ensuring compliance with users’ rules has become pivotal. Online platforms are increasingly becoming powerful gatekeepers of digital speech, which often substitute governmental authorities in deciding what is and what is not harmful speech online. Social media platforms and chatbots

reasons adduced by the national courts are “relevant and sufficient”, thus unanimously declaring the application inadmissible (Hogefeld v. Germany, 2000).

now function as *de facto* public squares where speech, political debate, protest, and cultural exchange take place. However, in these new digital public squares, traditional legal frameworks struggle to apply effectively. Unlike state actors, the private owners of digital platforms are not bound directly by constitutional free speech protections. Rather, they enforce their own terms of service and policies which are often unclear, inconsistent, and lack transparency and democratic safeguards. The internet follows its own interests, summarized by John Perry Barlow in his utopic 1996 *Declaration of the Independence of Cyberspace*[23] which clearly rejects government interference and control, and idealizes the internet and cyberspace as a new realm, with a new "Social Contract" completely detached from traditional legal frameworks, where freedom of speech reigns as sovereign (Bromell, 2021, pp. 30-31). While the objective of the internet seems to be to guarantee extreme protection of the right to freedom of speech, things in practice work very differently. Overtime, in fact, this libertarian approach to the digital space became aware of the dangers behind an illegitimate use of the internet, and gave rise to four competing objectives in the governance of the digital space, recognizing

[23] The *Declaration* reads in part:

> "Governments of the Industrial World, you weary giants of flesh and steel, I come from Cyberspace, the new home of Mind. On behalf of the future, I ask you of the past to leave us alone. You are not welcome among us. You have no sovereignty where we gather.
>
> We have no elected government, nor are we likely to have one, so I address you with no greater authority than that with which liberty itself always speaks. I declare the global social space we are building to be naturally independent of the tyrannies you seek to impose on us. You have no moral right to rule us nor do you possess any methods of enforcement we have true reason to fear. …
>
> Your legal concepts of property, expression, identity, movement, and context do not apply to us. They are all based on matter, and there is no matter here.
>
> Our identities have no bodies, so, unlike you, we cannot obtain order by physical coercion. We believe that from ethics, enlightened self-interest, and the commonweal, our governance will emerge. …
>
> In our world, whatever the human mind may create can be reproduced and distributed infinitely at no cost. …
>
> We will create a civilization of the Mind in Cyberspace. May it be more humane and fair than the world your governments have made before (Barlow, 1996, as cited in Bromell, 2021, pp. 30-31).

> (1) the need to "maintain a free, open and secure internet", (2) the need to "prevent abuse … that harms individuals and communities and damages democracy", (3) the need for "appropriate checks and balances on the use and abuse of the internet for commercial purposes", and (4) the need to "avoid or discourage" governments' "surveillance, censorship and suppression of dissent" (Bromell, 2021, p. 31).

In such a context, it becomes significantly harder to determine what are the legitimate restrictions on the right to freedom of speech which platforms may impose, and which are the standards of legality, necessity, and proportionality that they apply when removing content online. While governments must explicitly adopt these standards when restricting speech because they are bound by constitutional protections, online platforms are not obliged to offer equally rigorous justifications for their restrictions. This allows for new forms of restrictions to come to life, especially when speech is removed or sanctioned because it goes against the values or the principles of the big tech company behind the platform. These restrictions are less visible, less contestable than the ones imposed by physical governments: indeed, they are generally announced solely to the users whose content is being removed; they are hard to contest and there are generally no independent supervising bodies which may control whether the restriction is right or wrong. It is also important to highlight that, as private bodies, big tech companies' main goal is not to protect the fundamental rights of their users, but to increase their profits. Online service providers collect extensive personal data to sell targeted advertising, anticipating users' interest and shaping their thoughts, behaviors and purchasing decisions (Frenkel & Kang, 2021, as cited in Bromell, 2021, p. 56). As Bromell (2021, p. 56) explains the algorithms used for this purpose are neither transparent nor accountable and can potentially impinge on fundamental rights. Furthermore, the goal of these "attention economies" created by online platforms aim to sell to users what they like,

pushing them into the famous rabbit-holes of the internet which worsen users' confirmation biases and push their ideologies to become more extreme (Bromell, 2021, pp. 58-61). Clearly, with their enormous impact on public discourse, these platforms may well be sources of radicalization online and impinge upon the right to freedom of speech of many users, especially considering the lack of transparency of their algorithms.

Thus, as private tech companies take on quasi-governmental roles of these new digital public squares, without having either the same constitutional limitations or goals of physical governments, the challenge lies in establishing legal frameworks that ensure that the right to freedom of speech and other fundamental rights are respected and enforceable. With regard to this, many big tech companies have collaborated with governments and international organizations to remove extremist content online, especially in the European Union. Indeed, the *EU Code of Conduct on Countering Illegal Hate Speech Online* has been stipulated in 2016 between Facebook, Microsoft, Twitter and YouTube, (followed by Snapchat and Dailymotion in 2018, TikTok in 2020, and even Linkedin in 2021) and the European Commission to ensure the elimination of harmful content online (European Commission, n.d.b). Under this Code, regular monitoring of online platforms must be performed by specialized agencies in EU Member States, in order to ensure that all platforms are not removing legal content when trying to eliminate harmful ones (European Commission, n.d.b). However, such an initiative remains a code of conduct, lacking enforceability and consistency, and always leaving the last word to big tech companies. This is exactly why the European Union has been trying to develop a common legal framework for the promotion of a safe, and respectful digital sphere, with its regulations on counterterrorism first and, more recently, with its AI Act (European Commission, 2025a).

1.5 The Role of Online Platforms in Counterterrorism: Navigating Public-Private Responsibilities

In light of the increasing employment of online platforms by terrorist groups for incitement, planning, radicalization, and recruitment, major platforms like YouTube and Facebook have found themselves obliged to impose stricter regulations and enhance the removal of terrorist content. For instance, in Europe, they have signed the aforementioned *Code of Conduct* with the European Commission, striving to eliminate all harmful content from their digital public squares (European Commission, n.d.b). Particularly after the Christchurch attacks, then, all the major flaws in the existing moderation systems of online platforms came to be evident, and the state of New Zealand pushed for the creation of a new international organization, namely the Christchurch Call[24], "to eliminate terrorist and violent extremist content online", through community building, "understanding algorithms and developing interventions, responding to crises, (and) increasing transparency" (Christchurch Call, n.d.). Furthermore, Facebook, Microsoft, Twitter and YouTube, later joined by many other big tech companies have also created the GIFCT, an independent

[24] The Christchurch Call to Action reads:

> "A free, open and secure internet is a powerful tool to promote connectivity, enhance social inclusiveness and foster economic growth.
>
> The internet is, however, not immune from abuse by terrorist and violent extremist actors. This was tragically highlighted by the terrorist attacks of 15 March 2019 on the Muslim community of Christchurch – terrorist attacks that were designed to go viral. The dissemination of such content online has adverse impacts on the human rights of the victims, on our collective security and on people all over the world. …
>
> The events of Christchurch highlighted once again the urgent need for action and enhanced cooperation among the wide range of actors with influence over this issue, including governments, civil society, and online service providers, such as social media companies, to eliminate terrorist and violent extremist content online. The Call outlines collective, voluntary commitments from Governments and online service providers intended to address the issue of terrorist and violent extremist content online and to prevent the abuse of the internet as occurred in and after the Christchurch attacks.
>
> All action on this issue must be consistent with principles of a free, open and secure internet, without compromising human rights and fundamental freedoms, including freedom of expression. It must also recognise the internet's ability to act as a force for good, including by promoting innovation and economic development and fostering inclusive societies" (Christchurch Call, 2024).

organization “which brings together the technology industry, civil society, and academia to foster collaboration and information-sharing to counter terrorist and violent extremist activity online”, through (1) prevention of and (2) response to terrorist threats and (3) the adaptation of the technological industry to understand new terrorists’ exploitations of technology (GIFCT, n.d.).

Notwithstanding all these advances, the efforts of big tech companies in the field of counterterrorism remain scarce due to four main factors listed by Bromell (2021, p. 38), namely:

1. “Inconsistent moderation and enforcement of online service providers’ terms of service and community standards/guidelines;
2. Difficulties in detecting and enforcing the removal of harmful content, particularly where this is implicit or coded;
3. The lack of a common global regulatory framework that is consistent across platforms, responsive to local context and able to be operationalized to remove harmful content quickly; and
4. The persistence of digital content over time”.

Moreover, the privatization of counterterrorism efforts greatly undermines public accountability: most of the times, when platforms decide to remove harmful content, they do not have to reveal their reasons to their users, with their decisions resulting opaque and unreviewable. Furthermore, the moderation systems of these online platforms are set up to “work best in English, and in text-based content rather than videos, images and messaging apps” giving the possibility to those speaking other languages to quite easily bypass their control: in 2014, for instance, a review carried out by Samaratunge and Hattotuwa highlighted an abundant presence of harmful and illicit content on Facebook written in Sinhalese (Bromell, 2021, pp. 40-41). Facebook, indeed, can be taken as an example for all other social media platforms, like YouTube, aiming to regulate terrorist content

online; all other social media platforms, in fact, either follow its controversial approach to counterterrorism regulations or implement even milder regulations. X, for instance, has been known to be much more reluctant in eliminating any kind of speech from its feeds. A close inspection at Facebook's regulation of extreme or harmful content, therefore, can be illustrative of how weak the policy lines on the matter of counterterrorism are even in the most 'regulated' platform. Despite claiming to wanting to "block praise, support, and representation of white supremacism" after the 2017 Charlottesville Attacks, the platform failed to remove Neo-Nazi groups from its digital public square even after the 2019 Christchurch attacks; then, after the Taliban take-over of Kabul in 2021, Facebook, followed by YouTube and Twitter, "banned Taliban content on their platforms" signaling greater ease in banning Jihadist content inciting to terrorism than the same content posted by far-right movements (Bromell, 2021, pp. 39-40).

Clearly, while there have been increasing attempts at regulating harmful content online, the results have been scarce and controversial at best. Big tech companies must still develop greater safeguards in this respect even if, due to the same nature of the internet, it is hard to really remove content from the web. Moreover, if one considers the overlap between the private and the public sector in this field further problems arise. This is because "digitalization", as specified by Bromell (2021, pp. 36-37) "complicates distinctions between public and private space", and especially between "concealment and exposure" due to the "culture of oversharing" prompted by social media, and between "online anonymity and pseudonymity". Every crime, when committed online, is harder to punish, because in the digital space the legal responsibilities are not as clear as they are in real life. Furthermore, it is even harder to establish with certainty the people responsible for certain crimes, for some individuals might hide behind fake accounts or anonymity with great dexterity.

In cases of counterterrorism, then, these issues become even more contentious. When international organizations or governments oblige online platforms to apply more stringent content regulations and security policies and pressure them to comply with national security demands, these platforms face a great dilemma: complying with these demands, indeed, could clash with their business models and with the expectation of their users, possibly making them lose profit. In cases where they perceive their profits going down, their users diminishing or where they are feeling obliged to remove what they perceive to be lawful content under their standards, platforms might feel discouraged to submit to governments' requests. In the opposite cases, instead, where they perceive the content to be removed as illegal or even too controversial for their standards or where they see that most of the users are reacting badly to such a content, these platforms could very well accept these government and organizations' demands to preserve user trust and avoid reputational harm from being seen as vehicles for extremist content. Hence, in cases where both the public and the private sector benefit, cooperation appears possible and the tension between the two sectors diminish: public-private partnerships can "foster shared responsibilities in safeguarding communities" thanks to cooperation in measures of risk assessment and management strategies (Editorial Team, 2024).

However, the tension between the two sectors is, most of the time, high. This is mainly due to the fact that, when addressing counterterrorism online and protecting in the meantime fundamental rights, neither the public nor the private sectors can be considered fully accountable. The lack of clear international norms on this matter, then, only exacerbates this tension, with different countries and organizations taking very different stances on this matter. Worsening even more this tension, the new AI revolution adds complexities in the field of online counterterrorism and on how to balance security imperatives with the protection of fundamental rights.

2. The Ascent of AI and its Impact on Terrorism, Counterterrorism, and Freedom of Speech

2.1 The Rise of Artificial Intelligence: A Short Introduction

A revolution into the revolution, the spread of artificial intelligence has brought about new fundamental shifts in the way in which the real and the digital world interact among each other. As of today, AI permeates nearly every aspect of daily life: from virtual assistants like Siri or Alexa to personalized recommendations, and even from entertainment platforms such as Netflix or Spotify to targeted advertising from big tech companies such as Google, Amazon or Facebook, AI has truly become an essential part of every person's life (Kausik & Rashid, 2024, p. 2). Statista reports that revenues from the AI software market worldwide "grew beyond US $ 184 billion (only) in 2024" and it is projected to raise past US $ 826 billion in 2030, thus signaling the extensive use that is being made of AI systems (Thormundsson, 2024). These machine-based systems are in fact capable of mimicking "human cognitive functions, such as problem solving and decision making … to make intelligent decisions autonomously … (and) without human intervention" (Kausik & Rashid, 2024, p. 2). Large data sets lie behind the decisions and the reasoning of AI systems, which allow them to be easily integrated in many different sectors, such as education, healthcare, manufacturing, industrial automation, finance, cybersecurity, and even national security. This impressive diffusion of AI systems in such disparate sectors marks what Kausik and Rashid (2024, p. 1) term as the era of "Industry 4.0", a Fourth Industrial revolution with the employment of "AI … Big Data, IoT (Internet of Things), Sensors, CPS (Cyber-Physical

Systems, i.e., robots), and Blockchain" in different industries, all "to reduce cost, time and labor" and optimize industries by reducing the human workforce". This era has however already sunsetted, allowing for the new "Industry 5.0" era to rise, integrating societal and ethical values within the field of AI development and use to allow for human-centric, transparent, and sustainable AI innovation (Kausik & Rashid, 2024, p. 2). However, some legislative issues have already arisen. Unlike traditional softwares, AI systems do not operate based on "explicit designs" and fixed and interpretable code alone (p. 85). In other words, in the AI era, Lessig's early 2000s *dictum* "code is law" referring to the regulation of "software and protocols on the nascent Internet" no longer applies, since AI systems, "created through a massively resource-intensive training process of tuning trillions of parameters" are "opaque and not designed" (Judge et al., 2025, pp. 85-86). Due to the "black-box" nature of AI systems, developers and engineers cannot encode specific rules during the creation of these systems, but rather "must hope the model abides by the desired behavior after sufficient reinforcement" (Judge et al., 2025, p. 86). This creates a situation where "it is impossible to demonstrate compliance with a given regulatory specification", underscoring the growing need to regulate not only AI use but also AI development (Judge et al., 2025, p. 86). A pivotal example is the European Union's EU AI Act, a landmark regulation on AI which aims to balance technological advancement with the protection of fundamental rights, and which will be analyzed in depth in the next chapter.

This chapter, instead, focuses on describing how AI systems are increasingly being used in the field of counterterrorism, exploring the balance between security and freedom of speech in this sector. Thus, it first analyzes the AI's impact on freedom of speech, examining the role of big tech companies in arbitering what is and what is not permissible speech on their platforms; consequently, it examines how AI systems are being employed in terrorism, looking specifically

at its use in the processes of radicalization and recruitment. The chapter then inspects how these systems are being implemented in counterterrorism measures examining both advantages and disadvantages of this tactic and looking at which countries have already implemented AI in this field; consequently, it addresses the role of international organizations in setting ethical standards for AI in security. Finally, the chapter analyzes whether AI systems can effectively balance security imperatives and the protection of fundamental rights when employed for counterterrorism measures, the central question of this thesis.

2.2 The Impact of AI on Freedom of Speech

The pupil of classical liberalism and then the sovereign of the digital world, the right to freedom of speech is now under serious threat due to the growing adoption of artificial intelligence in online content moderation. Human supervision of these new AI systems is, indeed, only limited or inexistent, leaving the latter to analyze, assess, and remove vast amounts of content and data by themselves. While these systems are certainly quicker in carrying out these tasks than a human operator, they lack contextual understanding which would allow them to distinguish between the different nuances of speech existing between harmful speech, on the one hand, and legitimate speech, on the other.

This is exactly the problem of digital tools which this thesis has analyzed in its first chapter. However, while this problem has been evident since the early 2010s with the creation of social media's automated content moderation systems, the efforts to solve it have been minimal with the result that, now, the integration of AI systems in this field of content moderation has exacerbated the issue rather than solved it. For instance, AI systems may inadvertently remove lawful political discourse, or journalistic articles, or even activist content, simply because they are not able to

distinguish the context in which certain terms are used or because the systems may be prone to misclassifications and confirmation biases (Oh & Downey, 2024, p. 2). Indeed, content which includes references to "minority identity … (e.g., Arabs, Black, LGBT+)" or which contains explicit language or "messages with specific racial and ethnic dialects" are more likely to be wrongly "misclassified (by AI systems) as toxic regardless of their communicative contexts" (Gorwa et al., 2020; Thiago et al.,2021; Zhou, 2021, as cited in Oh & Downey, 2024, p. 2).

Hence, not only AI can severely impact the freedom of speech of individuals, but it can also augment the marginalization of minorities, allowing many states to employ such technologies online to eradicate dissent, and reduce the freedom of expression of those individuals or groups of individuals which they consider dangerous to their stability and security. As a matter of fact, in 2023, Freedom House reported that in 55 out of the 70 states covered by its project *Freedom of the Net,* and especially in China and Myanmar, "people faced legal repercussions for (simply) expressing themselves online", and that AI systems were used in at least 21 states "to remove disfavored political, social, and religious speech" (Funk et al., 2023). AI can in fact be employed "as an amplifier of digital repression, making censorship, surveillance and the creation and spread of disinformation easier, faster, cheaper, and more effective" (Funk et al., 2023). This is because algorithms are created by humans, who are biased by nature, and therefore they can sometimes be working with biased datasets which can perpetuate stereotypes or unfairly target specific groups.

All in all, the great problem of AI systems is their lack of transparency, because this allows both big tech companies and governments to deploy such systems in ways that may violate users' fundamental rights online without proper accountability. This lack of transparency and accountability, then, can generate worrisome consequences for users who, not only may be more prone to suffer from "online violence and harassment" or "privacy and data protection" breaches,

but also become victims of disinformation, online manipulation and great breaches of their right to freedom of expression and information (Siagian et al., 2023, pp. 553-554). Furthermore, the risks posed by "filter bubbles" on the right to freedom of expression are significant, in so far as AI systems and algorithms can force "users' exposure to only pre-existing views", extremizing their initial ideologies and precluding a variety of different perspectives to the users (Siagian et al., 2023, p. 555).

Crucial to addressing AI systems' negative impact on freedom of speech is therefore the implementation of "stronger regulations, better transparency, greater user control over their data, and ethical data usage" by platforms which will serve not only to increase users' trust in AI systems and big tech companies but will ensure that every individuals' digital rights are upheld and respected (Siagian et al., 2023, p. 555).

2.2.1 Who Sets the Rules: Big Tech Companies in Deciding What Is and Isn't Free Speech

The owners of most AI systems and models, big tech companies, are becoming the 21st century arbiters of the right to freedom of speech. Through their algorithms and content moderation policies, indeed, they can effectively set the rules on what is and what is not legitimate speech on their platforms. However, such companies are extremely unaccountable to their users about their decision-making processes: this means that, most of the time, users do not really know or understand why some content is removed over some other, since the moderation processes of these platforms are anything but transparent and clear. There could henceforth be instances of over-reporting or even excessive downranking, through which content aimed at stifling "the social progress that could be promoted by sharing sensitive content" is flagged or removed by algorithms because of their lack of contextual understanding (Ganapini et al., 2020, p. 9).

As highlighted in the first chapter of this thesis, private tech companies have very different standards and goals from national governments and must not even submit to the same checks and balances as the latter. Crucially, these companies generally value profit over their users' fundamental rights: henceforth, while they may claim to be safeguarding freedom of speech and other digital rights, their priority remains that of increasing their profits. In line with this, it is important to emphasize that these big tech companies are in the hands of a very small number of powerful individuals and corporate entities, who may have very different ideas on what is and what is not lawful speech online. These individuals and entities have been effectively given the power to control the right to freedom of speech and information of the billions of people worldwide who use their platforms, deciding what is extremist or harmful content and what is not by applying (most often than not) vague standards which may be influenced by their own biases, by their investors' biases or even by government agencies which fund their platforms. In short, in the digital realm the right to freedom of speech reigns as sovereign, until someone decides that it is no longer the case.

This appears even more worrisome if one considers that there are no effective supervisory bodies of the content moderation systems of these platforms which may swiftly control whether the content that has been banned is indeed harmful or incites to crime and terrorism. Furthermore, these AI systems and platforms are not immune to biases, which they often encode and then perpetuate in their work. Hasimi and Poniszewska-Marańda (2024, p.4) have deeply explored this matter and have identified 5 different types of biases present in all machine-based systems, namely:

1. "Sample bias (which) occurs when a dataset does not reflect the realities of the environment in which a model will run.

2. Exclusion bias (which can occur when) …deleting valuable data thought to be unimportant or due to the systematic exclusion of certain information.
3. Observer bias (which)... is the effect of seeing what you expect to see or want to see in data.
4. Racial bias (which) … occurs when data skews in favor of particular demographics. This has been seen in facial recognition and automatic speech recognition technologies. (And, finally)
5. Association bias (which) … occurs when the data for a machine learning model reinforces and/or multiplies a cultural bias".

Taken together, the unaccountability of these AI systems, the concentration of decision-making powers in the hands of a few very powerful corporations, and the biases which are embedded in their systems, all underscore the urgent need for a robust regulatory framework on AI. It is indeed important that, going forward, AI systems are developed to be more transparent, more accountable and more respectful of their users and of their fundamental rights. The following section aims exactly at proving how necessary this regulatory framework really is, bringing the case of Meta Platforms, Inc. and of the rules which apply in deciding what is and what is not free speech on its platforms.

2.2.2 What are The Rules set by the Big Tech Companies? The Case of Meta Platforms, Inc.

Meta Platforms, Inc. has been chosen as representative of other big tech companies like X and YouTube for its leading role in the development and employment of automatic content monitoring systems. Indeed, all these companies can unilaterally decide, often without meaningful supervision, what constitutes extremist or harmful content on their platforms. While every

platform works differently and has its own Terms of Use, they all set some minimum restrictions on what is permissible to share in their feeds. Arguably, Meta is the company which has set the most stringent restrictions when dealing with harmful or extremist content; and, due to this, it has been chosen as the case study of this paragraph.

Meta Platforms, Inc. has published a comprehensive set of Community Standards to govern user behavior on its platforms, namely Facebook, Messenger, Instagram and Threads, and "enable billions of people to freely express themselves across countries and cultures and in dozens of languages" (Meta, n.d.). These standards, based on users and experts' feedback, are designed to protect users from harmful content and, even more importantly, to allow users to talk about what they want, "even if some may disagree" (Meta, n.d.). Meta (n.d.) claims, in fact, in its *Commitment to Voice*:

> "Meta wants people to be able to talk openly about the issues that matter to them, whether through written comments, photos, music, or other artistic mediums, even if some may disagree or find them objectionable. In some cases, we allow content—which would otherwise go against our standards—if it's newsworthy and in the public interest. We do this only after weighing the public interest value against the risk of harm, and we look to international human rights standards to make these judgments. In other cases, we may remove content that uses ambiguous or implicit language when additional context allows us to reasonably understand that the content goes against our standards".

Therefore, when Meta (n.d) enacts some restrictions on the right to freedom of expression, it does so to ensure "the content people see is authentic, … (and does not) contribute to a risk of harm to the physical security of persons" and to uphold the privacy and dignity of its users. On its website,

Meta (n.d.) lists 26 content categories[25] where freedom of speech may be limited, including incitement to violence, hate speech, human exploitation, misinformation, and violations of election integrity. For the purposes of this thesis, it is important to consider the category of "violence and incitement" where Meta aims at preventing potential offline violence originating from online expression, "removing content which incites or facilitates violence, which poses threats to public and personal safety, and which provides offers of violent services, instructions on weapon or explosive use with violent intent, and threats linked to elections or public events" (Meta, n.d.).

To enforce its standards quicker and at scale, Meta has created an AI system called Few-Short Learner (FSL) which can "adapt to take action on new and evolving types of harmful content within *weeks* instead of months" (Meta, 2021). Trained on vast datasets to identify patterns associated with prohibited content, the system has been effective at identifying nuanced or implicit harmful content that would typically evade detection, such as veiled incitements to violence (Meta, 2021). While Meta reports successes in this implementation, believing that "FSL can, over time, enhance the performance of all of (their) integrity AI systems by letting them leverage a single, shared knowledge base and backbone to deal with many different types of violations", there is still work to do to ensure that its AI systems develop contextual understanding and are able to efficiently remove only truly harmful content (Meta, 2021).

Furthermore, to address concerns about content moderation decisions, Meta created an independent Oversight Board, with the aim of reviewing and deciding on content moderation

[25] The full list comprehends: "(1) coordinating harm and promoting crime, (2) dangerous organizations and individuals, (3) fraud, scams, and deceptive practices, (4) restricted goods and services, (5) violence and incitement, (6) adult sexual exploitation, (7) bullying and harassment, (8) child sexual exploitation, abuse, and nudity, (9) human exploitation, (10) suicide, self-injury, and eating disorders, (11) adult sexual solicitation and sexually explicit language, (12) hateful conduct, (13) privacy violations, (14) violent and graphic content, (15) account integrity, (16) authentic identity representation, (17) cybersecurity, (18) inauthentic behavior, (19) memorialization, (20) misinformation, (21) spam, (22) third-party intellectual property infringement, (23) using meta intellectual property and licenses, (24) additional protection of minors, (25) locally illegal content, products, or services, (and finally, 26) user requests" (Meta, n.d.).

cases: on April 2025, for instance, the Board has opened a procedure on the new hate speech policies implemented in January on Facebook, issuing 17 recommendations to Meta, including calls for the company to evaluate the "effectiveness of its new community notes system", and for clarifying its updated position on hateful ideologies (Zeff, 2025). Indeed, after Trump's re-election, Zuckerberg followed Musk's lead on X and relaxed certain hate speech and abuse rules on his platforms, allowing content that was previously banned under the guise of promoting more speech (Ortutay, 2025). Professor Leiner of the University of Virginia's Darden School of Business views "th(is) policy change (as) a tactic to earn favor with the incoming administration while also reducing business costs related to content moderation", which "will lead to real-world harm, not only in the United States where there has been an uptick in hate speech and disinformation on social media platforms, but also abroad where disinformation on Facebook has accelerated ethnic conflict in places like Myanmar" (Ortutay, 2025). This current shift in policies is even more alarming if one considers that the powers of the Oversight Board are limited, insofar that it can rule only on individual posts and not on entire policy updates (Zeff, 2025). This shortcoming in accountability puts users at great danger because it essentially allows platforms to arbitrarily decide what to remove and what to leave on their feeds. Despite the Oversight Board's advocacy, Meta is under no legal obligation to implement its recommendations. Meta appears, now more than ever with the new US administration, more motivated by financial incentives than by human rights concerns, with the "more speech" narrative serving as a rhetorical shield for policy changes that reduce moderation infrastructure and increase platform engagement.

It thus appears especially important that the international community pushes for greater accountability and transparency of AI systems, creating strong regulatory frameworks capable of addressing both security concerns and uphold freedom of speech protection. In so doing, the

international community could impose necessary constraints on the arbitrary and profit-driven practices of big tech companies, thereby promoting a safer and more human-centered digital sphere.

2.3 AI in Terrorism: Radicalization and Recruitment

With the diffusion of AI systems in every aspect of everyday life, it seemed rather obvious that terrorist organizations would have wanted to use such systems to advance their agendas. If used improperly, indeed, they may not only negatively impact on fundamental rights, but also "destabilize economies, political institutions, and international relations through targeted psychological impacts on people's consciousness" (Pashentsev, 2023, p. 1). Terrorist and extremist groups have been indeed capable of taking advantage of every step of the digital revolution: they started publishing "extremist websites in the late 1990s", and then they exploited the new social media platforms and encrypted messaging apps of the 2010s, such as "Facebook, Youtube, Twitter, Instagram, … Tik Tok (and) Telegram", but also "anonymous cloud storage platforms, and even the Dark Net … to communicate and coordinate worldwide operations with reasonable expectations of privacy and security", (Weimann et al., 2024, pp. 18-19). Another purpose for which extremist and terrorist groups have been using AI is to spread their radicalized messages and to welcome in their ranks new recruits: McGuffie and Newhouse (2020, as cited in Weimann et al., 2024, p. 18) have indeed highlighted a significant risk posed by generative AI and foundation models, such as Large Language Models (LLMs) like ChatGPT of being used as tools of "large-scale radicalization and recruitment" in their experiment where they fed ChatGPT with extremist content, finding that the system would negatively impact on the radicalization of users.

Deep fake videos, synthetic speech and automated chatbots can all be used to manipulate and radicalize users and for propaganda and recruitment purposes. AI systems are indeed able to create such emotional and persuasive materials that can be tailored to specific users using algorithmic profiling which are hard to resist for people. To avoid losing users, for instance, AI systems may deploy the famous rabbit-hole theory, increasingly exposing users to content they like (in this case, increasingly extreme content) through automated suggestions and reinforcing their ideology. Terrorist groups may easily take advantage of this "to influence public sentiment and expand the impact of their attacks" (Esmailzadeh & Motaghi, 2024). They could be doing this through a process of social engineering which aims at psychologically manipulating people "into performing actions or divulging confidential information" (Webroot, 2021, as cited in Bazarkina, 2023, p. 252). A sort of psychological blackmailing could therefore push AI users to join some terrorist cause, but there are also those users who just need a small push for joining, and they may find it on online chat rooms, social media pages or websites which incite to acts of terrorism and explain how to carry them out.

All in all, a 2023 report by the GIFCT summarizes all the greatest threats posed by the use of generative AI by terrorist or extremist groups, including:

> "(1) Propaganda: AI can be used to generate and distribute propaganda content faster and more efficiently than ever before. This can be used for recruitment purposes or to spread hate speech and radical ideologies. AI-powered bots can also amplify this content, making it harder to detect and respond to.
>
> (2) Interactive recruitment: AI-powered chatbots can interact with potential recruits by providing them with tailored information based on their interests and beliefs, thereby making the extremist group's messages seem more relevant to them.

(3) Automated attacks: Terrorists can use AI to carry out attacks more efficiently and effectively—for example, by using drones or other autonomous vehicles.

(4) Social media exploitation: AI can also be used to manipulate social media and other digital platforms to spread propaganda and recruit followers.

(5) Cyber-attacks: AI can be used by extremist groups to enhance their ability to launch cyber-attacks against targets, potentially causing significant damage" (Weimann et al., 2024, p. 19).

It appears thus evident that the integration of AI systems into terrorist strategies marks a dangerous evolution in terrorism, because such systems are capable, very easily and in a short amount of time, of amplifying not only the reach of their attacks but also their influence and propaganda worldwide. Appropriate safeguards and human supervision are therefore needed to avoid that AI systems enable even more radicalization and acts of terrorism, and indeed there are many countries which are already employing them for counterterrorism purposes.

2.4 The Role of AI in Counterterrorism: Advantages and Disadvantages

The increasing use of AI systems for terrorist purposes has induced states to adopt AI-driven measures for counterterrorism efforts, in an attempt to 'fight fire with fire.' AI systems have indeed already been deployed in many counterterrorism actions for "determining the structure of terrorist networks and organizations, identifying disputes and splits, recognizing incitement to terrorism online, (and) locating high value targets" (Ganor, 2018, p. 606). This is because AI systems are capable of analyzing and working with a large amount of data, "including social media posts, internet searches and financial transactions" which may be connected to terrorist organizations or actions (Esmailzadeh & Motaghi, 2024, p. 169). Indeed, the Computing Research

Institute in Qatar has been capable of discovering ISIS support in Iraq and Syria solely by analyzing around 3 million tweets on X with an AI algorithm able to "identify tweeters as opponents or supporters of ISIS with 87 percent accuracy" (Ganor, 2018, p. 606). The United Kingdom, instead, has employed a data retrieval system which is capable of making "connections between Person, Object, Location, and Event" and of "building a complete profile and network of associations of the monitored subjects" (Ganor, 2018, pp. 606-607). Exactly because they are able to analyze such large amounts of data, AI systems are also used in predictive policing to identify individuals at risk of being radicalized and areas at risk for terrorist activities: suspicious behavior or activities can be quickly reported to security agencies which can then act accordingly to the danger at hand. In the US, for example, the Military has been employing "big data" to detect terrorist movement through the use of drones in combination with data analysis: through AI systems, in fact, the Military is capable of analyzing "tens of thousands of tweets (posted) daily" by terrorists and supporters "to detect patterns and threats" (Ganor, 2018, pp. 606-607). AI's capability of analyzing great amount of data, moreover, can be deployed to address "the underlying social and political factors that drive terrorism and extremism" online by applying an holistic approach to counterterrorism, which aims to adjust the root causes of extremism and promote social justice and political engagement (Esmailzadeh & Motaghi, 2024, p. 177).

Another field of implementation of AI-systems for counterterrorism purposes is that of content moderation on social media. Predictive machine learning (ML) tools and automated-hash matching can identify and remove terrorist materials from different platforms (Gorwa et al, 2020, as cited in Gunton, 2022, p. 70). Furthermore, AI systems are increasingly being used for surveillance purposes, with measures such as facial recognition, gait analysis, "internet monitoring, bulk interception and collection of communications or telephone tapping" for tracking

alleged terrorists which are less costly than traditional measures of surveillance and act quicker (Dieu & Montasari, 2022, p. 32).

While there are clear advantages in using AI systems for counterterrorism purposes, there is also the other side of the coin to consider. The disadvantages of employing AI systems in this field are indeed many, especially because artificial intelligence is a sector still young, which necessitates much updating before being truly capable of resembling humans' thinking and way of speaking. Ganor (2018, p. 607) classifies all the arguments against the employment of AI systems in counterterrorism in three categories: "(1) generic arguments expressing concern about the growing use of AI and big data and the implications of these processes on human society as a whole; (2) utilitarian arguments that claim that it is impossible to use AI and big data effectively in the prevention of terrorism; and (3) ethical arguments that maintain that the possible damage that might be inflicted on innocent civilians due to the use of AI and big data in the field of counter-terrorism should rule out the use of this technology". Gunton (2022, p. 76), then, summarizes all the negative aspects of using AI systems in content moderation, arguing that such applications are not effective since they do not "take contextual information into account, causing inaccurate applications, such as false positives" and negatively impact on users' freedom of expression "by encouraging the removal of legal content". Indeed, it is much easier for a machine to impinge upon fundamental rights than it is for human analysts who are capable of better understanding the blurred lines between harmful and dissenting speech. In line with this, the use of AI in surveillance may raise concerns about privacy and civil liberties. Moreover, while the use of AI systems surely reduces the burden of human analysts working in the field of counterterrorism, if used in an unregulated manner, it also raises questions about reliability and fairness: algorithmic biases and opaque decision-making processes can in fact undermine transparency and accountability of

security organizations in that there can be errors in data collection or even biased training datasets which may lead to racial profiling, misidentification and even the illegitimate suppression of speech. Other threats which may be amplified with the use of AI in counterterrorism are "the spread of misinformation", propaganda, and "the loss of privacy" (Esmailzadeh & Motaghi, 2024, p. 178). Lastly, the lack of human oversight in many AI systems is clearly a flaw of these systems, because this can erode due process guarantees and accountability and favor the use of such systems for illegitimate purposes.

Notwithstanding the numerous shortcomings highlighted above on the use of AI systems for counterterrorism purposes, numerous states, beyond the US and the UK, have already adopted such measures in their security strategies. This choice highlights some states' realist stance in the international arena, for which the pursuit of security often takes precedence even at the expense of fundamental rights. The international community, and especially many international organizations, have been highly vociferous on the use of AI in the field of counterterrorism to counter this realist stance of most states at the international level and to achieve a meaningful regulation on AI systems, capable of effectively protecting both security and freedom of speech when AI is employed for counterterrorism purposes.

2.4.1 The International Community's Standards on the Use of AI in Counterterrorism

In response to the growing integration of artificial intelligence into the field of online counterterrorism, many international organizations worldwide have intensified efforts to develop safe, transparent and accountable standards on the use of AI in this field, striving to safeguard both security and fundamental rights. The CTED, for instance, cooperates with many organizations in this field, such as the United Nations Interregional Crime and Justice Research Institute and the World Economic Forum, in trying to follow every "development in the use of AI-powered

algorithms by technology platforms (including GIFCT companies) to support their content moderation efforts" (United Nations, n.d.). The UN-Secretary-General, António Guterres, then, has published the 2020 *Road Map on Digital Cooperation*, recognizing AI's potential to promote peace even in the field of counterterrorism (United Nations, 2020). Furthermore, the GIFCT works strenuously to combat terrorism online, promoting AI-driven solutions to detect and remove terrorist content, while encouraging transparency and independent oversight: it has implemented "beyond locality-sensitive hashes for images and videos to develop a process for hashing PDFs based on Text Locality Sensitive Hashing (TLSH)" and it is "constantly looking to make it a more perfect solution while also investing in areas like AI with faculty.ai, better integration with our systems using Hasher-Matcher-Actioner (HMA) with Meta, and better content moderation triage processes working with Jigsaw and Tech Against Terrorism" (GIFCT, 2023). Other fundamental initiatives and treaties at the international level to support an ethical and safe AI governance are: the *Asilomar AI Principles*[26], the *Montreal Declaration for Responsible AI*[27], the principles of the *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems*[28] and the European Group on Ethics in Science and New Technology's

[26] Developed at the Beneficial AI 2017 conference by the Future for Life Institute, the Asilomar Principles are 23 principles divided in three categories, namely Research Issues, Ethics and Values, and Longer-Term Issues, the most important of which are: (6) Safety, (8) Judicial Transparency, (9) Responsibility, (11) Human Values, (12) Human Privacy, (13) Liberty and Privacy, (16) Human Control, (19) Capability Caution, (22) Recursive Self-Improvement, and (23) Common Good (Asilomar Principles, 2017).

[27] The principles outlined in the Montreal Declaration (2018, p. 4) are: (1) Well-Being, (2) Respect for Autonomy, (3) Protection of Privacy and Intimacy, (4) Solidarity, (5) Democratic Participation, (6) Equity, (7) Diversity Inclusion, (8) Caution, (9) Responsibility, and (10) Sustainable Development. In establishing such principles, the Declaration, "addressed to any person, organization and company", had three objectives, namely: (1) the creation of "an ethical framework for the development and deployment of AI"; (2) guiding "the digital transition so everyone benefits from this technological revolution"; and (3) opening "a national and international forum for discussion to collectively achieve equitable, inclusive, and ecologically sustainable AI development (Montreal Declaration, 2018, pp. 5-6).

[28] The three pillars of the Ethically Aligned Design designed by the Institute of Electrical and Electronics Engineers Standards Associations (IEEE SA) are (1) the protection of "universal human values", (2) the upholding of "political self-determination and data agency", and (3) the enhancement of "technical dependability (IEEE, 2017, p. 10). Henceforth, the principles outlined here are: "(1) Human Rights, (2) Well-Being, (3) Data Agency, (4) Effectiveness, (5) Transparency, (6) Accountability, (7) Awareness of Misuse, (and, 10) Competence" (IEEE, 2017, p. 11).

Statement on Artificial Intelligence, Robotics and 'Autonomous Systems'[29] (Floridi & Cowls, 2019, pp. 4-5). All these instruments, among others, provide over 47 overlapping principles for transparent and accountable AI development and use (Floridi & Cowls, 2019, p. 4). Nonetheless, all these principles set by different organizations "threaten to become overwhelming and confusing" generating two issues: either the disparate principles "are similar, leading to unnecessary repetition and redundancy, or, if they differ significantly, confusion and ambiguity will" arise generating a "market of principles", where AI developers and providers are led "to shop for the most appealing ones" (Floridi, 2019 as cited in Floridi & Cowls, 2019, p. 2). This situation therefore calls for increased international discussions and cooperation on the issue of AI regulation. Recently, for instance, there has been the third annual AI Action Summit, held in Paris in February 2025 which, despite falling short of achieving a unified global approach to AI, with countries like the US and the UK rejecting stronger international regulation and opposing cooperation with authoritarian regimes like China, served as a vital forum for great discussion and international information exchange (Milmo, 2025).

In Europe, some regulatory frameworks such as the European Union's 2022 Digital Services Act (DSA)[30] have been implemented to enhance transparency and accountability in

---

[29] Commissioned by the European Commission, this Statement "calls for the launch of a process that would pave the way towards a common, internationally recognized ethical and legal framework for the design, production, use and governance of artificial intelligence … based on the values laid down in the EU Treaties and the Eu Charter of Fundamental Rights" (European Group, 2018, p. 5). The ethical principles and democratic prerequisites highlighted by the Statement are: "(a) Human Dignity, (b) Autonomy, (c) Responsibility, (d) Justice, Equity, and Solidarity, (d) Democracy, (f) Rule of Law and Accountability, (g) Security, Safety, Bodily and Mental Integrity, (h) Data Protection and Privacy, (and) (i) Sustainability" (European Group, 2018, pp. 16-19).

[30] Regulation (EU) 2022/2065 of the European Parliament and of the Council 'on a Single Market for Digital Services' alongside Regulation (EU) 2022/1925 or Digital Market Act (DMA) form "the Digital Service Act Package" consisting of comprehensive rules "for the online platforms with the specific purpose of creating a safer digital space where the fundamental rights of users are protected and to establish a level playing field for businesses" (Chiarella, 2023, p. 34). More specifically, the DSA provides a conditional liability exemption framework for intermediary service providers; introduces targeted due diligence obligations for specific categories of intermediaries; lays out rules for enforcement and cooperation among national and EU authorities (Chiarella, 2023, p. 44). It applies horizontally,

content moderation practices (Trujillo et al., 2025). For instance, "Article 15 demands that (every online service providers, and (also) Very large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs)[31] release periodic transparency reports (which) … must contain updated aggregated figures on active recipients, content removals, user appeals, use of automation and AI in moderation, and timeliness of interventions, among other information" (Trujillo et al. 2025, p. 2). For what concerns online terrorist content, Article 21[32] of the 2017 EU Terrorism Directive allows "those responsible for investigating or prosecuting such offence … to make use of effective investigative tools such as those which are used in combating organised crime or other serious crimes … in accordance with national law (and with) the principle of proportionality", while Article 22[33] of the same Directive pushes EU Member States to cooperate with third sources

---

to all services offered within the EU, regardless of where their provider is based. Article 14 requires providers to establish clear terms and conditions on their platforms while Article 15 requires them to publish annual transparency reports on content moderation and complaints (Regulation (EU) 2022/2065, 2022, pp. 49-50). Furthermore, they must implement measures against misuse (Art. 23), explain how their content is recommended and prioritized (Art. 27) and set up internal complaint-handling systems for users to contest platforms decision (Art. 20) (Regulation (EU) 2022/2065, 2022, pp. 53-59).

[31] VLOPs and VLOSEs "which have a number of average recipients of the service in the Union equal to or higher than 45 million" (Art. 33) face stricter rules than other service providers under the DSA (Regulation (EU) 2022/2065, 2022, p. 63). They must provide systems of risk assessment (Art. 34), of risk mitigation (Art. 35 as well as crisis response mechanisms (Art. 36) (Regulation (EU) 2022/2065, 2022, pp. 64-67). Furthermore, they "shall be subject … at their own expense and at least once a year, to independent audits to assess compliance" with the DSA (Art. 37) and "shall provide the Digital Services Coordinator of establishment or the Commission, at their reasoned request and within a reasonable period specified in that request, access to data that are necessary to monitor and assess compliance with this Regulation" (Art. 40) (Regulation (EU) 2022/2065, 2022, pp. 67-72).

[32] In its entirety, Article 21 of the EU Terrorism Directive recites:

> "To ensure the success of investigations and the prosecution of terrorist offences, offences related to a terrorist group or offences related to terrorist activities, those responsible for investigating or prosecuting such offences should have the possibility to make use of effective investigative tools such as those which are used in combating organised crime or other serious crimes. The use of such tools, in accordance with national law, should be targeted and take into account the principle of proportionality and the nature and seriousness of the offences under investigation and should respect the right to the protection of personal data. Such tools should, where appropriate, include, for example, the search of any personal property, the interception of communications, covert surveillance including electronic surveillance, the taking and the keeping of audio recordings, in private or public vehicles and places, and of visual images of persons in public vehicles and places, and financial investigations" (Directive (EU) 2017/541, 2017).

[33] In its entirety, Article 22 of the EU Terrorism Directive recites:

> An effective means of combating terrorism on the internet is to remove online content constituting a public provocation to commit a terrorist offence at its source. Member States should use their best endeavours to

in online content moderation to remove content which promotes terrorist acts. Under the 2016 General Data Protection Regulation (GDPR)[34], then, EU citizens "have the right not to be subject to a decision based solely on automated processing that produces legal effects concerning them or similarly significantly affecting them (Benizri et al., 2023). Even the EUROPOL has published a report on the use of AI systems in law enforcement measures, recognizing both AI use's

> cooperate with third countries in seeking to secure the removal of online content constituting a public provocation to commit a terrorist offence from servers within their territory. However, when removal of such content at its source is not feasible, mechanisms may also be put in place to block access from Union territory to such content. The measures undertaken by Member States in accordance with this Directive in order to remove online content constituting a public provocation to commit a terrorist offence or, where this is not feasible, block access to such content could be based on public action, such as legislative, non-legislative or judicial action. In that context, this Directive is without prejudice to voluntary action taken by the internet industry to prevent the misuse of its services or to any support for such action by Member States, such as detecting and flagging terrorist content. Whichever basis for action or method is chosen; Member States should ensure that it provides an adequate level of legal certainty and predictability for users and service providers and the possibility of judicial redress in accordance with national law. Any such measures must take account of the rights of the end users and comply with existing legal and judicial procedures and the Charter of Fundamental Rights of the European Union (the Charter)" (Directive (EU) 2017/541, 2017).

[34] Regulation (EU) 2016/679 of the European Parliament and of the Council on 'on the protection of natural persons with regard to the processing of personal data and on the free movement of such data' has been issued

> "In order to ensure a consistent level of protection for natural persons throughout the Union and to prevent divergences hampering the free movement of personal data within the internal market, a Regulation is necessary to provide legal certainty and transparency for economic operators, including micro, small and medium-sized enterprises, and to provide natural persons in all Member States with the same level of legally enforceable rights and obligations and responsibilities for controllers and processors, to ensure consistent monitoring of the processing of personal data, and equivalent sanctions in all Member States as well as effective cooperation between the supervisory authorities of different Member States. The proper functioning of the internal market requires that the free movement of personal data within the Union is not restricted or prohibited for reasons connected with the protection of natural persons with regard to the processing of personal data. To take account of the specific situation of micro, small and medium-sized enterprises, this Regulation includes a derogation for organisations with fewer than 250 employees with regard to record-keeping. In addition, the Union institutions and bodies, and Member States and their supervisory authorities, are encouraged to take account of the specific needs of micro, small and medium-sized enterprises in the application of this Regulation. The notion of micro, small and medium-sized enterprises should draw from Article 2 of the Annex to Commission Recommendation 2003/361/EC" (Art. 13) (Regulation 2016/679, 2016, p. 3).

Entered into force on 25 May 2018, the GDPR "equalized the rules of data protection" among EU Member States , thus eradicating "the fragmentation of data protection across EU Member States and the resulting legal uncertainties (which) were considered (as) an obstacle to the pursuit of economic activities at EU level and (as) a distortion of competition" (Voigt & Von Dem Bussche, 2017p. 2). The GDPR provides some key organizational requirements which all the companies working within the EU Single Market must respect, namely: (1) provision of "records of processing activities", as per Art. 30; (2) "designation of a Data Protection Officer", as per Artt. 37-39; (3) provision of "Data Protection Impact Assessments (DPIA)", as per Artt. 35-36; (4) "data protection by design and by default", as per Art. 25; (5) "data subject rights", as per Artt. 12-23; (6) "data breach notification", as per Artt. 33-34 (7) "appointment of a representative by NON-EU entities", as per Art. 27; and (8) "codes of conduct and certification" as per Artt. 40 and 42 (Voigt & Von Dem Bussche, 2017, pp. 3-5).

advantages and disadvantages, and calling for greater regulation of such systems (Europol, 2024, pp. 9-12). Furthermore, the Council of Europe has created the *Framework Convention on Artificial Intelligence and Human Rights*, trying to regulate the use of AI systems, and calling for increased legal safeguards and robust impact assessments (Council of Europe, n.d.). Meanwhile, the European Commission's AI strategy has given birth to the EU Artificial Intelligence Act, the first comprehensive attempt to regulate AI in a rights-based and risk-sensitive manner worldwide, which aims to set global standards on the regulation of AI (European Commission, 2025a). The Act has been the result of all the calls for action at the European level, trying to ensure that AI development and employment goes hand in hand with the protection of users' fundamental rights.

2.5 Can AI Effectively Protect Both Security and Freedom of Speech When Used in Counterterrorism?

The question "can AI effectively protect both security and freedom of speech when used in counterterrorism?" lies at the heart of this thesis. The increasing use of AI systems for security purposes, in fact, carries with it great risks in light of the many shortcomings possessed by AI systems and which still have not been addressed. These systems present, indeed, both "alignment (and) control problem(s)": "as AI systems become more capable", the dangers of "misalignment, when the behavior of the system does not align with the values of its human creators", escalate, and it becomes increasingly harder to "actively manage and regulate the(ir) behavior" (Judge et al., 2025, p. 91). Without changes to the biases of AI systems and to their content moderation systems, indeed, AI risks becoming a tool not just for counterterrorism, but for the suppression of dissent, activism, and minority voices. Ethical considerations should therefore be included both in the development and in the use of AI systems for counterterrorism purposes, keeping in mind

"the potential for AI to harm human values and interests" (Esmailzadeh & Motaghi, 2024, p. 178). While AI systems have great potential in this field, meaningful regulations are needed to ensure the balance between security and freedom of speech.

Henceforth, AI regulation should be increased, in an effort to enact "clear guidelines" for its use and development and reduce its malicious use (Esmailzadeh & Motaghi, 2024, p. 178). Not only this, but also to promote the development of safer and more reliable technologies the regulatory frameworks should be developed (Judge et al., 2025, p. 91). For what concerns the protection of fundamental rights, "transparency in data collection and clear consent" is of the utmost importance: "implementing strong security measures and encryption (in fact) is a shared responsibility to protect" users from human rights breaches from AI systems (Siagian et al., 2023, pp. 556-557). Furthermore, "awareness of individual rights in the digital environment and education about data protection practices" should be meaningfully enhanced, so as to ensure that individuals know their rights not only in the physical but also in the digital world (Siagian et al., 2023, p. 557). Education and awareness of users should hence be incremented, "through educational programs, public awareness campaigns, and media coverage" which allow them to make "more informed decisions about" AI use (Esmailzadeh & Motaghi, 2024, p. 178). Concurrently, human oversight of AI systems' decision-making processes should be increased as should the redress mechanisms for ensuring that content wrongfully removed may be re-posted. Hence, a holistic approach to this issue is urgently needed, relying on the cooperation of "governments, technology companies, civil society, and individuals … to develop policies and practices that ensure freedom of speech and human rights" guarantees in the new Industry 5.0 (Siagian et al. 2023, p. 555). To ensure that AI systems serve the public good, while preserving democratic values, human dignity and fundamental rights, especially when used in the field of

counterterrorism, Floridi and Cowls (2019, pp. 5-8) have underlined five fundamental principles which AI developers and providers should strive to uphold, namely: (1) the principle of beneficence, to promote "the well-being of people and the planet with AI"; (2) the principle of non-maleficence, to prevent privacy infringements, security breaches, misuse and all the negative consequences of AI systems; (3) the principle of autonomy, to balance the decision-making capabilities of AI systems with that of humans and ensure that humans retain their power to decide when and how to delegate decisions to AI; (4) the principle of justice, to enhance fairness and equitable access to AI while also eliminating possible discrimination; and finally, (5) the principle of explicability, to make individuals "understand and hold accountable the decision-making processes of AI".

Only with all these safeguards, AI systems will be capable of being employed for counterterrorism purposes without the risk of impinging upon fundamental rights, like the right to freedom of speech. Henceforth, a holistic approach to AI in counterterrorism should be promoted at the international level and implemented, balancing "security with privacy, human rights, and social justice" (Esmailzadeh & Motaghi, 2024, pp. 177-178). Such a holistic approach should also take in consideration the evolving reality in which artificial intelligence tools are being developed. In less than half a century, society has transitioned from an analogue one to a digital society, and now to a cybersociety: such a rapid transformation has deeply challenged traditional notions of the state, the separation of powers, and the very foundation of fundamental rights, which must all be reinterpreted through this new socio-technological perspective. The 'silent' and 'friendly' revolution brought about by the digitalization of life and by the employment of artificial intelligence has created a digital democracy which no longer can be controlled by any existing constitutions which, as Balaguer argues in *La Constitución del Algoritmo*, regulate "a world that

partly no longer exists or is socially irrelevant" (Balanguer, 2023, p. 14 as cited in Palombino, 2022, pp. 1-2). Balaguer thus calls for the adoption of a "Constitution of the Algorithm" capable of "analyzing digital reality from the perspective of the ruptures it is generating, which have a constitutional dimension", as well as of "proposing solutions that can mitigate these ruptures and facilitate a constitutional response" (Balanguer, 2023, pp. 16–17 as cited in Palomino, 2022, p. 2). According to Balanguer, "five distinct constitutional ruptures" justify this call, namely:

> "(1.) the rupture of the cultural context of the constitution", since "the rules governing technology companies are based on priorities that do not correspond to those protected by the Constitution", mainly revolving around profit-driven logics;
>
> "(2.) the rupture between physical and virtual reality, the latter of which is not capable of guaranteeing the protection of individual freedoms";
>
> "(3.) the rupture in the configuration of reality, with "the destruction of a shared social perception of reality, brought about by new modes of communication and the biased nature of those who mediate them" and, consequently,
>
> "(4.) the rupture of a unified cultural reference" and "(5) of the economic constitution", which reflect existing constitutions' inability to alt "the algorithm's intrusion into the protection of rights and the legitimization of power" and the "globalization processes and the growing economic power of non-state actors" (Palomino, 2022, pp. 2-4).

Recognizing the urgency of these developments, the European Union has enacted the EU AI Act, with the aim of promoting a human-centered and transparent approach to artificial intelligence which respects both security imperatives and fundamental rights. Fundamentally, this AI governance strategy will be unified among all 27 EU Member states, but it will also possibly set global standards in this rapidly evolving field.

3. The Artificial Intelligence Act: A New Digital Legal Framework for the European Union

3.1 What is the EU Artificial Intelligence Act (EU AI Act)?

While there are certainly many advantages to the employment of AI systems in everyday life for both people and companies, their unregularized use carries with it also major risks, like privacy violations, inappropriate data collection and retention, or the creation of deep fakes which may endanger even the national security of a country. Consequently, regulating AI appears as a necessary step towards ensuring that technological advances align with core societal values and with the protection of human rights, especially when AI systems are used for security purposes. For this reason, the European Commission enacted a legislative proposal in April 2021 with the scope of guaranteeing the protection of fundamental rights and of ensuring safety, transparency and accountability in AI development and use in Europe (Krüger, 2025). The proposal aimed to support both innovation and investment in the field of AI development, while positioning the European Union as a global leader in human-centered and trustworthy AI (Krüger, 2025). Indeed, when the EU AI Act (Regulation (EU) 2024/1689 laying down harmonized rules on artificial intelligence) was enacted on August 1, 2024, the European Commission prided itself describing it as “the first-ever comprehensive legal framework on AI worldwide” (European Commission, 2025a). To this end, the Act provides a comprehensive set of definitions in Art. 3 relating to the AI environment, the most important of which is that of an AI system, described 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 input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments" (Regulation (EU) 2024/1689, 2024, p. 46).

Providing specific definitions of all the words connected with the AI environment is of great importance, especially if one considers that most EU members have not approved a national legal definition of AI or any laws regulating its use and development prior to the adoption of the AI Act.

The Act, which will officially apply from August 2, 2026, introduces a risk-based approach to regulating AI systems, creating a four-tiered risk pyramid which categorizes systems as carrying, from the bottom up, minimal, limited, high, or unacceptable risk (European Commission, 2025a). Most AI systems, like spam filters, fall within the minimal risk category and are not subject to specific regulation (European Commission, 2025a). Those systems which fall instead within the limited risk category, such as Chatbots and synthetic media (e.g., deep fakes) but also Generative AI (e.g., AI-generated texts or images) are obliged to comply with transparency obligations, clearly labeling and disclosing to their users each time they are interacting with AI (European Commission, 2025a). Going up the pyramid, there are those systems employing AI for measures carrying a high risk of negatively impacting fundamental rights or safety, including when AI is used in sectors such as education, employment, law enforcement, medicine, migration, and the judiciary; providers of high risk systems must meet onerous obligations, providing risk management strategies and high-quality datasets, human oversight and transparency, documentation and traceability, as well as robust cybersecurity and accuracy standards (European

Commission, 2025a). At the top of the risk pyramid, there are eight recognized practices which have been banned for the unacceptable risk they carry for their users,

> "namely: (1) harmful AI-based manipulation and deception, (2) harmful AI-based exploitation of vulnerabilities, (3) social scoring, (4) individual criminal offence risk assessment or prediction, (5) untargeted scraping of the internet or Closed-Circuit Television (CCTV) material to create or expand facial recognition databases, (6) emotion recognition in workplaces and education institutions, (7) biometric categorisation to deduce certain protected characteristics, (and finally) (8) real-time remote biometric identification for law enforcement purposes in publicly accessible spaces" (European Commission, 2025a).

While the Act has still not entered into force, the Commission has launched the AI Pact, a voluntary initiative inviting AI providers and deployers to start implementing key AI Act obligations before the formal application of the whole Act; furthermore, some provisions have already entered into force or will do so very soon, including the part of the Act dealing with 'prohibited practices and AI literacy" (in force since February 2, 2025) and the General-Purpose AI (GPAI) rules which will become applicable on August 2, 2025 (European Commission, 2025a). Moreover, to facilitate the implementation at the EU level of the Act, the President of the European Commission, Ursula von der Leyen, approved, on February 2025, "*InvestAI*, an initiative to mobilise €200 billion for investment in AI, including a new European fund of €20 billion for AI gigafactories … (and to render) Europe an AI continent" (Hickman & Lorenz, 2024). More recently, in April 2025, the European Commission has also launched the *AI Continent Action Plan* to bolster AI development across the EU and facilitate the smooth implementation of the Act, focusing on five key areas:

> "(i) building a large-scale AI computing infrastructure; (ii) increasing access to high-quality data; (iii) promoting AI in strategic sectors; (iv) strengthening AI skills and talents; and (v) simplifying the implementation of the EU AI Act" (Hickman & Lorenz, 2024).

All in all, the Act is a truly revolutionary and unique regulatory framework for the development, deployment and use of AI, which renders the European Union a global leader in the promotion of human-centered and trustworthy AI (Krüger, 2025). As a matter of fact, by adopting a risk-based approach which safeguards both fundamental rights and democratic values, the AI Act will not only ensure that AI systems entering or developed within the European market are safe and transparent, but it will also offer a model that may inspire similar regulatory efforts in other countries which seek to offer in their markets the same rights-based safeguards for their citizens on the development and use of AI systems.

This chapter first examines the Act's impact on big tech companies and digital regulation, and then its implementation within the national jurisdiction of four key EU Member States. In doing so, it will consider how Italy, Germany, France and Spain are planning on enacting the Act as countries which have already been using AI systems for security purposes without clear regulations: this allows to understand how the AI Act may be implemented to fill existing regulatory gaps within the Member States' jurisdictions. Moreover, these four countries are among the largest and most influential within the European Union, and their implementation of the AI Act will likely provide examples to other EU Member States; finally, within these four countries, relevant discussions on AI development and use have already been faced, especially in matters concerning digital rights, state surveillance, and ethical use of AI for security purposes. Subsequently, the chapter moves to the fundamental problem dealt with by this thesis, considering how this Act and its implementation can impact the balance between security imperatives and the

protection of fundamental rights in the field of online counterterrorism, concluding with some recommendations for public authorities to ensure that the Act is applied upholding this delicate balance.

3.2 The AI Act's Impact on Big Tech and Digital Regulation

Since the EU AI Act applies extraterritorially, its provisions target both EU-based and foreign-based companies operating within the EU market and employing AI in their work (Art. 2(1)) (Regulation (EU) 2024/1689, 2024, p. 45). It thus affects mainly technology companies working in the field of artificial intelligence and obliges them to assume a high degree of accountability, especially since non-compliance can result in administrative fines of up to €35 million or 7% of the company's total worldwide annual turnover, whichever is higher (Art. 5) (Regulation (EU) 2024/1689, 2024, pp. 51-53). The extraterritorial aspect of this Act propels Europe to become the world's leader in regulating AI use and development, since every company which carries out its business in the Old Continent must respect its rules, much like it happened when Europe approved the GDPR and extended its regulatory influence beyond its borders. Clearly, where democratic values and human rights are at risk due to new technologies, the European Union strives to position itself as their protector, increasingly taking on a leading global role in the protection of digital rights.

The Act draws particular attention when dealing with high-risk AI systems, being those the ones most at risk of impinging upon fundamental rights. In particular, Art. 16[35] lists all the

[35] The full Art. 16 recites: "Providers of high-risk AI systems shall: (a) ensure that their high-risk AI systems are compliant with the requirements set out in Section 2; (b) indicate on the high-risk AI system or, where that is not possible, on its packaging or its accompanying documentation, as applicable, their name, registered trade name or registered trade mark, the address at which they can be contacted; (c) have a quality management system in place which complies with Article 17; (d) keep the documentation referred to in Article 18; (e) when under their control,

obligations of providers of high-risk AI systems, including ensuring compliance with the Act's provisions, taking corrective action where necessary to guarantee the correct application of the Act, and maintaining updated technical documentation (Regulation (EU) 2024/1689, 2024, p. 62). Art. 53[36], then, lists all the obligations of GPAI models, which must respect almost all the same

keep the logs automatically generated by their high-risk AI systems as referred to in Article 19; (f) ensure that the high-risk AI system undergoes the relevant conformity assessment procedure as referred to in Article 43, prior to its being placed on the market or put into service; (g) draw up an EU declaration of conformity in accordance with Article 47; (h) affix the CE marking to the high-risk AI system or, where that is not possible, on its packaging or its accompanying documentation, to indicate conformity with this Regulation, in accordance with Article 48; (i) comply with the registration obligations referred to in Article 49(1); (j) take the necessary corrective actions and provide information as required in Article 20; (k) upon a reasoned request of a national competent authority, demonstrate the conformity of the high-risk AI system with the requirements set out in Section 2; (l) ensure that the high-risk AI system complies with accessibility requirements in accordance with Directives (EU) 2016/2102 and (EU) 2019/882" (Regulation (EU) 2024/1689, 2024, p. 62).

[36] The full Art. 53 recites: "2. The obligations set out in paragraph 1, points (a) and (b), shall not apply to providers of AI models that are released under a free and open-source licence that allows for the access, usage, modification, and distribution of the model, and whose parameters, including the weights, the information on the model architecture, and the information on model usage, are made publicly available. This exception shall not apply to general-purpose AI models with systemic risks. 3. Providers of general-purpose AI models shall cooperate as necessary with the Commission and the national competent authorities in the exercise of their competences and powers pursuant to this Regulation. 4. Providers of general-purpose AI models may rely on codes of practice within the meaning of Article 56 to demonstrate compliance with the obligations set out in paragraph 1 of this Article, until a harmonised standard is published. Compliance with European harmonised standards grants providers the presumption of conformity to the extent that those standards cover those obligations. Providers of general-purpose AI models who do not adhere to an approved code of practice or do not comply with a European harmonised standard shall demonstrate alternative adequate means of compliance for assessment by the Commission. 5. For the purpose of facilitating compliance with Annex XI, in particular points 2 (d) and (e) thereof, the Commission is empowered to adopt delegated acts in accordance with Article 97 to detail measurement and calculation methodologies with a view to allowing for comparable and verifiable documentation. 6. The Commission is empowered to adopt delegated acts in accordance with Article 97(2) to amend Annexes XI and XII in light of evolving technological developments. 7. Any information or documentation obtained pursuant to this Article, including trade secrets, shall be treated in accordance with the confidentiality obligations set out in Article 78. 2. The obligations set out in paragraph 1, points (a) and (b), shall not apply to providers of AI models that are released under a free and open-source licence that allows for the access, usage, modification, and distribution of the model, and whose parameters, including the weights, the information on the model architecture, and the information on model usage, are made publicly available. This exception shall not apply to general-purpose AI models with systemic risks. 3. Providers of general-purpose AI models shall cooperate as necessary with the Commission and the national competent authorities in the exercise of their competences and powers pursuant to this Regulation. 4. Providers of general-purpose AI models may rely on codes of practice within the meaning of Article 56 to demonstrate compliance with the obligations set out in paragraph 1 of this Article, until a harmonised standard is published. Compliance with European harmonised standards grants providers the presumption of conformity to the extent that those standards cover those obligations. Providers of general-purpose AI models who do not adhere to an approved code of practice or do not comply with a European harmonised standard shall demonstrate alternative adequate means of compliance for assessment by the Commission. 5. For the purpose of facilitating compliance with Annex XI, in particular points 2 (d) and (e) thereof, the Commission is empowered to adopt delegated acts in accordance with Article 97 to detail measurement and calculation methodologies with a view to allowing for comparable and verifiable documentation. 6. The Commission is empowered to adopt delegated acts in accordance with Article 97(2) to amend Annexes XI and XII in light of evolving technological developments. 7.

rules of the providers of high-risk AI systems, but benefit from more freedom since their systems do not pose serious risks to fundamental rights (Regulation (EU) 2024/1689, 2024, pp. 84-85). Among providers which supply GPAI models are companies such as Google, Meta, Microsoft, OpenAI and even Amazon, all of which must rethink their AI development strategies if they want to keep operating in Europe. Indeed, these providers work extensively in sectors like consumer service and advertising, where a wrongful application of AI can result in consumers' privacy and fundamental rights' breaches. It is therefore important that these companies regularly submit reports, conduct impact assessments and develop risk management systems, all in line with the EU AI Act. A company that respects the provisions of the Act, indeed, signals its clear intentions of adhering to standards that prioritize an ethical and human-centered approach to AI as well as the protection of fundamental rights.

3.3 The Implementation of the AI Act in the Member States

The implementation of the AI Act in the European Member States aims at generating a common framework and uniform standards of "clarity and consistency" across all the EU countries, thus facilitating "cross-border operations" and equipping AI system providers with a strong foundation to develop and implement GPAI model-solution (Krüger, 2025). Obviously, promoting a continent-wide and clear legal framework not only favors more awareness in the innovation field within Europe, but it also pushes international AI providers to invest in Europe, where a stable and legalized environment can be helpful in testing and developing new technologies in a safe manner (Krüger, 2025),

---

Any information or documentation obtained pursuant to this Article, including trade secrets, shall be treated in accordance with the confidentiality obligations set out in Article 78" (Regulation (EU) 2024/1689, 2024, pp. 84-85).

To oversee the enforcement and the implementation of the AI Act, the European Commission established, in February 2024, the European AI Office which, in so doing, must cooperate alongside the national authorities designated to supervise compliance with the Act within each Member State (European Commission, 2025b). The AI Act requires Member States, with great discretion and autonomy, to designate three national authorities to check the enactment of the Act: in particular, under Art. 3(26), a Market Surveillance Authority is needed to ensure "that only products compliant with EU law are made available on the Union market" based on Regulation (EU) 2019/1020[37]; under Artt. 3(19) and 28(1), a Notifying Authority is needed "for establishing and performing the procedure for assessment, designation and notification of conformity assessment bodies and for their monitoring"; lastly, under Art. 77(2), a National Public Authority is needed to "enforce the respect for fundamental rights obligation in Member States in relation to high-risk AI systems referred to in Annex III" (Overview, 2025). The authorities dealing with the implementation of the Act must all be designated by August 2, 2025; nevertheless, 16 EU members, among which France, still lack all three authorities, while all the other countries have appointed at least one authority (Overview, 2025). On the other hand, the authorities responsible for the protection of fundamental rights had to be appointed by November 2, 2024, but only 13 EU Members have actually chosen the designated authority, signaling a grave delay in ensuring fundamental rights' protection in most of the European countries (Overview, 2025). While this decentralized approach allows for flexibility, it also increases the risk of fragmented application, especially in politically sensitive areas such as counterterrorism and freedom of speech online. It

[37] Regulation (EU) 2019/1020 of the European Parliament and of the Council 'on market surveillance and compliance of products' "shall apply to products that are subject to the Union harmonisation legislation listed in Annex I ('Union harmonisation legislation'), in so far as there are no specific provisions with the same objective in the Union harmonisation legislation, which regulate in a more specific manner particular aspects of market surveillance and enforcement" (Art. 1(1)) (Regulation (EU) 2019/1020, 2019, p. 12). It indeed strive to ensure product safety, by improving the capabilities of the market surveillance authorities.

is thus important to understand how EU Member States are dealing with the implementation of the AI Act.

3.3.1 Italy's Path to AI Governance: From Regulatory Vacuum to the EU AI Act

Italy does not have any specific laws or regulations on the matter of AI and will therefore strongly rely on the EU AI Act to implement a just and transparent legal framework governing artificial intelligence (White & Case, 2024). However, the Council of Ministers approved, on April 23, 2024, the draft law n. 1146 (DDL – Disegno di Legge sull'Intelligenza Artificiale) on artificial intelligence, a DDL which is way too general in nature perhaps because the country did not want to create unnecessary inconsistencies with the upcoming AI Act (Capone & Iaselli, 2024). Importantly, however, the DDL, composed of 26 articles, aims to "identify the principles concerning the research, experimentation, development, adoption, and application of artificial intelligence systems and models, and to promote the correct, transparent, and responsible use of artificial intelligence" (Capone & Iaselli, 2024). Connected to the topic of this thesis are Articles 3, 4 and 6 of the DDL: in particular, Art. 3[38] emphasizes the respect for fundamental rights

[38] Article 3 on 'General Principles" of the DDL n. 1146 recites:

1. "The research, experimentation, development, adoption, application, and use of artificial intelligence systems and models must take place in compliance with the fundamental rights and freedoms established by the Constitution, by European Union law, and in accordance with the principles of transparency, proportionality, security, protection of personal data, confidentiality, accuracy, non-discrimination, gender equality, and sustainability.
2. The development of AI systems and models must be based on data and processes whose correctness, reliability, security, quality, appropriateness, and transparency must be ensured and monitored, according to the principle of proportionality in relation to the sectors in which they are used.
3. AI systems and models must be developed and applied in respect of human autonomy and decision-making power, harm prevention, explainability, and the principles set out in paragraph 1.
4. The use of AI systems must not compromise the democratic functioning of institutional and political life.
5. In order to ensure respect for the rights and principles in this article, cybersecurity must be guaranteed throughout the entire lifecycle of AI systems and models as an essential precondition, following a proportional and risk-based approach. This includes the adoption of specific security controls to ensure resilience against attempts to alter their use, expected behavior, performance, or security settings.
6. This law ensures that persons with disabilities have full access to AI systems and their functionalities or extensions, on an equal basis and without any form of discrimination or prejudice, in compliance with the

enshrined in the Italian Constitution and EU law in the "development, … application, and use of artificial intelligence systems and models"; Art. 4[39] reinforces the commitment to the protection of the right to freedom of speech, while requiring AI systems to uphold transparency measures in their processing of data; Art. 6[40], then, deals with the deployment of AI in national security and defense, and excludes the activities related to intelligence, cybersecurity and the military, from the scope of the DDL, however pushing them to always uphold the general principles expressed in

provisions of the United Nations Convention on the Rights of Persons with Disabilities, adopted in New York on December 13, 2006, and ratified in Italy under Law No. 18 of March 3, 2009" (DDL n. 1146, 2024).

[39] Article 4 on 'Principles on Information and Personal Data Confidentiality) of the DDL n. 1146 recites:

1. "The use of AI systems in the field of information must occur **without undermining the freedom and** pluralism of the media, freedom of expression, and the objectivity, completeness, impartiality, and fairness of information.
2. The use of AI systems must ensure the lawful, fair, and transparent processing of personal data, and compatibility with the purposes for which such data were collected, in accordance with European Union law on personal data and privacy protection.
3. Information and communications regarding data processing associated with the use of AI systems must be conveyed in clear and simple language, to ensure the user's full understanding and the ability to object to the improper use of their personal data.
4. Access to AI technologies by minors under the age of fourteen requires the consent of those holding parental responsibility. Minors aged fourteen and over may give their own consent for the processing of personal data associated with the use of AI systems, provided that the information and communications referred to in paragraph 3 are easily accessible and understandable" (DDL n. 1146, 2024).

[40] Article 6 on 'Provisions on National Security and Defense' of the DDL n. 1146 recites:

1. "The activities referred to in Article 3, paragraph 1, carried out for national security purposes in accordance with the goals and methods outlined in Law No. 124 of August 3, 2007, by the bodies identified in Articles 4, 6, and 7 of the same law, as well as cybersecurity and resilience activities described in Article 1, paragraph 1, letters a) and b) of Decree-Law No. 82 of June 14, 2021, converted with amendments by Law No. 109 of August 4, 2021, performed by the National Cybersecurity Agency to protect national security in cyberspace, and also those carried out for national defense purposes by the Armed Forces, are excluded from the scope of this law. However, these activities must still be conducted in compliance with the fundamental rights and freedoms established by the Constitution, and in accordance with Article 3, paragraph 4.
2. The development of AI systems and models must respect the conditions and purposes outlined in Article 3, paragraph 2. The processing of personal data through the use of AI systems by the bodies referred to in Articles 4, 6, and 7 of Law No. 124 of 2007 is subject to Article 58, paragraphs 1 and 3, of the Personal Data Protection Code, pursuant to Legislative Decree No. 196 of June 30, 2003. The processing of personal data through the use of AI systems by the National Cybersecurity Agency is governed by Article 13 of the aforementioned Decree-Law No. 82 of 2021.
3. By means of a regulation adopted pursuant to Article 43 of Law No. 124 of 2007, the implementation procedures of the principles and provisions of this article will be defined, as they apply to the activities mentioned in Article 3, paragraph 1, carried out by the bodies identified in Articles 4, 6, and 7 of Law No. 124 of 2007, as well as to AI-related activities functional to the operations of those bodies and performed by other public or private entities exclusively for national security purposes. Similarly, for the National Cybersecurity Agency, this will be done through a regulation adopted in accordance with Article 11, paragraph 4, of Decree-Law No. 82 of 2021" (DDL n. 1146, 2024).

Art. 3. These provisions, and the DDL more in general, delineate Italy's high commitment to develop a rights-based approach to AI, even when these systems are used for security and counterterrorism purposes, and will be pivotal for the national implementation of the AI Act.

To ensure an adequate implementation of the AI Act, the Italian government will also have to consider "the GDPR and the Italian Data Privacy Code (Legislative Decree no. 196/2003), (the) EU and Italian competition law (Articles 101 and 102 TFEU and Law no. 287/1990), the Italian Consumer Code (Legislative Decree no. 206/2005)", and even the aforementioned DDL n. 1146, all laws which may indirectly affect AI systems as well as their regulation and application (White & Case, 2024). In particular, the Italian Data Protection Authority (Garante per la Protezione dei Dati Personali - DPA), who supervises Italy's compliance with the GDPR, has already been vociferous with respect to the implementation of the AI Act and has taken concrete actions to prevent privacy violations and promote a transparent and ethical use of AI systems in Italy. For instance, in March 2022, he has banned the use of biometric data by sanctioning the US-based company Clearview AI for €20 million for having applied biometric monitoring techniques on the Italian territory, obtaining Italians' personal data without informing users and also without respecting any purpose and storage limitations (GPDP, 2022). Furthermore, the Italian DPA has underlined some key principles to facilitate the Act's transparent and ethical enactment, emphasizing the importance of "data protection when AI is used in healthcare services"; identifying "tax risks from the point of view of combating tax evasion"; and recognizing the potential use of "facial recognition (technologies) for law enforcement purposes (White & Case, 2024).

In light of these principles, the *Italian Strategy for Artificial Intelligence 2024-2026* has been enacted, considering the potential of AI in boosting national productivity, public service

efficiency, healthcare, environmental sustainability, demographic decline and national security as markedly high (AGID, 2024). The *Strategy*, advocating for a human-centered development of AI, emphasizes the need for investing in AI research and infrastructure, for public-private partnerships to drive innovation, and for finding tailored AI solutions adapted to Italy's unique industrial and social fabric (AGID, 2024). It has indeed designated the sectors of research, public administration, education, and enterprises as the four strategic areas of AI development, underscoring the importance of AI's future role in the Made in Italy industry, the Digital and Financial industry, the protection of the territory and also of privacy and security of individuals (AGID, 2024). Moreover, unlike many other EU members, Italy has already chosen the National Cybersecurity Agency (Agenzia per la Cybersicurezza Nazionale, ACN) as the Act's Market Surveillance Authority and the Agency for Digital Italy (Agenzia per l'Italia Digitale, AgID) as its Notifying Authority, positioning itself at the forefront in the enactment of the AI Act in Europe (Overview, 2025).

Notwithstanding Italy's steps forward regarding the implementation of the AI Act on its territory, this process risks, if not carried out carefully, to receive great criticism. This is mainly because Italy has been frequently using AI for security purposes, especially in the context of counterterrorism and of monitoring online radicalization after the 2015 terrorist attacks in Europe (US Department of State, 2016). For instance, Decree Law No. 7 of February 18, 2015 (*Urgent Measures for the Fight against Terrorism, Including International Terrorism)* strengthened the government's "authority to collect personal data related to the perpetration of terrorist crimes" and "to instruct internet service providers to block access to websites identified by authorities as being used for terrorist recruitment activities" (US Department of State, 2016). The National Police even acquired an Automatic Image Recognition System (SARI), to use facial recognition softwares to combat crime, but the Italian DPA, alongside other NGOs, has lamented the fact that this system

"would implement a form of mass surveillance" thus banning its use (Carrer & Coluccini, 2021). Furthermore, this unregulated way of implementing AI for counterterrorism may result in disproportionate restrictions on the right to freedom of speech, which is not only protected by both the EU Charter and the ECHR but also by Article 21[41] of the Constitution of the Italian Republic which recites: "Everyone has the right to freely express their ideas through speech, in writing and by any other means of communication" (Constitution of the Italian Republic, Article 21). While "in exceptional cases of necessity and urgency, strictly defined by law", the right to freedom of speech, as all other personal freedoms, may be restricted, it is important that in all the cases where AI is used for counterterrorism purposes, the Italian government does not exploit Art. 13 without the necessary safeguards, impinging too much or for too long on Italians' fundamental rights (Constitution of the Italian Republic, Article 13).

In light of all this, it appears clear that Italy will have to calibrate its use of AI for security purposes with the new goals of the EU AI Act which will require certain AI systems, incompatible with fundamental rights, to be banned as well as greater transparency from both developers and providers.

[41] The whole text of Article 21, dealing also with the freedom of press, recites:

> "Everyone has the right to freely express their ideas through speech, in writing and by any other means of communication.
> The press shall not be subjected to authorization or censorship.
> Seizure shall be permitted only by a measure for which reasons must be stated issued by the judicial authority, in the case of offences for which the law governing the press grants express authorization, or in the case of violation of its provisions concerning the disclosure of the identity of those responsible for such offences.
> In such cases, when it is a matter of extreme urgency and when prompt intervention of the judicial authority is not possible, periodical publications may be seized by officers of the judicial police, who shall immediately, and in any case within twenty-four hours, report the matter to the judicial authority. If the latter does not confirm the seizure order within the following twenty-four hours, the seizure shall be deemed to be withdrawn and null and void.
> The law may introduce general provisions for the disclosure of the financial sources of the periodical publications.
> Printed publications, public performances and any other events contrary to public decency are forbidden. The law shall provide for appropriate measures for the prevention and repression of all violations". (Constitution of the Italian Republic, Article 21).

3.3.2 Germany's Implementation of the AI Act: Balancing Innovation and Regulation

Over the past decade, Germany has developed a national AI approach which strives to uphold ethical standards while ensuring further industrial innovation and international competitiveness in the sector. In 2014, under Chancellor Angela Merkel's leadership, Germany introduced its *Digital Agenda 2014-2017* recognizing, for the first time at the national level, the sector of AI as a critical area for investment and national economic growth (Global Institute for National Capability, 2024). Considering AI as a fundamental mean to accelerate Germany's "cross-departmental innovation strategy", the *Agenda* was then renovated in 2017, with a great emphasis on innovating and expanding Germany's AI infrastructure primarily in the industrial sector[42], but also in other key developmental sectors such as energy, education, public administration, and healthcare (Global Institute for National Capability, 2024), highlighting that "digitization and interconnectivity in these areas help to boost the productivity of the basic systems used in (the German) community" (The Federal Government, 2014, pp. 13-14). The 2017 *Agenda* also supported projects like the 2015 *High-Tech Strategy*, boosting Germany's funds for AI

[42] With regard to the digitalization of the industrial sector, the Digital Agenda aims to develop: "the establishment and expansion of research and technology programmes with high transferability to industry, for example, the areas of autonomic technology, 3D, big data, cloud computing and microelectronics; the initiation of new business models and innovative services by fostering the development and distribution of big data and cloud applications that offer greater security and data privacy; reinforcing security and confidence in relation to the use of digital services, including measures to strengthen the German digital security sector; assisting small and medium-sized IT enterprises with their internationalisation efforts and facilitating their access to growth capital; the promotion of norms and standards to ensure the seamless integration of traditional industry with ICT"; furthermore, with regard to the application of AI in other key sectors of development, the Agenda specifies that it will strive in "developing centres of excellence to provide information and demonstrations of best practice for Industry 4.0 and smart services to the SME and skilled craft sector and also supporting user-friendly applications and services (usability); supporting smart home applications; facilitating ICT-based support for (electro-)mobility; supporting the digital transformation in the media and creative industries, opening up sizeable opportunities for new customer groups; promoting sustainability and climate protection (environmental awareness in IT and in the use of IT); supporting the digitisation of construction; further developing legal specifications for the integration of telemedicine; expanding the eHealth initiative, enhancing links with the innovations delivered by health care businesses and ensuring the interoperability and security of IT systems; … drafting an "Intelligent Connections" strategy to create additional opportunities for growth and efficiency through ICT in the education, energy, health, transport and administration primary industries. (The Federal Government, 2014 pp. 13-14).

innovation and development, and the *Industry 4.0 Project*, "integrating AI into automated production systems, supply chain management, and smart manufacturing" (Global Institute for National Capability, 2024). Following this, in November 2018, Germany developed its *National AI Strategy*, committing €3 billion, increased to €5 billion in 2020, to promote AI research and development by 2025 while ensuring an ethical and legal application of AI in both business and society, a commitment further reinforced by subsequent initiatives such as the 2019 *AI Made in Germany Strategy* and the 2020 *AI Standards for Industry 4.0* (European Commission, 2021).

While Germany seemed to be at the forefront in Europe in the field of AI innovation and development, its approach remained mostly policy oriented, but policy makers did not even bother creating a unified definition of "AI" under German law, aligning with the definition provided in the EU Council's December 2022 *Mandate for the EU AI Act* (Lorenz, 2024). Prior to the enactment of the EU AI Act, indeed, the country lacked a comprehensive legal framework addressing artificial intelligence, with references to AI solely within German Labor Law and more specifically within the Works Constitution Act of 25 September 2001 (Betriebsverfassungsgesetz – BetrVG), as amended by Article 1 of the Act of 19 July 2024 (Works Constitution Act, 2024). The Act deals with the topic of AI in its Section 80(3), Section 90(1) No. 3 and Section 95(2a), in particular: section 80(3) requires that, in all the cases in which AI is used by the works council to "carry out its tasks" an expert of AI must be involved to give advice on best practices; section 90(1) No. 3 deals more specifically with the role of the employer, and states that where the latter wants to introduce new "working procedures and operations including the use of artificial intelligence", they must inform the works council; finally, Section 95(2a), making Section 95(1) and (2) applicable also when artificial intelligence is used by employers, specifies that, when AI tools are used to draft or create the selection guidelines for hiring, transferring, promoting or

dismissing employees (even in larger companies with more that 500 employees), the works council still has the right to approve them, and that if there is no agreement between employer and works council, even if the guidelines were prepared with AI assistance, the matter still goes to the conciliation committee for a binding decision (Works Constitution Act, 2024).

Finally, in 2023, Germany signed the EU AI Act, thus adding to its policy approach to AI a new legally regulated framework in line with EU standards. To facilitate the implementation of the Act, the German Federal Republic has not only established the Federal Network Agency (Bundesnetzagentur) to supervise the implementation of the Act as the Market Surveillance Authority (Overview, 2025), but it has also launched the 2023 *AI Action Plan*, allocating more funds into AI research and development than the AI Act requires to foster public awareness and industry collaboration within a technological "eco-system that is both cutting-edge and accountable", and which upholds the values of ethics and transparency against harmful manipulation of AI (Krüger, 2025). Germany's Independent Federal Anti-Discrimination Commissioner and the Federal Commissioner for Data Protection and Freedom of Information have strongly supported the need to balance effective AI regulation with the preservation of innovation with regard to the 2023 *AI Action Plan*, reiterating the need for the new legal framework to be aligned with the GDPR, "human rights and fundamental rights … cybersecurity, product safety, transparency" and data quality (Lorenz, 2024). Germany's implementation of the AI Act and its focus on innovation must henceforth be reconciled also with its Basic Law (*Grundgesetz*) which, in Article 5[43] guarantees the right to freedom of expression to all its citizens (Basic Law,

[43] Article 5 of the German Basic Law states:

> "(1) Every person shall have the right freely to express and disseminate his opinions in speech, writing and pictures and to inform himself without hindrance from generally accessible sources. Freedom of the press and freedom of reporting by means of broadcasts and films shall be guaran teed. There shall be no censorship. - (2) These rights shall find their limits in the provisions of general laws, in provisions for the protection of young persons and in the right to personal honour. (3) Arts and sciences, research and teaching shall be free.

Article 5). This right, which must not be abused "to combat the free democratic basic order" (Art. 18), can be restricted, as per Article 19, "by or pursuant to a law" without affecting its "essence" (Basic Law, Articles 18 & 19). Hence, when AI is used for online counterterrorism, the right to freedom of speech may be restricted, but such restrictions must be proportionate and not lead to over censorship. In the field of counterterrorism and security, the Federal Criminal Police Office (BKA), the Rhineland-Palatine State Criminal Police Office (LKA), and the German Research Center for Artificial Intelligence have all signed a contract to employ AI for intelligence and police purposes, focusing on "the pre-selection and relevance assessment of immense amounts of data and the analysis of unstructured raw data" for enhancing predictive analysis, online surveillance, and online counter terrorism, by promoting narratives which aim to prevent radicalization by discrediting terrorist and extremist propaganda (DFKI AI, 2021). In light of the new AI Act, however, Germany will need to recalibrate its use of AI in counterterrorism, aligning existing practices with stricter transparency and fundamental rights safeguards.

3.3.3 The EU AI Act in France: Balancing Security Practices with Regulations

In France, the EU AI Act will serve as the primary legal framework addressing AI development and usage. However, it seems possible that the country will enact some tailored regulations in specific fields in conjunction with the enactment of the AI Act: an example is the legislative proposal made to revise the country's Intellectual Property Code (IPC) to take in consideration AI (Hainsdorf & Liard, 2024). The country has indeed made already extensive use of AI: as of 2025, France is "the third country in the world in terms of the number of AI researchers" and it "has more than 1000 startups, including gems like Mistral AI, H Company, as

The freedom of teaching shall not release any person from allegiance to the constitution" (Basic Law, Article 5).

well as applications such as Alan, Pigment, Doctolib, and more", thus becoming "the leading hub for generative AI in Europe" (Elysee, 2025). It thus appears logic that its representatives have been highly vociferous in the drafting of the AI Act, signaling that France wants to maintain its leading role in AI development and governance (Elliot & Silverman, 2024, p. 4). However, only rare mentions of AI had been made within French national law before 2023: an example can be found in Law No. 2016-1321 of 7 October 2016, enacted to promote innovation and support the growth of the digital economy, which gave the French Data Protection Authority (CNIL) the responsibility of examining the ethical and societal implications of emerging technologies, and AI systems (Elliott & Silverman, 2024, p. 4).

The CNIL, established in 1978 as France's independent authority on data protection, has even released an *AI Action Plan*, in May 2023, a guide for overseeing AI systems which revolves around 4 main goals, namely:

> "(1) understanding the functioning of AI systems and their impacts for people,
>
> (2) enabling and guiding the development of AI that respects personal data,
>
> (3) federating and supporting innovative players in the AI ecosystem in France and Europe,
>
> (and finally) (4) audit(ing) and control(ling) AI systems and protect(ing) people" (CNIL, 2023).

Therefore, while ensuring that fundamental rights are protected, the CNIL ensures transparent AI, monitors compliance with regulations on enhanced video surveillance, and investigates complaints concerning AI systems. It has also "launched … (AI) innovative projects" in the fields of health in 2021 and education in 2022 (CNIL, 2023). In 2023, then, the CNIL has launched a new support program "to assist innovative companies in their compliance with the GDPR", and has chosen five of its representatives to create its own AI Department with the goals of

"(1) improving the understanding of AI systems,

(2) strengthening the CNIL's expertise for the identification and prevention of privacy risks related to the implementation of AI systems,

(3) preparing for the implementation of the future EU AI Act, and

(4) developing relationships with other stakeholders in the AI ecosystem" (CNIL, 2023; Elliott & Silverman, 2024, p. 4).

In light of the massive work the CNIL is carrying out in the field of AI, the Agency is expected to continue to be France's primary oversight body for the implementation of the EU AI Act (Elliott & Silverman, 2024, p.4). However, France has not yet designated neither a National Public Authority or a Notifying Authority or a Market Surveillance Authority under the AI Act, even if at the EU AI Board Meeting of November 2024, the country was represented by its Directorate General of Enterprises (Overview, 2025).

The CNIL has also been one of the most important bodies supervising France's use of artificial intelligence for security purposes, especially in the field of counterterrorism. In fact, while France still lacks clear laws regulating AI, the country has not shied away from using mass surveillance strategies and AI systems for security purposes following the declaration of the State of Emergency after the 2015 terrorist attacks. The Loi No. 2015-912 (Relative au Renseignement) of 24 July 2015 or *Intelligence Act* expanded the surveillance capabilities of the country, authorizing techniques such as real-time metadata collection, geolocation tracking and the use of black boxes to examine potential terrorist threats through automated data analysis (Loi n° 2015-912, 2015). Of course such law raised serious concerns with respect to the rights to "privacy, correspondence and the inviolability of the home" and the CNIL has often repeated "that there is a need for "more precise implementation conditions that limit violations of these fundamental

rights, on the one hand, and for effective monitoring methods that are adapted to the nature of these violations, on the other", henceforth calling for better protection during data collection (Rees, 2015). France's 2018 Villani Report *For a Meaningful Artificial Intelligence* followed the footprint left by its 2015 Intelligence Act, stating that AI systems will be pivotal "to ensure security missions, maintain the ascendancy against (France's) potential adversaries (and) maintain (the country's) position in relation to allies" (Thibout, 2018). More recently, in light of the 2024 Paris Olympic Games, France enacted Loi No. 2023-380 and allowed AI-driven video surveillance systems around Paris, by deploying approximately 300 AI cameras designed to detect anomalies such as crowd surges, abandoned objects, and fires by scanning both athletes and spectators (Dal Bello et al., 2024). Here, again, there have been many complaints by the CNIL but also by some NGOs, such as La Quadrature du Net which legally challenged the act (Carpentier, 2024).

In light of France's extensive use of AI for surveillance and security purposes, the introduction of the AI Act may generate some tension between the country's security goals and the European Union's objectives and principles. Tension may arise also in the protection of fundamental rights, and especially the right to freedom of speech, when AI systems are used for online counterterrorism purposes. While protected under Art. 11[44] of the 1789 Declaration of the Rights of Man and of the Citizens which holds constitutional values since 1971, and also under Art. 4(3)[45] of the 1958 Constitution of the Fifth Republic, in cases of emergency, the balance between fundamental rights protection and security has tilted dramatically towards the latter in

[44] Article 11 of the Declaration of the Rights of Man and of the Citizen recites:

> "The free communication of ideas and of opinions is one of the most precious rights of man. Any citizen may therefore speak, write and publish freely, except what is tantamount to the abuse of this liberty in the cases determined by Law" (Declaration of the Right of Man and of the Citizen, Article 11).

[45] Article 4(3) of the French Constitution recites:

> "Statutes guarantee the pluralistic expression of opinions and the equitable participation of political parties and groups in the democratic life of the Nation" (The Constitution of the Fifth Republic, Article 4).

recent years when dealing with matters of counterterrorism: after the 2015 terrorist attacks, France has indeed proclaimed a State of Emergency[46] grounded in Loi n° 55-358 which was then "extended five times, until November 2017, when many of the emergency powers and measures were codified and written into ordinary law (and) restrictions introduced during the temporary state of emergency became permanent" (Kilpatrick, 2020, p. 4). During the State of Emergency the Minister of Interior could:

> "• Place under house arrest persons whose actions prove dangerous for security and public order; …
>
> • Take any measure necessary to block online materials condoning or inciting acts of terrorism;
>
> • Disband by decree (via the Council of Ministers) any association or group involved in, facilitating, or inciting the commission of acts which pose a serious threat to public order;
>
> • Restrict freedom of movement, of persons and vehicles, within the state by imposing defence and security zones, and imposing a curfew in certain areas; …
>
> • Prohibit certain public meetings and demonstrations;
>
> • Provisionally close certain meeting places;
>
> • Authorise administrative searches" (Kilpatrick, 2020, pp. 9-10).

[46] France's constitutional and legal framework provides for three distinct levels of emergency powers, depending on the gravity of the situation at hand. The state of siege, never enacted, is grounded in Article 36 of the Constitution and the Defense Code and it is reserved for situations of foreign war or armed insurrection and must be decreed by the Council of Ministers, because it transfers certain powers from civilian to military authorities, with any extension beyond 12 days requiring parliamentary approval (The Constitution of the Fifth Republic, 1958). The state of emergency, frequently enacted, is regulated by the 1955 Act and it is triggered by serious threats to public order or disasters. It allows the government to impose curfews, and other restrictions on movement or assembly (Loi n° 55-358). Finally, there are the "exceptional powers" under Article 16 of the Constitution, which give the President of the Republic sweeping authority when the Republic's institutions, territorial integrity, or national independence face an immediate and serious threat, and when normal constitutional functioning is disrupted (The Constitution of the Fifth Republic, 1958).

However, while these measures should have tackled only terrorist threats, they quickly expanded and started more broadly to be used to "pacify dissent" and restrict civil liberties, thus impinging on the right to privacy and the freedoms of movement, assembly, and speech especially of Muslim minorities: with regard to freedom of speech, for example, "ninety per cent of reports to the Collectif Contre L'Islamophobie en France (Collective Against Islamophobia, CCIF) concerning discrimination in primary schools in 2015 related to mothers barred from accompanying school outings because of the supposedly "proselytizing character" of their clothing" (Kilpatrick, 2020, pp. 11-13). This normalization of the securitization of the legal order, furthered by the inclusion of emergency restrictions into ordinary law, such as allowing "places of worship to be closed based on "serious reasons", rather than on material proof", have negatively impacted on fundamental rights, enhancing stigmatization and marginalization of certain groups (Kilpatrick, 2020, pp. 16-17). Such normalization of security measures risks become even more embedded within the legal system when applied to the use of AI in counterterrorism without stringent regulations. Therefore, important adjustments will have to be made during the implementation of the AI Act to ensure a careful calibration between security and freedom of speech whenever AI is used in the field of counterterrorism.

3.3.4 Spain's Adaptation of the EU AI Act Between Digital Rights and Security Risks

Like many other European countries before the AI Act, Spain did not possess a legal framework addressing the use and development of artificial intelligence. However, in November 2020, the country published its *National Artificial Intelligence Estrategy 2020-2025*, a document which, alongside the country's Charter of Digital Rights enacted in July 2021, constitutes the so-called "Spanish Guidelines" for the enactment of the European AI Act (Calvo et al., 2024). The

country's *National Strategy*, supporting regulatory planning and development of AI, holds seven goals, namely:

> "(1) scientific excellence and innovation in artificial intelligence;
>
> (2) projection of the Spanish language;
>
> (3) creation of qualified employment;
>
> (4) transformation of the productive system;
>
> (5) trust environment in relation to artificial intelligence;
>
> (6) humanistic values in artificial intelligence; (and, finally)
>
> (7) inclusive and sustainable artificial intelligence" (Gobierno de España, 2020, p.16).

All these goals clearly align with the EU AI Act's goals of creating a trustworthy and human-centric AI, which Spain plans to achieve through its *National AI Strategy* by enacting six different pillars of AI development: the first pillar aims to "promote scientific research, technological development and innovation in AI"; the second to "promote digital capabilities, empower national capabilities and attract global skills in the field of AI"; the third to "develop data platforms and technological infrastructures in support of AI"; the fourth to "incorporate AI into value chains to transform the economic fabric"; the fifth to "enhance the use of AI in government administration and in national strategic missions"; and finally, the sixth to "establish an ethical and regulatory framework that reinforces the protection of individual and collective rights, in order to guarantee inclusion and social welfare" (Gobierno de España, 2020, pp. 18-19). Of course, the sixth pillar will be developed through the adoption and enactment of the AI Act, while following the framework established by the Charter of Digital Rights, a non-binding document which adapts Spain's fundamental rights to the digital age. Some of the most important rights recognized by the Charter are: the right to freedom of expression and information (Art. XIV), the right to privacy

and data protection (Art. III), the right to non-discrimination (Art. VIII) and that of digital inclusion (Art. XI), the right to cybersecurity (Art. VI), and also the rights regarding artificial intelligence (Art. XXV) (Gobierno de España, 2021, pp. 8-27). The latter rights are particularly relevant when considering how Spain is adjusting its policy and legal landscapes to welcome the AI Act in its territory: Art. XXV, indeed, outlines key principles for ensuring that AI systems respect fundamental rights and values, such as the right to non-discrimination, and that they are human-centric, transparent and accountable; crucially, the right also guarantees that users have the ability to challenge decisions made by AI systems through human intervention, especially when such decisions affect their rights or assets (Gobierno de España, 2021, p. 27).

Spain has also established the Spanish Artificial Intelligence Supervisory Agency (AESIA) in September 2023 "as an autonomous agency of the Spanish Department of Digital Transformation" which will be responsible for overseeing the implementation and the enactment of the AI Act acting as both the Market Surveillance Authority and the National Public Authority (Overview, 2025). Furthermore, to ensure the smooth application of the Act, Spain has passed the Royal Decree 817/2023 "which establishes a controlled test environment in compliance with the EU AI Act ("RD Sandbox") … designed to enable participants to implement high-risk AI systems under the EU AI Act with the aim of obtaining guidance on achieving compliance (Calvo et al., 2024). Another important step taken by Spain was the creation of the Artificial Intelligence Advisory Council (Consejo Asesor de Inteligencia Artificial) with the National Order ETD/670/2020, an independent body responsible of advising the Ministry of Digital Transformation on AI policy and of fostering experts' discussions on the future of AI in public policymaking (Calvo et al., 2024). Lastly, and more recently, Spain's Council of Ministers has enacted the *Artificial Intelligence Strategy 2024*, in order to promote ethical and transparent AI

development, develop sustainable infrastructure, strengthen supercomputing capabilities (with an already active €90 million investment in the MareNostrum supercomputer) as well as AI talent through specialized training and scholarships, but also to establish a clear legal framework starting from the AI Act which will be comprehensive of laws on cybersecurity as well (Council of Ministers, 2024).

While the country seems to be advancing in the fields of AI development and regulation, however, some security issues have already emerged in the sector of regional security operations in politically sensitive areas like Catalonia: the Spanish government seems to have used Pegasus spyware against 60 Catalan politicians, lawyers and activist to enact a deep surveillance of the latter and suppress the Catalonian independence movement (Farrow, 2022). The main problems therefore remain that of predictive policing and deep surveillance which risk impinging on the civil liberties, fundamental rights, and privacy of the affected individuals, and risk generating discriminatory practices with the scope of reducing and eliminating political dissent. The right to freedom of speech is thus particularly at risk when Spain uses AI for national security or counterterrorism strategies, even if this is protected under Art. 20[47] of the Spanish Constitution,

[47] Article 20 of the Spanish Constitution reads:

"1. The following rights are recognised and protected:

a) the right to freely express and disseminate thoughts, ideas and opinions trough words, in writing or by any other means of communication;

b) the right to literary, artistic, scientific and technical production and creation;

c) the right to academic freedom;

d) the right to freely communicate or receive accurate information by any means of dissemination whatsoever.

The law shall regulate the right to invoke personal conscience and professional secrecy in the exercise of these freedoms.

2. The exercise of these rights may not be restricted by any form of prior censorship.

3. The law shall regulate the organisation and parliamentary control of the social communications media under the control of the State or any public agency and shall guarantee access to such media to the main social and political groups, respecting the pluralism of society and of the various languages of Spain.

4. These freedoms are limited by respect for the rights recognised in this Title, by the legal provisions implementing it, and especially by the right to honour, to privacy, to personal reputation and to the protection of youth and childhood.

5. The confiscation of publications and recordings and other information media may only be carried out by means of a court order" (Spanish Constitution, Article 20).

because Article 55(1)[48] recognizes the possibility of restricting such right, among others, for individuals or groups during a state of emergency or in other at-risk situations in connection with investigations of terrorist activity (Spanish Constitution, Articles 20 & 55). Article 55 has been consistently applied in the past to fight against Basque terrorism, and particularly against organizations such as Euskadi ta Askatasuna (ETA) (De La Cuesta, 2007, p. 3). More recently, however, it has also been employed in the fight against cyberterrorism with the implementation of Organic Law 13/2015, aiming at strengthening the fight against jihadist online terrorism (Garrigues, 2015).

All in all, as Spain implements the AI Act, it will need to ensure that freedom of speech and other fundamental rights are not only formally protected but also effectively respected, especially in cases of national and regional security and counterterrorism.

3.4 The AI Act's Impact on the Balance Between Security and Freedom of Speech in Counterterrorism

The EU AI Act aims to balance security imperatives with the protection of fundamental rights across all domains where AI systems are deployed. This balanced regulatory approach to AI should therefore extend to politically sensitive areas such as national security and counterterrorism

---

[48] Article 55 in its entirety of the Spanish Constitution reads:

> "1. The rights recognised in Articles 17 and 18, clauses 2 and 3, Articles 19 and 20, clause 1, subclauses, a) and d) and clause 5, Articles 21 and 28, clause 2, and Article 37, clause 2, may be suspended when the state of emergency or siege (martial law) is declared under the terms provided in the Constitution. Clause 3 of Article 17 is excepted from the foregoing provisions in the event of the proclamation of a state of emergency. 2. An organic law may determine the manner and the circumstances in which, on an individual basis and with the necessary participation of the Courts and proper Parliamentary control, the rights recognised in Articles 17. clause 2, and 18, clauses 2 and 3, may be– suspended as regards specific persons in connection with investigations of the activities of armed bands or terrorist groups. Unjustified or abusive use of the powers recognised in the foregoing organic law shall give rise to criminal liability where it is a violation of the rights and liberties recognised by the law" (Spanish Constitution, Article 55).

online, where some fundamental rights, such as the right to freedom of speech are particularly at risk of being infringed. However, most European countries have already adopted a security-prioritized stance in the fields of counterterrorism and in monitoring online radicalization, using also many AI systems without clear regulations establishing correct and incorrect practices.

The AI Act, born as a compromise between 27 diverse countries, many of which lean toward the protection of security over fundamental rights in the current geopolitical climate, reflects this reality. While protecting fundamental rights and requiring that AI development is safe, transparent and human-centered, the Act allows for broad national security exemptions for AI systems developed or used exclusively for military or national security purposes: Art. 2(3) underlines indeed that the Act "... shall not, in any event, affect the competences of the Member States concerning national security (Regulation (EU) 2024/1689, 2024, p. 45). For counterterrorism applications, the Act is relevant because it addresses AI use in biometric surveillance, predictive policing, and content moderation, all areas in which AI systems have been previously used without clear legal boundaries. However, and interestingly enough, the Act never mentions words such as counterterrorism or radicalization, possibly including the latter under the larger, and unclearly defined umbrella-term, 'national security'. Art. 5(1)(d) bans "the placing on the market" or "the use of an AI system for making risk assessments of natural persons in order to assess or predict the risk of a natural person committing a criminal offence" while Art. 5(1)(h) forbids real-time biometric identification in publicly accessible spaces; Art. 5(1)(c), then, prohibits AI systems used for social scoring

> "leading to (i) detrimental or unfavorable treatment of certain natural persons or groups of persons in social contexts that are unrelated to the contexts in which the data was originally generated or collected; (and) (ii) detrimental or unfavorable treatment of certain natural

> persons or groups of persons that is unjustified or disproportionate to their social behavior or its gravity" (Regulation (EU) 2024/1689, 2024, pp. 51-52).

However, such articles make clear exceptions for law enforcement purposes, for dealing with people who already possess a criminal record, for a targeted search of a specific victim, for the prevention of an imminent threat, or for the identification of suspects in serious crimes (Regulation (EU) 2024/1689, 2024, pp. 51-52). All these exceptions, even if strictly regulated, still carry the risk of infringing on fundamental rights such as the right to freedom of speech: for instance, when an AI system takes down some online content it identifies as incitement to terrorism, it may inadvertently remove material that solely discusses or reports on terrorism without promoting it due to the system's limited capability to accurately distinguish between legitimate and illegitimate content.

Another problem which emerges from the Act's national security exemption is the lack of oversight: Recital 24 of the Act, indeed, states that

> "If, and insofar as, AI systems are placed on the market, put into service, or used with or without modification of such systems for military, defence or national security purposes, those should be excluded from the scope of this Regulation regardless of which type of entity is carrying out those activities, such as whether it is a public or private entity" (Regulation (EU) 2024/1689, 2024, p. 7).

The Act, therefore, does not strictly regulate AI systems used for security purposes, leaving a great gap in safeguards and oversight in one of the most sensitive areas of AI application. This exemption could in fact allow that more intrusive AI systems used for military purposes, without being noticed, may be deployed without the safeguards required by the Act, thus potentially infringing on civil liberties and democratic processes (European Center, 2022).

The Act also deals with AI's impact on freedom of expression, retaining it one of the fundamental rights which might be impinged upon by AI and especially high-risk AI systems (Regulation (EU) 2024/1689, 2024, p. 13). Art. 5(1)(a) prohibits, indeed,

> "(a) the placing on the market, the putting into service or the use of an AI system that deploys subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective, or the effect of materially distorting the behaviour of a person or a group of persons by appreciably impairing their ability to make an informed decision, thereby causing them to take a decision that they would not have otherwise taken in a manner that causes or is reasonably likely to cause that person, another person or group of persons significant harm";

For this reason, the transparency obligations must obligatorily be met by AI providers without impinging upon their users' freedom of expression. Art. 50 obliges AI providers to be transparent with their users when they provide AI-generated content, including "synthetic audio, images, video, or text" (Art. 5(2)) or when they generate deep fakes (Art. 5(4)); but, the Act also requires that "compliance with this transparency obligation should not be interpreted as indicating that the use of the AI system or its output impedes the right to freedom of expression". Even in the protection of fundamental rights such as the right to freedom of expression, however, the Act allows exceptions for national security reasons, or in cases where AI systems are used by law enforcement for detecting, preventing, investigating, or prosecuting criminal offenses, provided such use is authorized by law (Cabrera, 2024).

All in all, while the EU AI Act represents a significant step toward responsible AI governance which balances security imperatives with the protection of fundamental rights, its exemptions for national security reasons may create regulatory grey areas which distort the

security-fundamental rights balance in favor of security. More specifically to the field of counterterrorism, then, the Act remains too unclear and broad, potentially blurring the line between online counterterrorism and unregulated censorship. Indeed, without careful oversight, AI tools may silence marginalized voices or suppress dissent, which are both forms of expression allowed in a democracy but also very difficult for a machine to distinguish from actual incitement to terrorism. Furthermore, since the AI Act will possibly set a global precedent for AI regulation, its exemptions may allow other countries to exploit similar regulatory grey areas. This could then lead to a fragmented international legal framework on AI which is always more dangerous than a cohesive one, because it may impinge upon both security and human rights negatively, allowing states and tech companies to 'cherry-pick' the regulations they prefer. While this scenario is likely to unfold outside the EU borders, within the European Single Market both states and companies will have to adapt their AI development and use to the more stringent legislative framework established by the EU AI Act.

3.4.1 What can Public Authorities do to Better Ensure a Fair Balance Between Security and Freedom of Speech When AI is Employed in Counterterrorism?

While Big Tech companies and AI providers must strictly follow the rules set out in the Act, public authorities must remain vigilant to ensure that a good balance is stricken between security and the protection of fundamental rights when employing AI systems for national security and counterterrorism purposes. Public authorities must hence always "verify the legality and proportionality of the interference, (and) … evaluate the adverse effect on the protected rights" since, according to the CJEU, "a fundamental objective of general interest, such as that of crime

prevention, investigation, detection and prosecution, is not sufficient in itself to justify an interference with other recognised rights" (Casaburo & Marsh, 2024, p. 1570).

The three national authorities which every EU member must elect to oversee the rightful implementation of the Act, namely the National Public Authority, the Notifying Authority and the Market Surveillance Authority, have a pivotal role in ensuring a good balance between security imperatives and fundamental rights protection. They should be at the forefront in ensuring independent oversight of law enforcement and military AI systems and tools; they should conduct frequent risk assessments on AI tools, before, during and after their implementation; and they should also guarantee accessible legal remedies for all the persons whose rights have been infringed from an AI system. Importantly, "risk assessment on the use of AI in law enforcement need to be carried out regularly and approaches adjusted accordingly, in order to center law enforcement policies and practices on human rights and fundamental freedoms" (OSCE, 2024, p. 4). This is important because, as aforementioned, AI systems embed numerous biases which increase the opaqueness of their operations. Checking for such biases and regularly controlling the deployment of AI tools is thus important to ensure that users' rights are upheld. The oversight of public authorities is crucial, because AI systems lack contextual understanding and may thus, in the field of online content moderation, remove dissenting content which is lawful but that such systems read as inciting to violence or terrorism: the balancing of conflicting interests and rights, hence, "cannot, in a rule-of-law state, be delegated to any form of automatism" (Pollicino & Dunn, 2024, p. 7 as cited in Antonucci, 2024, p. 1). "Once the non-neutrality of AI is acknowledged, it becomes necessary to steer it toward the values of equality": henceforth public authorities and legislators must have "the possibility of allowing the processing of sensitive data (usually

prohibited), in order to eliminate discriminatory effects and thereby respond to a collective need" (Pollicini & Dunn, 2024 as cited in Antonucci, 2024, p. 5).

It is also very important that public author supervise the transparency and ethics of every AI system, especially of those employed in politically sensitive areas such as counterterrorism, so as to allow for a human-centered and rightful application of the AI Act, which creates a good balance between fundamental rights protection and security imperatives. By publishing transparency reports and openly declaring to users when AI systems are being used to moderate online content, public authorities should be capable of enhancing both transparency and accountability of AI systems. Furthermore, they should strive to increase their cooperation with the private sector in this field by "shar(ing) information, resources, and expertise to develop effective solutions and strategies for addressing the social impacts of AI" (Esmailzadeh & Motaghi, 2024, p. 178). The Organization for Security and Co-operation in Europe (OSCE) has indeed stressed that "sustainable strategies (to ensure the balance between security and fundamental rights protection when AI systems are employed the field of online counterterrorism) can only be found when online platforms and service providers are engaged alongside (public authorities and) other stakeholders such as civil society, media, international actors and academia" to ensure transparency, accountability and fundamental rights as the standards for the development and use of AI systems (OSCE, 2024, p. 8).

4. Defining the Future of AI in Counterterrorism: A Comparison of the EU and the US's AI Governance Models

The extensive adoption of AI across almost every aspect of human life has prompted a global competition among states to shape AI governance and regulation. On one hand, some states, follow the European Union, advocating stronger regulation of AI systems. With the implementation of the EU AI Act, in fact, the European Union is striving to provide a comprehensive legal framework which promotes a safe, transparent and human-centric AI. With this Act, the EU has positioned itself as a leader in global AI regulation, offering an easy-to-adapt model for any state to reconcile security imperatives with fundamental rights protection, especially in the sensitive field of online counterterrorism. On the other hand, some states prefer to follow the more libertarian and market-driven approach to AI governance of the United States which, from across the Atlantic, is striving to become a global leader in AI as well. The US considers AI not only as a very useful tool for remaining a global leader in the sector of technological innovation (especially in light of its strategic competition against China) but also as a weapon for the protection of its national security. Thus, one of the first acts of the new Trump administration was the enactment of "an *Executive Order for Removing Barriers to American Leadership in AI* ("Removing Barriers EO[49]") which effectively abrogated "President Biden's *Executive Order for*

[49] Section 1, 2 and 4 of Trump's Removing Barriers EO read:

> "By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:
>
> Section 1. Purpose. The United States has long been at the forefront of artificial intelligence (AI) innovation, driven by the strength of our free markets, world-class research institutions, and entrepreneurial spirit. To maintain this leadership, we must develop AI systems that are free from ideological bias or engineered social agendas. With the right Government policies, we can solidify our position as the global leader in AI and secure a brighter future for all Americans.

the Safe, Secure and Trustworthy Development and Use of AI ("Biden EO[50]")" in order to "enhance America's global AI dominance", by either revising or discarding any previous policies perceived as obstacles to that goal (Anderson & Comstock, 2025).

This order revokes certain existing AI policies and directives that act as barriers to American AI innovation, clearing a path for the United States to act decisively to retain global leadership in artificial intelligence.

Sec. 2. Policy. It is the policy of the United States to sustain and enhance America's global AI dominance in order to promote human flourishing, economic competitiveness, and national security. …

Sec. 4. Developing an Artificial Intelligence Action Plan. (a) Within 180 days of this order, the Assistant to the President for Science and Technology (APST), the Special Advisor for AI and Crypto, and the Assistant to the President for National Security Affairs (APNSA), in coordination with the Assistant to the President for Economic Policy, the Assistant to the President for Domestic Policy, the Director of the Office of Management and Budget (OMB Director), and the heads of such executive departments and agencies (agencies) as the APST and APNSA deem relevant, shall develop and submit to the President an action plan to achieve the policy set forth in section 2 of this order. …" (Exec. Order No. 14179, 2025).

[50] Section 1 and 2 of the Biden EO read:

"Sec. 1. *Purpose.* Artificial intelligence (AI) holds extraordinary potential for both promise and peril. Responsible AI use has the potential to help solve urgent challenges while making our world more prosperous, productive, innovative, and secure. At the same time, irresponsible use could exacerbate societal harms such as fraud, discrimination, bias, and disinformation; displace and disempower workers; stifle competition; and pose risks to national security. Harnessing AI for good and realizing its myriad benefits requires mitigating its substantial risks. This endeavor demands a society-wide effort that includes government, the private sector, academia, and civil society. …

Sec. 2. *Policy and Principles.* It is the policy of my Administration to advance and govern the development and use of AI in accordance with eight guiding principles and priorities. When undertaking the actions set forth in this order, executive departments and agencies (agencies) shall, as appropriate and consistent with applicable law, adhere to these principles, while, as feasible, taking into account the views of other agencies, industry, members of academia, civil society, labor unions, international allies and partners, and other relevant organizations:

(a) Artificial Intelligence must be safe and secure. …

(b) Promoting responsible innovation, competition, and collaboration will allow the United States to lead in AI and unlock the technology's potential to solve some of society's most difficult challenges. …

(c) The responsible development and use of AI require a commitment to supporting American workers. …

(d) Artificial Intelligence policies must be consistent with my Administration's dedication to advancing equity and civil rights. …

(e) The interests of Americans who increasingly use, interact with, or purchase AI and AI-enabled products in their daily lives must be protected. ….

(f) Americans' privacy and civil liberties must be protected as AI continues advancing. …

(g) 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. …

(h) The Federal Government should lead the way to global societal, economic, and technological progress, as the United States has in previous eras of disruptive innovation and change. …" (Exec. Order No. 14110, 2023).

The last chapter of this thesis analyzes and compares these two very different approaches to AI governance in the field of counterterrorism, seeking to understand future trends in digital security and fundamental rights protection. It thus provides an overview of the US approach to the protection of the right to freedom of speech and online content regulation as well as to AI and its use in online counterterrorism. Consequently, it assesses the impact of the US approach on the balance between security and freedom of speech in online counterterrorism and identifies its similarities and differences with the European approach analyzed in Chapter 3. Finally, the chapter concludes with a discussion on the advantages and limitations of each governance model, trying to understand whether the EU's rights-based framework might emerge as a benchmark for international regulation in this evolving field or whether the US' more libertarian approach will succeed.

4.1 The US Approach to the Protection of the Right to Freedom of Speech and Content Regulation Online

"...The land of the free and the home of the brave", the United States has always tried to embody the ideals of freedom and bravery in its governance, legal frameworks, and societal values, considering the right to freedom of expression as a foundational element of its democratic identity (Key, 1814). This has been possible thanks to the US First Amendment[51] which guarantees the core rights of the American democracy, namely the rights to freedom of religion, speech, of the press, to peacefully assemble and to petition the government (US Const. Amend. I). These are the

[51] The First Amendment reads: "Congress shall make no law respecting an establishment of religion or prohibiting the free exercise thereof; or abridging the freedom of speech, or of the press; or the right of the people peaceably to assemble, and to petition the Government for a redress of grievances" (US Const. Amend. I).

rights which rendered the US the ‘land of milk and honey’: a biblical reference from the *Book of Exodus* often used in popular culture, such as in the Italian Crialese’s 2006 movie, *Nuovomondo* (“Golden Door”), where the United States is described as a land of boundless opportunities where people swim in milk and where every citizen can be whoever and say whatever they want (Crialese, 2006). While the US does not have a national religion, the right to freedom of speech has acquired an almost sacrosanct status in American legal and cultural discourse and in the minds of all the American citizens, regardless of their ethnicity, religion, political association or group of belonging. It appears thus evident that, when it comes to freedom of speech, the US has always taken a lighter regulatory approach when compared to the European Union, protecting even controversial or offensive expressions with only very limited exceptions.

This is the case even in the protection of digital speech with the government striving to respect the First Amendment by reducing at a minimum online government interference and leaning toward platform self-regulation. This approach has been sealed by Section 230(c) of the

1996 Communications Decency Act (CDA)[52] which provides "a broad immunity[53] from suit for (any) provider or user of an interactive computer service", specifying, in section 230(c)(1), "that service providers may not be treated as the publisher or speaker of any information provided by another information content provider" and, in section 230(c)(2), that "service providers may not be held liable for voluntarily acting to restrict access to objectionable material" (Brannon & Holmes, 2024). Most Courts have thus allowed the "early dismissal of many legal claims against interactive computer service providers, preempting lawsuits and statutes that would impose liability on third-party content", basing their decisions on the legal precedent set by the Fourth Circuit's 1997 decision in *Zeran v. America Online, Inc.*[54] which solidified broad Section 230

[52] The CDA, comprising findings (230(a)) and policy statements (230(b)), is part of the 1996 Telecommunications Act and was "intended to modernize the existing protections against obscene, lewd, indecent or harassing uses of a telephone" (Brannon & Holmes, 2024). The relevant sections for this thesis are © and (d), which read:
"(c)Protection for "Good Samaritan" blocking and screening of offensive material
(1)Treatment of publisher or speaker
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.
(2) Civil liability
No provider or user of an interactive computer service shall be held liable on account of—
(A)any action voluntarily taken in good faith to restrict access to or availability of material that the provider or user considers to be obscene, lewd, lascivious, filthy, excessively violent, harassing, or otherwise objectionable, whether or not such material is constitutionally protected; or
(B)any action taken to enable or make available to information content providers or others the technical means to restrict access to material described in paragraph (1).
(d)Obligations of interactive computer service
A provider of interactive computer service shall, at the time of entering an agreement with a customer for the provision of interactive computer service and in a manner deemed appropriate by the provider, notify such customer that parental control protections (such as computer hardware, software, or filtering services) are commercially available that may assist the customer in limiting access to material that is harmful to minors. Such notice shall identify, or provide the customer with access to information identifying, current providers of such protections" (47 US Code §230).

[53] This immunity, however, does not apply to "suits brought under federal criminal law, intellectual property law, any state law 'consistent' with Section 230, certain electronic communications privacy laws, or certain federal and state laws relating to sex trafficking" (Brannon & Holmes, 2024).

[54] This case stemmed from the defamation of Ken Zeran by an anonymous user on AOL who posted advertisements falsely claiming that Zeran was selling offensive t-shirts related to the Oklahoma City bombing (Zeran v. America Online, 1997). These posts included Zeran's phone number, leading to a flood of harassment and death threats, which made Zeran sue AOL for defamation, notwithstanding the fact that the platform had eventually removed all these posts against Zeran (Zeran v. America Online, 1997). The legal question was whether AOL could be held liable for defamatory content posted by a third party, and whether Section 230 of the Communications Decency Act (47 U.S.C. § 230) shielded them from such liability (Zeran v. America Online, 1997). Both the District Court and the Fourth Circuit Court of Appeals ruled in AOL's favor, holding that Section 230 grants immunity to internet service providers from being treated as the publisher or speaker of third-party content (Zeran v. America Online, 1997).

protections for online platforms. Notwithstanding some exceptions which have been developed to these Section 230 protections, American Courts have largely upheld the *Zeran* precedent, even in cases involving online terrorist content. Notably, in *Gonzalez v. Google LLC*[55] (2023), the US Supreme Court reaffirmed that recommender systems are covered by Section 230 protections, thereby shielding platforms from liability for terrorism-related content posted by users (Brannon & Holmes, 2024).

This legal framework empowers private big tech companies to act as *de facto* regulators of content on their platforms, allowing them broad discretion over what constitutes permissible or harmful speech online. In giving such a great decisional power to private tech companies which are not bound by the same constitutional constraints as the US government, the US approach to platform self-regulation has raised many concerns about transparency and consistency in content moderation. In light of these concerns, the US Department of Justice has enacted a review of Section 230 of the CDA "to encourage a safer and more open internet" by outlining four key areas for reformation (Department of Justice, 2023). The first focuses on "incentivizing online platforms to address illicit content" by implementing "bad Samaritan carve-outs" which would remove the immunity from platforms that intentionally facilitate illegal activity, and other carve-outs for content on "child abuse, terrorism and cyber-stalking"; the second area of reform requires a clarification of "federal government enforcement capabilities to address unlawful content"; the

---

[55] This case stemmed from the death of Nohemi Gonzalez, a 23-year-old US citizen killed during ISIS' 2015 terrorist attacks in Paris. Her parents filed a lawsuit against Google alleging that the platform was "directly and secondarily responsible for the terrorist attack" that killed Gonzalez: essentially, "the plaintiffs claimed that Google aided and abetted ISIS and conspired with the terrorist organization" (Gonzalez v. Google, 2023). After the Ninth Circuit Court, "the Supreme Court affirmed the Ninth Circuit decision by stating that petitioners failed to plausibly allege that "Google reached an agreement with ISIS", as required for conspiracy liability, and that Google's acts were "intended to intimidate or coerce a civilian population, or to influence or affect a government", as required for a direct-liability claim under §2333(a)" (Gonzalez v. Google, 2023).

third advocates the imposition of limitations of Section 230's applicability in federal antitrust cases; the last area, instead, aims to foster "open discourse and greater transparency", by "replac(ing) vague terminology in Section 230(c)(2)", by "provid(ing) definition of good faith" and by "explicitly overrul(ing) *Stratton Oakmont*[56]" to avoid the moderator's dilemma, clarifying that "a platform's removal of content pursuant to Section 230(c)(2) or consistent with its terms of service does not, on its own, render the platform a publisher or speaker for all other content on its service. (Department of Justice, 2023).

Looking ahead, however, a reform of Section 230 seems rather unattainable (Leary, 2025, p. 106). The tech industry, in fact, has strategically invested in political campaigns and lobbied policymaking to safeguard the current underregulated legal framework on online content moderation: a recent example can be seen in the high contributions made by many big tech companies like Meta, Apple and Google to Trump's 2024 Presidential campaign, in an effort to gain favors from the new administration which already advocates platform deregulation (Leary, 2025, p. 106-107). Evidently, this *quid pro quo* dynamic between the tech industry and the political institutions has led to many policy concessions in favor of tech companies for political support,

[56] The *Stratton Oakmont, Inc. v. Prodigy Services Co.* is a 1995 Supreme Court case in which the platform Prodigy was sued by the plaintiffs for defamation due to some statement posted on the platform by a third party (Samson, n.d.). According to the litigants, "Prodigy's liability (should have been) tested by the standards applicable to publishers of information, as opposed to distributors" the difference between which lies in this statement by the court:
"one who repeats or otherwise republishes a libel is subject to liability as if he had originally published it. In contrast, distributors such as bookstores and libraries may be liable for defamatory statements of others only if they knew or had reason to know of the defamatory statement at issue. A distributor, or deliverer of defamatory material is considered a passive conduit and will not be found liable in the absence of fault" (Samson, n.d.).
In the end, the Court considered Prodigy liable as a publisher rather than a distributor, because it "exercised (some) control over the content of its bulletin boards" (Samson, n.d.). This case of 1995 is the one which sparked the enactment of the "Good Samaritan" clause of Section 230(c)(1) of the CDA in 1996 which protects users and providers from liability "for actions to restrict or to enable restriction of access to objectionable online material". (Brannon & Holmes, 2024). Importantly, from 1995 to 1996 a shift in the perception of online platforms and online content moderation had occurred, with the Courts overruling their judgment in *Stratton* in the 1997 *Zeran* case.

creating a particularly hard-to-sever relationship between the private and the public sector which renders reforms of Section 230 almost impossible.

4.2 The US Approach to AI and its Implementation in the Field of Online Counterterrorism

The United States has extended its lighter regulatory approach also in the fields of AI regulation and AI use for counterterrorism. The First Amendment protections, the strong US culture of innovation and its market-driven approach to technology, alongside the legacy of the War on Terror and the US focus on national security, have allowed the country to develop a markedly different framework from the EU in these fields. As a matter of fact, the US does not have any "comprehensive federal legislation or regulations" on the development and usage of AI, relying on a decentralized and sector-specific approach to AI governance which emphasizes flexibility and private-sector leadership over prescriptive regulatory frameworks (Anderson et al., 2025). The former Biden administration attempted to introduce greater federal oversight on AI use and development, issuing not only the afore-mentioned Biden EO but also the 2022 *White House Blueprint for an AI Bill of Rights*, "to help guide the design, use and deployment of "automated systems"" emphasizing safeguards such as algorithmic fairness, data privacy, user transparency, and human oversight mechanisms (The White House, 2022, pp. 4-7). However, this more rights-oriented approach has been reversed under the new Trump administration which issued the Removing Barriers EO, abrogating the Biden EO and leaving up to the single states[57] the task of

[57] In West US, the State of Colorado established Regulation SB24-205, obliging "developers of high-risk AI systems to use … "reasonable care" to protect consumers from algorithmic discrimination", while the state of California has enacted the SB-942 California AI Transparency Act which compels "companies developing generative AI systems to provide AI detection tools free of charge and enables users to identify and mark that content has been AI-generated"; In Southwest US, the state of Utah has issued the S.B. 149 Artificial Intelligence Policy Act which

creating laws on AI. State laws, however, must not constitute a source of impairment for President Trump's goal of establishing US global dominance in AI development (Anderson et al., 2025). This goal, alongside the President's focus on national security and border control, allows for an increased use of AI for security purposes without stringent regulations.

The US government has long been advancing toward greater automation in the field of security, recognizing the vast potential of AI technologies: the 2017 Department of Defense's Project Maven inaugurated the use of AI systems for counterterrorism, automating the analysis of intelligence, surveillance, and reconnaissance (ISR) data, but the US intelligence has since then increased the use of AI in all its operations (Rassler, 2021, p. 40). Furthermore, having found in AI a very useful tool to pursue its security goals both offline and online, the US Department of Homeland Security (DHS) has right away implemented this technology in its counterterrorism operations while trying to uphold the fundamental rights of its citizens (Homeland Security, n.d.). It is increasingly investing in AI research to understand new ways to use AI for national security while creating an AI risk register and "a leading AI community by developing strong relationships with AI experts and organizations across public and private industry" capable of understanding future threats and developing new ways to employ AI in the security field (Homeland Security, n.d.). Furthermore, the agency cooperates with the Cybersecurity and Infrastructure Security Agency (CISA) and the Science and Technology Directorate (S&T) in addressing possible threats and advancing the use of AI for national security purposes: CISA, in particular, "uses machine learning and natural language processing models to collect and sort vulnerability data before it is

"creates liability for AI use not properly disclosed that ultimately violates consumer protection laws"; in Midwest US, the state of Illinois established Regulation HB 3773 which "prohibits employers from using AI that could potentially subject employees to unlawful discrimination based on what the Illinois Human Rights Act defines as protected classes"; in Southeast US, the state of Tennessee issued the HB 2091 Ensuring Likeness, Voice and Image Security (ELVIS) Act: which forbids "individuals from using AI to mimic a person's voice without permission"; finally, in Northeast US, the state of New Hampshire has created Regulation HB 1688 which "prohibits state agencies from using AI to surveil or manipulate members of the public" (Holland, 2024).

presented to human analysts" giving the latter the possibility of focusing on more complex evaluations and improving both the efficiency and accuracy of their operations; the S&T, instead, assists security missions by implementing AI to "reduce risk and make data-driven decisions", "to identify patterns that may indicate organized criminal activity" and "to detect illegal …weapons" (Homeland Security, n.d.). Additionally, the DHS is deploying defensive AI technologies, such as malware reverse engineering "to disrupt adversaries' malware development life cycle" and automated cyber vulnerability reporting, to improve threat detection and overall cybersecurity resilience (Homeland Security, n.d.). All these strategies have been paired with the use of AI systems, such as the content classification system which the US has developed in cooperation with the UK Home Office for the "identification, moderation and removal of online terror content by social media and technology companies" (Rassler, 2021, p. 41).

Clearly, the US is already extensively using AI for security purposes but, in many instances, this strategy raises important concerns regarding the protection of fundamental rights, and especially of the right to freedom of speech, due to the limited legal safeguards and oversight in place.

4.3 The Impact of the US Approach on the Balance Between Security and Freedom of Speech in Online Counterterrorism

The United States' growing reliance on AI systems in the field of counterterrorism has significantly affected the fragile balance between security imperatives and the protection of fundamental rights, particularly the right to freedom of expression. This balance had already been severely affected during the War on Terror started after the terrorist attacks of 9/11, a period marked by an extraordinary rise in prosecutions of government whistleblowers and in

"surveillance of journalists and threats of criminal liability directed at the press", juxtaposed by "new legislation targeting terror-related speech and judicial reluctance to vigorously enforce existing free speech protection[58]" (Raban, 2018, p.156). Although the War on Terror officially ended on August 30, 2021, with the withdrawal of the US military forces from Afghanistan, national security remains one of the highest, if not the highest, priority in the United States which has often conceptualized security as a foundational precondition for the enjoyment of other rights rather than a right among others. Hence, with the emergence of artificial intelligence, the US government has integrated such technologies in most of its counterterrorism measures, and particularly in online surveillance and content moderation, with only limited legal supervision due to the lack of a comprehensive legal framework on the use of AI or on the use of AI for security purposes (Anderson et al., 2025). This, paired with the country's approach of platform self-regulation enabled by the broad immunity granted to online service providers by Section 230(c) of the CDA, has often allowed big tech companies broad discretion over deciding what constitutes terrorist or harmful content, often in coordination with government priorities. In the fields of

[58] 9/11 pushed the US to prioritize national security over First Amendment protections, generating an environment where the boundaries of permissible speech became ambiguous. During this period, the government made extensive use of the Espionage Act, "a draconian federal statute … that imposes heavy criminal penalties for the disclosure of classified information", to prosecute whistleblowers, including CIA and State Department officials, while also surveilling journalists and, in some cases, labeling them as co-conspirators through the NSA' "secret surveillance program of electronic communications (Raban, 2018, pp. 142-149). Even the judiciary showed deference to the executive during this period, due to the latter's increased powers to effectively combat the War on Terror. Prosecutions for the dissemination of jihadist propaganda, such as that of Ali al-Timimi, who was "charged and convicted based on his advocacy of the use of force against American troops", revealed a tendency by courts to bypass the 1969 *Brandenburg v. Ohio* standard, which traditionally protects advocacy unless it incites imminent lawless action (Raban, 2018, pp. 152-153). This standard was further eroded by the Supreme Court's decision in *Holder v. Humanitarian Law Project*, which criminalized a group of American NGOs "that counseled the Partiya Karkeran Kurdistan ("PKK") and the Liberation Tigers of Tamil Eelam ("LTTE") … on how to advance their causes by peaceful means" because these associations were dealing with designated terrorist organizations (Raban, 2018, p. 154). Subsequent cases, including that of Tarek Mehanna, demonstrated the government's great power of discretion in penalizing speech even if such speech was not necessarily harmful or inciting to terrororism: Mehanna was, indeed, an American pharmacist "convicted of "knowingly provid[ing] material support" to Al-Qaeda for activities that included translating Arab-language jihadist materials into English and posting the translations on a jihadist website" even if the sole proof "of coordination between Mehanna and a terrorist organization consisted in the fact that the jihadist internet site on which he posted his translations was also used for recruitment by Al-Qaeda" (Raban, 2018, p.155).

security and counterterrorism, in fact, a sort of *quid pro quo* dynamic has emerged, with private tech companies often aligning their content moderation policies with governmental objectives in exchange for regulatory leniency (Leary, 2025, pp. 106-108).

The broad and unregulated applications of AI systems for online surveillance and content moderation have been particularly criticized by civil society organizations such as the American Civil Liberties Union (ACLU). Notably, ACLU's National Security Project's deputy director, Patrick Toomey, has often demanded greater "safeguards and transparency" in the use of AI systems by national security agencies, remarking that:

> "If developing national security AI systems is an urgent priority for the country, then adopting critical rights and privacy safeguards is just as urgent. Without transparency, independent oversight, and built-in mechanisms for individuals to obtain accountability when AI systems err or fail, the policy's safeguards are inadequate and place our civil rights and civil liberties at risk" (ACLU, 2024).

Furthermore, the implementation of automated decisions in online surveillance and content moderation have been highly criticized for their lack of contextual understanding of the different nuances between harmful speech and legitimate expression, often leading to over-censorship or to the disproportionate suppression of speech, particularly of marginalized communities (Funk et al, 2023). All the negative repercussions on free speech of the use of AI systems in online content moderation and counterterrorism which have been highlighted in section 2.2 of this thesis are amplified in the US context, where the regulations of online platforms are much lighter than the ones imposed within the European Single Market.

Moreover, this great privatization of content moderation paired with the absence of comprehensive federal legislation on AI has created a fragmented oversight landscape, especially

when AI is used for security purposes: the Privacy and Civil Liberties Oversight Board (PCLOB), which has been created to ensure that counterterrorism policies respect American citizens' privacy and civil liberties, is not even equipped to oversee or review any "AI systems for national security purposes", thus leaving a great gap in the protection of fundamental rights in security operations (Patel & Toomey, 2024). Since the PCLOB has always "promoted transparency and reform of opaque government programs", Patel and Toomey (2023) argue that its jurisdiction and mandate should be increased to cover "all AI used in national security systems" in an effort to promote the respect of fundamental rights, and especially of the right to freedom of speech. In particular, they underline that this enlarged mandate should: (1) "explicitly include protection of civil rights"; (2) "make clear that the AI oversight authority is charged with reviewing the full AI lifecycle, including the initial decision to deploy the AI system, and any impact assessments and audits that have been performed"; and (3) "be able to recommend halting AI systems in national security contexts if their risks to privacy and civil liberties outweigh their benefits, something which the PCLOB has already done in its previous recommendations for counterterrorism programs" (Patel & Toomey, 2023). Another positive step toward a more balanced approach to security and fundamental rights in the use of AI can be read in the Joint Statement of the Consumer Financial Protection Bureau, the Department of Justice, the Equal Opportunity Commission, and Federal Trade Commission (FTC) on *Enforcement Efforts Against Discrimination and Bias in Automated Systems* which aims to safeguards American citizens "against discrimination and bias in automated systems" by underlining that legal protections should apply to AI-driven practices just as they do to traditional methods (Benizri et al., 2023; Chopra et al., 2023, pp. 1-2).

Although the Biden administration seemed more sensitive to the fragility of the balance between security and fundamental rights whenever AI is used for counterterrorism measures with

the implementation of the Biden EO and of the *Blueprint for an AI Bill of Rights*, these instruments were non-binding and lacked enforcement mechanisms (Anderson & Comstock, 2025). The character of such instruments allowed the current Trump administration to strongly move away from an approach to AI which emphasized the alignment with democratic values and fundamental rights. To achieve "America's global AI dominance in order to promote human flourishing, economic competitiveness and national security", security imperatives can effectively trump the protection of fundamental rights, and especially of the right to freedom of speech, whenever AI is used in counterterrorism, further putting the US' approach to AI in complete antagonism with the one established by the EU with the AI Act (Exec. Order No. 14179, 2025).

4.4 A Comparative Analysis of the US and EU Approaches…

While both the EU and the US recognize the great potential of AI across various sectors, and especially in the field of online counterterrorism and anti-incitement to terrorism, their regulatory approaches differ significantly, particularly in balancing national security imperatives with the protection of fundamental rights, such as the right to freedom of speech.

4.4.1 … *In Protecting Freedom of Speech*

First of all, it is important to underline that while in Europe the right to freedom of speech is a qualified right both under Art. 10 of the ECHR and Art. 11 of the EU Charter, in the US, the First Amendment provides near-absolute protection to the right, even if more stringent limitations have been applied to the right after the beginning of the War on Terror in 2001 as discussed above. This occurs notwithstanding the fact that, under international law, the right to freedom of speech, is considered at all effects a qualified right under Art. 19 of the ICCPR. The US, a ratifying member

of the ICCPR since 1992, obviously recognizes that certain limitations can be implemented with regard to the right to freedom of speech, but its constitutional protection of the right remains strong. This is due to the US' mostly dualist approach to international law, which considers "all forms of international law as being entirely separate from domestic legal processes and" which believes "that international law merely operates at an international level in a community of independent states or that, if it operates at all domestically, it must be implemented or transformed by some formal conduct of a domestic political entity" (Paust, 2013, p. 246). The European Union, instead, exhibits monist tendencies in its approach to international law, as many of its Member States, like France, do, considering "international law (as) … the apex in terms of law's validity and primacy" and embedding most of its regulations within their legal systems without changing them (Paust, 2013, p. 245). More recently, however, the European Court of Justice (ECJ) has shown some dualist tendencies in its decisions, without treating international law or treaties as directly applicable within the EU legal order or by subordinating international law to EU law, following the lead of many American courts. A landmark case in this respect is the *Kadi* case[59], where the ECJ affirmed that international law cannot always override the constitutional principles of the EU, thus prioritizing EU fundamental rights over conflicting international obligations (Kadi and Al

[59] After the 9/11 terrorist attacks, the United Nations Security Council adopted Resolution 1267 (1999), imposing economic sanctions on individuals and entities allegedly associated with Osama Bin Laden, Al-Qaeda, and the Taliban, listed by the UN Sanctions Committee (Kadi and Al Barakaat v Council of the European Union and Commission, 2008). To give effect to this resolution within the EU legal order, the Union, which has exclusive competence in matters of internal market and common foreign and security policy, adopted Regulation 467/2001, which was then challenged by Mr. Kadi and the Al Barakaat Foundation before the Court of First Instance (CFI), arguing that it violated their fundamental rights, including the right to property, the right to be heard, and the right to effective judicial review (Kadi and Al Barakaat v Council of the European Union and Commission, 2008). This case posed a fundamental dilemma for the EU legal order: if the Court denied jurisdiction, it would demonstrate deference to international law, but at the cost of undermining the autonomy of the EU legal system and accepting a potential gap in fundamental rights protection. On 3 September 2008, the ECJ overturned the decision of the CFI that the EU did not have jurisdiction in this case, and held that no international agreement can override the constitutional principles of the EU, which include respect for fundamental rights; furthermore, it found that Kadi's right to be heard and the right to effective judicial review, had been disregarded as well as his right to privacy (Kadi and Al Barakaat v Council of the European Union and Commission, 2008).

Barakaat v Council of the European Union and Commission, 2008). This approach completely overturned the previous approach of the Court which, since the 1998 *A. Racke GmbH&Co. v Hauptzollamt Mainz*[60] case, always underlined "necessity for the (then European Commission) EC to respect international law" (Govaere, 2009, p. 2). In *Kadi*, the ECJ's defense of fundamental rights can be understood in light of the implicit acceptance of the autonomy of EU law by the constitutional courts of the Member States and by the ECHR: the Court had indeed taken this decision also to ensure that its autonomy would not be undermined and that jurisdiction on these matters would not pass to the national courts of the Member States (Govaere, 2009, p. 12). This autonomy becomes, for the ECJ, the "authority to determine the validity and interpretation of EU law as a self-contained and self-referential legal system distinguishable and independent from national and international law"; it is thus "not relative" and "it has the purpose and effect of creating the jurisdictional element of sovereignty" (Eckes, 2020, p. 2). The jurisprudence of the European Union is thus an evolving one, which aims to balance its commitments under international law with its own legal and constitutional standards.

4.4.2 … *In Addressing Content Moderation*

The contrasting legal cultures of these two jurisdictions also influence how content moderation online is approached, reflecting a different vision of the state's role in digital governance. The US model, rooted in First Amendment's protections and free-market logic, favors

[60] In the 1998 A. Racke GmbH&Co. v Hauptzollamt Mainz case, "the Bundesfinanzhof (Federal Finance Court) referred to the Court for a preliminary ruling under Article 177 of the EC Treaty two questions concerning the validity of Council Regulation (EEC) No 3300/91 of 11 November 1991 suspending the trade concessions provided for by the Cooperation Agreement between the European Economic Community and the Socialist Federal Republic of Yugoslavia" for "the importation into Germany of certain quantities of wine originating in the" latter (1. -2.) (A. Racke GmbH&Co. v Hauptzollamt Mainz, 1998). The "Examination of the questions referred has disclosed no factor of such a kind as to affect the validity of Council Regulation (EEC) No 3300/91 of 11 November 1991 suspending the trade concessions provided for by the Cooperation Agreement" (A. Racke GmbH&Co. v Hauptzollamt Mainz, 1998).

speech and innovation, sometimes at the expense of safeguards against discriminatory AI practices; the EU model, instead, emphasizes legal certainty, rights protection, and institutional accountability. Indeed, the European Union has increasingly tried to shield its citizens from breaches of privacy and fundamental rights with the enactment of the GDPR, the DSA and the AI Act, thus trying to impose stricter regulations on the work of big tech companies. Contrarily, the US has allowed greater discretion to the companies thanks to Section 230(c) of the CDA, signaling the government's great commitment to the protection of the private sector, of the market and of innovation. The same tension which arises in both jurisdictions between the public and the private sector appears much greater in the European Union, which puts stronger emphasis on transparency and rights and tries to ensure that platforms respect the rights of its citizens. In the US, instead, the balance between these sectors appears much more skewed towards the private one, with the federal government providing only limited oversight and limited judicial review of the platforms' work. This lighter regulatory approach of content moderation online depends also on the US' market and innovation-driven approach to new technologies, which contrasts with the EU's greater attention to the protection of fundamental rights. Furthermore, the *quid pro quo* relationship between big tech companies and the government is less strong in the European Union than it is in the US, where tech companies lobby in policymaking and fund presidential campaigns to have lighter regulatory frameworks (Leary, 2025, p. 106-107). It is thus evident why big tech companies operating in the US have greater possibility of maneuvering even when implementing AI tools in both its content moderation and its users-engagement strategies than those operating within the European market where there are more stringent rules, for instance on the protection of users' privacy under the GDPR.

4.4.3 … *In AI Governance and Regulation*

The approach taken regarding the protection of the right to freedom of speech and content moderation is also reflected in the two jurisdictions' governance of AI. While the European Union aims to provide a harmonized and risk-based legal framework on the use and development of AI for its Member States with the implementation of the AI Act, the federal government of the United States has preferred to adopt a decentralized, sector-specific and innovation-led approach to the issue. Importantly, while the EU AI Act has provided a single uniform definition of AI, the US government has refrained from adopting a single definition, with only the National Artificial Intelligence Initiative, and consequently the Removing Barriers EO, describing it 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. Artificial intelligence systems use machine and human-based inputs to-
>
> (A) perceive real and virtual environments;
>
> (B) abstract such perceptions into models through analysis in an automated manner; and
>
> (C) use model inference to formulate options for information or action" (15 U.S.C. 9401(3); Art. 3 Exec. Order No. 14179, 2025).

Differently from the EU Member States, the majority of which did not possess any legal definition of AI before the enactment of the AI Act, many of the 50 US states have enacted state-level privacy laws on the definition of AI[61]. For instance, the state of Texas, which created an AI advisory

[61] Other states include the definition of AI in that of 'profiling:' the state of Connecticut created its Public Act No. 22-15 describing profiling as "any form of automated processing performed on personal data to evaluate, analyze or predict personal aspects related to an identified or identifiable individual's economic situation, health, personal preferences, interests, reliability, behavior, location or movements"; the California Consumer Privacy Act (CCPA), instead, considers as profiling "any form of automated processing of personal information, [...] to evaluate certain personal aspects relating to a natural person and in particular to analyze or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behavior, location, or movements" (Anderson et al., 2025). Additionally, Its National Assembly Bill 1008 clarifies that the

council, defines AI as "an algorithm, including an algorithm incorporating machine learning or other artificial intelligence techniques, that uses data-based analytics to make or support governmental decisions, judgments or conclusions" (Anderson et al., 2025). Furthermore, while the EU AI Act defines both enforcement powers and penalties with regard to the use and development of AI, with the creation of the European AI Office and with the requirement to designate at the national level a Market Surveillance Authority (Art. 3(26)), a Notifying Authority (Artt. 3(19) & 28(1)) and a National Public Authority (Art. 77(2)) to oversee the implementation of the AI Act, the US has adopted a fragmented approach leaving up to the states the issues of enforcement and penalties (European Commission, 2025b; Overview, 2025). In Colorado, for example, the Attorney General is the sole authority designated by the *Colorado AI Act* to supervise its enforcement while, in California, the Assembly Bill 2655 (or *Defending Democracy from Deepfake Deception Act*) enables the "Attorney General, any district attorney, or any city attorney (to) seek injunctive relief to compel removal of materially deceptive content" (Anderson et al., 2025). Furthermore, the Senate Bill 942 (or C*alifornia AI Transparency Act*) sets the maximum penalties for violations of the Act at "US$5,000 per violation per day", which is clearly a much lower penalty if compared to the one adopted at the European level (Anderson et al., 2025).

All in all, the jurisdictions' approaches to AI regulation and governance strongly reflect their approaches to fundamental rights protection and to content moderation as well as the balance which they strike between the public and the private sector in these fields. The US enacts a more innovation-driven and libertarian approach, while the EU is adopting an increasingly rights- and risk-based approach to AI development and usage, perfectly reflecting the standards that these

CCPA applies to personal data in any format, including abstract digital information generated by AI systems, while its National Senate Bill 1223 includes in the CCPA also neural data as "sensitive personal information" (Anderson et al. 2025).

distinct jurisdictions apply in their markets. Importantly, then, the new Trump Administration is strongly adapting this innovation-driven approach to AI with limited regulations, to favor the US' global dominance in the field (Exec. Order No. 14179, 2025). This is clearly shown in the worlds of Vice President JD Vance[62] at the February 2025 Paris AI summit, who underlined the US commitment to avoid overregulation of AI systems not to stifle innovation, and to go against the stricter regulation proposed by the EU (Madhani & Adamson, 2025). Indeed, even if former President Biden's *Blueprint for an AI Bill of Rights* has not yet been abrogated, it is hard to imagine that the new President will ever sign it, given his hard stance against overregulation of AI. Indeed, under Article 1 Section 7 of the US Constitution:

> "Every Bill which shall have passed the House of Representatives and the Senate, shall, before it become a Law, be presented to the President of the United States; If he approve he shall sign it, but if not he shall return it, with his Objections to that House in which it shall have originated, who shall enter the Objections at large on their Journal, and proceed to reconsider it" (US Const. art. 1(7)).

4.4.4 … *In the Use of AI for Counterterrorism Purposes*

Turning to the domain of online counterterrorism and anti-incitement to terrorism, both the EU and the US jurisdictions have already implemented AI-powered content moderation, risk analysis, and surveillance measures in their intelligence strategies. Both jurisdictions, in fact, knowing that many other countries worldwide are already using AI systems[63] in their

[62] Vice President JD Vance's exact words were: "at this moment, we face the extraordinary prospect of a new industrial revolution, one on par with the invention of the steam engine. … But it will never come to pass if overregulation deters innovators from taking the risks necessary to advance the ball" (Madhani & Adamson, 2025).

[63] With regard to this, the 2024 DHS' Homeland Threat Assessment specifies: " the proliferation of accessible artificial intelligence (AI) tools likely will bolster our adversaries' tactics. Nation-states seeking to undermine trust in our government institutions, social cohesion, and democratic processes are using AI to create more believable mis-,

counterterrorism strategies and, most importantly, that terrorist groups are deploying these new technologies to accrue their ranks by radicalizing new members, preparing attacks and sending impactful messages to their victims, are employing AI in counterterrorism to safeguard their security. These jurisdictions' efforts to intensify their counterterrorism strategies have certainly arisen as effects of the War on Terror: the shift of attention towards security imperatives has been so great that it still influences decision-making in most countries worldwide, and especially in the US and the EU which have both suffered acts of terror within their territory (Howell & Lind, 2009, p. 47). National security is therefore often prioritized in matters of counterterrorism when weighed against the protection of fundamental rights. This happens especially in the US, but also within many Member States of the European Union, such as France, which, in light of serious attacks or concerns, can establish a state of emergency limiting the protection of fundamental rights of their citizens to ensure their safety and security. Nevertheless, the European Union pushes its Member States to maintain a high level of protection of fundamental rights even during periods of emergency, supervising their limitations: while Art. 52(1) of the EU Charter does not require the EU Member States to explicitly notify the EU institutions of their derogations, the Court of Justice of the European Union can review the limitations imposed by the states especially in relations to EU obligations; furthermore, Art. 15(3)[64] of the ECHR deals exactly with "derogation(s) in time

dis-, and malinformation campaigns, while cyber actors use AI to develop new tools and accesses that allow them to compromise more victims and enable larger-scale, faster, efficient, and more evasive cyber-attacks" (DHS, 2024, p. 7).

[64] Art. 15(3) reads in particular:

> "Any High Contracting Party availing itself of this right of derogation shall keep the Secretary General of the Council of Europe fully informed of the measures which it has taken and the reasons therefor. It shall also inform the Secretary General of the Council of Europe when such measures have ceased to operate and the provisions of the Convention are again being fully executed" (ECHR, 1950).

of emergency" requiring that every limitation made under this article be "officially notified to the Secretary General of the Council of Europe" (ECHR, 1950).

Overall, thanks to the enactment of the AI Act, when AI systems are implemented in counterterrorism in Europe, they will be subjected to more public scrutiny, and governments will have to justify their deployment through risk assessments and fundamental rights impact analysis (Art. 73(6)) (Regulation (EU) 2024/1689, 2024, p. 102). Nevertheless, the Act has been criticized for containing certain evident loopholes which may allow certain AI systems to bypass the control of the Act, potentially undermining the protection of fundamental and digital rights of users and the protection of the rule of law (European Center for Not-for-Profit Law, 2024). This is mainly due to the fact that

> "(1) gaps and loopholes can turn prohibitions into empty declarations;
>
> (2) AI companies' self-assessment of risks jeopardizes fundamental rights protections;
>
> (3) standards for fundamental rights impact assessments are weak; (and)
>
> (4) the use of AI for national security purposes will be a rights-free zone" (European Center for Not-for-Profit Law, 2024).

The Act, however, has still not been applied, hence it will only be truly possible to analyze the real consequences of its applications after August 2026. In the US, instead, security programs can be highly opaque, with AI tools often used in bulk surveillance or secretive watchlists, with limited avenues for redress: for instance, systems of Facial Recognition Technology (FRT) have been implemented by the Federal Bureau of Investigations (FBI), the US Department of Justice (DOJ) and the DHS without any standardized policy for FRT use and without transparent supervision (US Commission on Civil Rights, 2024, p. 100). The lighter supervision of counterterrorism measures in the US depends on all the successes achieved in the field of counterterrorism since

9/11 thanks to the exploitation of "data to better understand and degrade terror networks" (Rassler, 2021, p. 31). In the future, and surely during this Trump's Administration, it is unlikely that the US will increase the control of AI because this move would go against all the goals that this Administration wants to achieve: in particular, more stringent rules on AI would decrease innovation and therefore economic profit, possibly reducing the US' leverage against China in the technological field and they would also decrease the ability of law enforcement agencies to use AI systems in their strategies for increasing national security. Perhaps, in 2028, with the new Presidential elections, there will be a new shift in the US approach to AI and to AI for security purposes which might align the country's strategy with that of the European Union or distance it from the latter even more. Even in the short term, a shift in the current Administration's perspective on this matter could happen, for instance considering the Midterm Elections. These are however just political speculations which may or may not materialize. From a legal perspective, what matters is the current position.

4.5 Pros and Cons of the EU and US AI Governance Models: Insights on Protecting Security and Freedom of Speech in Counterterrorism

The comparative analysis of the European and the American approaches to AI governance and regulation has demonstrated that no perfect model of AI governance worldwide exists yet, and that the global competition among states to shape international standards on this sector may continue for some years. When considering these two Transatlantic approaches, indeed, both models have inherent flaws but also great advantages.

In the US, the absence of both a unified legal framework and of an adequate federal oversight on AI development and use has allowed different states to enact different legislations,

creating a fragmented regulatory framework of AI governance, full of legal gaps and grey areas which can be exploited to use AI systems for security purposes without legal restraints and supervision. This approach can thus lead to major inconsistencies in accountability mechanisms and transparency requirements, posing increased risks to the protection of fundamental rights. On the other hand, the model promoted by the AI Act within the European Single Market depends strongly on the EU Member States' enforcement capabilities which may differ based on the single countries' resources, technical expertise, and commitment to the cause. Moreover, as noted by critics and by the US Vice President, JD Vance, the Act may lead to overregulation, hindering AI innovation and technological progress (Madhani & Adamson, 2025).

Contrarily, the US permissive regulatory framework to AI governance certainly does not obstruct but instead promotes AI innovation by providing developers with flexibility and fewer compliance burdens. This is obviously a strength of the US model which will possibly allow the country to maintain a dominant position in global tech development, particularly vis a vis its strategic competition with China. In 2024 alone, the US devoted "more than 6 billion dollars … including 4.1 billion from the Defense Advanced Research Projects Agency and 2 billion dollars from other related agencies" to foster innovation in the field of advanced technologies, while the European Innovation Council, the agency responsible for this goal in Europe, allocated only 256 million euros in advanced technologies, investing "only 4% of what the US spends" on artificial intelligence (European Commission, 2025c). The European Union, clearly, is not so much focusing on innovation rather than on promoting a comprehensive risk-based AI regulation, which safeguards users' fundamental rights by imposing stringent limitations on high-risk AI systems and requiring frequent mandatory oversight and impact assessments. This is important especially due to the increased use of AI for security and counterterrorism purposes, because it will constrain

AI developers and providers to respect fundamental rights in all their operations. The extraterritoriality of the Act, then, will ensure that also foreign companies working within the EU Single Market will have to uphold its provisions, possibly influencing their work also in other countries worldwide.

All in all, the AI Act, notwithstanding some of its flaws, represents a pivotal attempt to set global standards on AI use and development, following the Union's broader agenda of promoting responsible and human-centric digital governance, as already shown with its implementation of the GDPR and the DSA. Contrarily, the US approach is more focused on promoting its own global leadership in the field of AI development than on setting global standards on AI development and use, emphasizing market freedom and innovation. It thus remains unclear whether a "true global standard for AI regulation" will emerge and whether it will follow the European or the American standards (Benizri, 2023). Perhaps, even a new hybrid model could emerge, combining the strengths of both models, for "a harmonization effort between the two sides of the Atlantic would be necessary to address the challenges posed by the digital ecosystem" (Antonucci, 2024, p. 2). As of now, the AI Act represents the world's 'best bet,' in trying to ensure responsible, transparent and human-centric AI development and use, especially in the field of counterterrorism, where it strives to balance security imperatives with the protection of fundamental rights. Yet, any such regulatory effort must ultimately be tested against the backdrop of time. The rapid pace of technological innovation, especially in AI and other advanced technologies, and the continuous social changes it brings about, have created a situation where static, long-standing rules are no longer sufficient nor effective without updates. Today, legislation and legislators are constantly chasing new developments, trying to regulate new emerging aspects of innovation, even if the legislative processes are much slower than are the advances of technology. This means that for

any legislative framework regulating AI not to become outdated right after it has been enacted or even while it is being decided, laws must be not only coherent but also adaptable and flexible. Perhaps, there will never be a time when law will surpass technological development, but it surely must try to keep up the pace, by reinventing and adapting itself to this current dual reality. Thus, even though the analysis offered in this thesis is relevant now, it will have to be constantly updated and revisited as reality and innovation continue to evolve.

Conclusions

Once seen as a distant technological innovation, artificial intelligence is no longer a vision of the future: everyone knows what it is, everyone is using it and it "is already having a major impact on society", especially when employed in the field of online counterterrorism (Floridi & Cowls, 2019, p. 2). As extremist and terrorist groups increasingly exploit digital platforms and AI systems for propaganda, radicalization, recruitment and incitement to terrorism, indeed, counterterrorism measures have included AI systems to detect, prevent and halt these threats. AI systems are capable of analyzing tons of data more efficiently and in less time than any human being, thus enhancing online content moderation and surveillance, in an attempt to uphold security imperatives. However, the unregulated deployment of such automated systems often lacks transparency and accountability safeguards and can thus potentially infringe users' fundamental rights, and especially their right to freedom of speech. In light of this, the tension between security and fundamental rights in the field of online counterterrorism has grown exponentially, and the need to provide comprehensive legal frameworks on the use and development of AI has become evident. The European Union has thus enacted the AI Act, a first comprehensive attempt worldwide at AI regulation which establishes a risk-based approach to AI development and use, imposing transparency and accountability safeguards on AI developers and providers, such as oversight mechanisms and human rights impact assessments (European Commission, 2025a). This approach which aims at setting global standards for safe and human-centered AI, however, could face many enforcement challenges across Member States and stall Europe's ability to keep pace with AI innovation and technological development. On the other hand, the US approach to AI governance remains more fragmented and deregulatory, striving to foster increasing AI innovation

and technological advancement but providing less safeguards for the protection of fundamental rights especially when AI is used for security purposes. A comparative analysis of the two models thus shows that no perfect model of AI governance and regulation exists yet, but that the EU AI Act is a pivotal step in the right direction to ensure an ethical and rights-based approach to AI, especially in light of its extraterritorial application. However, it will be possible to fully judge the Act's effectiveness only after its full implementation after August 2026.

The coming struggle of this century appears thus evident: providing transparent and ethical AI regulation, perhaps drawing from the EU AI Act, to ensure that, when used for counterterrorism and other security purposes, AI systems respect and protect democratic values, human dignity and fundamental rights, and especially the right to freedom of speech which has always been described as one of the founding milestone of the digital realm. For this to happen, lawmakers must integrate in their law-making process the five principles of "beneficence", "non-maleficence", "autonomy", "justice" and "explicability" for the ethical inclusion of artificial intelligence within human society (Floridi & Cowls, 2019, pp.5-8). These principles must be respected especially when trying to regulate AI use in counterterrorism, because they strive to preserve fundamental rights and democratic values while also allowing a responsible application of AI systems for security purposes. Therefore, any regulation of AI use in the field of counterterrorism, in line with the EU AI Act, should impose more stringent human oversight of AI systems and increased transparency obligations, guaranteeing the accountability of the automated systems by ensuring users' rights to challenge automated decisions and actions in front of an independent oversight body capable of reviewing states and private companies' use of AI. Not only this, but any comprehensive AI regulation should: first, "require formal verification", thus providing mathematical reassurances about system behavior; second, "mandate independent monitoring", with a dedicated regulatory

body capable of overseeing and of intervening on AI systems development and even of recalling unsafe systems; and third, define "red lines" which AI system ought not to cross, "regardless of context or user intent" (Judge et al., 2025, p. 92). As innovation moves forward and new AI systems are developed, new challenges and risks will surely arise, perhaps increasing the possibility for cross-border digital harm or the lack of transparency of AI's decision-making processes. Furthermore, the role of the private sector in the field of AI governance will grow exponentially, enhancing the privatization of online counterterrorism and, consequently, the existing tensions between public and private regulatory responsibilities and between the protection of fundamental rights and security in this field.

Ultimately, the future of AI legislation, especially in the fields of security and counterterrorism, is not just a legal project but a societal one: cooperation among governments, AI and technological experts, organizations, and individuals is indeed pivotal to ensure an ethical development and use of AI by "shar(ing) information, resources, and expertise to develop effective solutions and strategies for addressing the social impacts of AI" (Esmailzadeh & Motaghi, 2024, p. 178). The dangers imposed by AI systems are global and henceforth must be addressed by reinforcing existing international efforts of AI regulation. Not only this, but also the human dimension of AI's technological development should be enhanced, since "best results come from interdisciplinary design processes that incorporate knowledge from human and social sciences right from the beginning" (Salo-Pöntinen & Saariluoma, 2022, pp. 19-20).) The majority of the existing development frameworks on AI is, indeed, solely technical, and can thus limit the extent to which human factors can be meaningfully considered in AI application (Salo-Pöntinen & Saariluoma, 2022, pp. 19-20). Implementing a human dimension in the development of AI from the start would instead ensure a holistic approach to AI development, thus reducing biases,

misclassifications and errors made by automated systems as well as other "negative consequences" in the adoption of AI systems in social life (Salo-Pöntinen & Saariluoma, 2022, pp. 19-20). Importantly, so long as AI systems are developed solely through a technical approach and on "black-box data-driven systems", the legal frameworks regulating AI "will remain approximate and incomplete", in so far as they will not provide transparency and accountability safeguards, and they will more easily fail in protecting fundamental rights such as the right to freedom of speech (Judge et al., 2025, p. 86).

This in turn will negatively impact also the very structure and functioning of democratic governance, of governments and even of states since, beyond fundamental rights, the rise of artificial intelligence also challenges the very structure and functions of democratic governments. Studies have shown that "LLMs are able to write messages that persuade on political issues" during elections campaigns: for instance, "messages crafted by GPT-3 increased support among a representative sample of US voters for a ban on smoking, or a tightening of gun control policy, by about 2-4% on average" (Summerfield et al., 2024, p. 4). Furthermore, it has been found that "AI generated misinformation and disinformation will likely increase voter confusion, create false perceptions of candidates, and fuel cynicism toward the entire electoral process" (Csernatoni, 2024). The employment of AI systems to influence public discourse, information access, and electoral processes may increase political biases and consequently, political polarization: LLMs can indeed trap users into "filter bubbles, (where their prejudices are constantly reinforced)" and "echo chambers (where they are insulated from the discomfort of contrary views)", thus enhancing existing social divides (Summerfield et al., 2024, p. 6). Such instruments in the hands of undemocratic or authoritarian governments may even be used to control the population's political choices and reinforce the hold of the government over power. Notably, China is using AI tools "to

bolster the state's authority", becoming a worldwide "exporter of digital authoritarianism" especially to countries such as "Bangladesh, Colombia, Ethiopia, Guatemala, the Philippines and Thailand"; while countries such as "Iran, Russia and Venezuela are purposefully experimenting with and weaponizing generative AI to manipulate the information space and undermine democracy" (Csernatoni, 2024). Nevertheless, even when employed within democratic legal orders, AI systems used in the fields of public administration or national security without sufficient oversight and transparency and accountability safeguards, may erode the democratic legitimacy of elected assemblies and governments, while also reducing the centrality of parliament and of traditional governmental and constitutional systems: when key decisions are delegated to AI and other digital tools, in fact, this can create a form of technocratic rule, also referred to as 'algocracy', which bypasses traditional democratic safeguards and lacks legitimacy and stringent oversight standards (Volkov, 2025, p. 2). Current constitutions, indeed, were implemented to regulate an analogue world which no longer exists and must necessarily either evolve or embrace new "Constitutions of the Algorithm", in Balanguer's words, to ensure the protection of the democratic order, of fundamental rights, and of digital rights in this new and constantly evolving dual world (Palombino, 2022, pp. 1-2).

Perhaps due to the AI's extraordinary ability to mimic humans, and with them also policy and law makers, individuals tend to forget that those technologies have been developed to serve rather than replace humanity. Henceforth, a more "human compatible approach to AI" should be developed, with the aims of maintaining AI systems under human oversight and decreasing "risks of harm to individuals and society to an acceptable level" (Judge et al., 2025, p. 93). This is especially important for ensuring a fair balance between security imperatives and fundamental rights protection when AI is used in online counterterrorism strategies, an area of action where

often rights infringements are carried out in favor of upholding security and in which the deployment of AI systems only augment this danger due to the opacity of their online decision-making and surveillance processes and to their inherent biases. A human-based approach to AI development and legislation would also limit AI's unduly influence on democratic governance.

Consequently, AI governance in this field should transform from mere reactive policy making into a more anticipatory legal approach, which should be capable of easily adapting to current and future challenges posed by AI systems and evolving alongside technological advancement. AS of now, in fact, legislation is trying to catch up with AI innovation, which goes so much faster than the former, that regulators struggle to provide comprehensive and exhaustive laws on AI. Instead, anticipatory and flexible regulations would actively forecast and prevent potential threats of AI, by mandating constant oversight before, during and after the development of (especially high-risk) AI tools and while the latter are employed. Current AI policy frameworks concentrate solely on the negative aspects of AI, trying to uphold "the right to human decision-making ("human-in-the-loop") and the right to privacy ("data minimization") as the primary solutions that will safeguard the public against the dangers of technology"; however, future regulations should also focus on regulating the positive aspects of AI use by investing on ethical AI development with more "representative and better data to perform accurately and fairly" (Lobel, 2023, p. 1076).

This task undoubtedly resembles that of Sisyphus, condemned by the gods to "ceaselessly rolling a rock to the top of a mountain, whence the stone would fall back of its own weight" (Camus, 1942, p. 75). AI development and innovation, in fact, is just like Sisyphus' rock, because by the time AI laws and regulations are enacted, the technology has already developed, and

traditional legal *iters*, lengthy by nature, struggle to keep up with its advancement. However, in this struggle, there is human dignity: Camus indeed writes that

> "At each of those moments when he leaves the heights and gradually sinks toward the lairs of the gods, he is superior to his fate. He is stronger than his rock" (Camus, 1942, p. 77).

Camus' Sisyphus is unwavering in his effort, making it both the driving force and the purpose of his existence. Just as Sisyphus watches his rock falling back down and goes back to the foot of the mountain to continue his punishment with resolve, so must lawmakers continue refining legal frameworks on ethical and human-centered AI, pushing 'their rock' with a clear vision and making their effort both the driving force and the purpose of their work. Their endeavor must indeed be that of ensuring a fair balance between security imperatives and the protection of fundamental rights. Achieving this balance may be hard but, fortunately, not impossible: with diligent oversight, public accountability, and a rights-based AI regulatory framework, lawmakers can protect both security and fundamental rights, while ensuring an ethical innovation of AI going forward. Importantly, the need for continuous updating extends also to the fields of policymaking and of academic research: given the unprecedented speed of AI and digital innovation, it is not only the traditionally slow legislative processes that must adapt, but also academics and policymakers. Even the analysis of this thesis, as aforementioned, will need to be constantly updated to understand the complex landscape of AI regulation, AI use in counterterrorism, and the consequent effects on the balance between security imperatives and the protection of fundamental rights. Academics and policymakers, whose work is generally more rapid than those of legislators will have to support the latter so as to ensure an adaptable and flexible approach to AI regulation, in an effort to create 'living instruments' of AI regulation which do not risk becoming outdated the moment they are enacted. AI legislation will thus need to be capable of easily evolving to keep up

with AI and digital innovations, ensuring the respect of fundamental rights and of democratic order, if it wants to pass the test of time.

As Bromell (2021, p. 32) has underlined, “the internet displays the best and worst of human enterprise—and a great deal of the mediocre middle”. It will be up to human beings to decide whether the revolution brought about by the development and use of artificial intelligence, especially in the field of counterterrorism, will be one of the best or worst of these enterprises, by either becoming a testament to our highest ideals or a symptom of our inability to govern transformative technologies and to fairly balance security imperatives with the protection of fundamental rights in AI-led counterterrorism.

References

ACLU. (2024). ACLU Warns that Biden-Harris Administration Rules on AI in National Security Lack Key Protections. *ACLU.* https://www.aclu.org/press-releases/aclu-warns-that-biden-harris-administration-rules-on-ai-in-national-security-lack-key-protections.

AGID. (2024). *Italian Strategy for Artificial Intelligence.* Dipartimento Per la Trasformazione Digitale. https://www.agid.gov.it/sites/agid/files/2024-07/Italian_strategy_for_artificial_intelligence_2024-2026.pdf.

Anderson, H., Comstock, E. & Hanson, E. (2025). *AI Watch: Global Regulatory Tracker – United States.* White & Case. https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-united-states.

Antonucci, C. (2024). [Review of the Book *Intelligenza Artificiale e Democrazia. Opportunità e Rischi di Disinformazione e Discriminazione,* by O. Pollicino & P. Dunn]. *Nomos, 3,* pp. 1-5. https://www.nomos-leattualitaneldiritto.it/wp-content/uploads/2025/02/2.-Antonucci_Recensione.pdf.

A. Racke GmbH&Co. v Hauptzollamt Mainz, Case C-162/96. (1998). https://www.ilsa.org/Jessup/Jessup15/Second%20Batch/Racke.pdf.

Asilomar AI Principles, August 11, 2017. Future for Life Institute. https://futureoflife.org/open-letter/ai-principles/.

Bąkowski, P. (2023). *Understanding EU Counter-Terrorism Policy.* European Parliament. https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739395/EPRS_BRI(2023)739395_EN.pdf.

Basic Law for the Federal Republic of Germany, May 23, 1949 (As amended on 22 March 2025). https://www.btg-bestellservice.de/pdf/80201000.pdf.

Bazarkina, D. (2023). Current and Future Threats of the Malicious Use of Artificial Intelligence by Terrorists: Psychological Aspects. In E. Pashentsev (Ed.), *The Palgrave Handbook of Malicious Use of AI and Psychological Security* (pp. 251-272). Palgrave Macmillan. https://link.springer.com/book/10.1007/978-3-031-22552-9.

Bejan, T.M. (2019). The Concepts of Freedom (of Speech). *Proceedings of the American Philosophical Society, vol. 163(2),* pp. 95-107. https://www.amphilsoc.org/sites/default/files/2020-03/attachments/Bejan.pdf.

Benizri, I., Evers, A., Mercer, S.T., Jessani, A.A. (2023). *A Comparative Perspective on AI Regulation.* Lawfare. https://www.lawfaremedia.org/article/a-comparative-perspective-on-ai-regulation.

Brannon, V.C. & Holmes, E.N. (2024). *Section 230: An Overview*. Congress.Gov. https://www.congress.gov/crs-product/R46751.

Bromell, D. (2021). *Regulating Free Speech in a Digital Age - Hate, Harm and the Limits of Censorship.* Springer. https://doi.org/10.1007/978-3-030-95550-2.

Cabrera, L.L. (2024). *EU AI Act Brief - Pt. 3, Freedom of Expression.* Center for Democracy and Technology. https://cdt.org/insights/eu-ai-act-brief-pt-3-freedom-of-expression/.

Calvo, A., Pena, C. & Soberon, M. (2024). *AI Watch: Global Regulatory Tracker - Spain.* White & Case. https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-spain.

Camus, A. (1942). *The Myth of Sisyphus.* Translated by O'Brien, J. (1955). Dominican House of Studies. https://dhspriory.org/kenny/PhilTexts/Camus/Myth%20of%20Sisyphus-.pdf.

Capone, F. & Iaselli, V. (2024). *La Privacy nell'AI Act e nel DDL Italiano Sull'Intelligenza Artificiale: Perché è un Tema Ineludibile.* Agenda Digitale EU. https://www.agendadigitale.eu/sicurezza/la-privacy-nellai-act-e-nel-ddl-italiano-sullintelligenza-artificiale-perche-e-un-tema-ineludibile/?utm_campaign=ad-sicurezza_nl_20240705&utm_source=ad-sicurezza_nl_20240705&utm_medium=email&sfdcid=*%7CSFDCID%7C*.

Carpentier, A. (2024). *How Algorithmic Video Surveillance was Used During the Paris Olympics.* Le Monde. https://www.lemonde.fr/en/france/article/2024/08/16/how-algorithmic-video-surveillance-was-used-during-the-paris-olympics_6716745_7.html.

Carrer, L. & Coluccini, R. (2021). Italy Wins: DPA Blocks Facial Recognition System and MP Proposes Moratorium. *Reclaimyourface.* https://reclaimyourface.eu/italy-dpa-blocks-facial-recognition-system-mp-proposes-moratorium/.

Casaburo, D. & Marsh, I. (2024). Ensuring Fundamental Rights Compliance and Trustworthiness of Law Enforcement AI Systems: The ALIGNER Fundamental Rights Impact Assessment. *AI Ethics, 4,* pp. 1569-1582. https://doi.org/10.1007/s43681-024-00560-0.

CEP. (2025). *Extremist Content Online: Pro-ISIS TikTok Account Spreading Bomb-Making Information and Knife Attack Instructions Located.* Counter Extremism Project. https://www.counterextremism.com/press/extremist-content-online-pro-isis-tiktok-account-spreading-bomb-making-information-and-knife.

Chiarella, M.L. (2023). Digital Market Act (DMA) and Digital Services Act (DSA): New Rules for the EU Digital Environment. *Athens Journal of Law, 9(1),* pp. 33-58. https://heinonline.org/HOL/Page?handle=hein.journals/atnsj9&id=33.

Christchurch Call. (2024). *The Christchurch Call to Action.* The Christchurch Call. https://www.christchurchcall.org/content/files/2024/06/Christchurch-Call-full-text-English-1.pdf.

Christchurch Call. (n.d.). *Our Work.* The Christchurch Call. https://www.christchurchcall.org/our-work/.

Chopra, R., Clarcke, K. Burrows, C.A. & Khan, L.M. (2023). Joint Statement on Enforcement Efforts Against Discrimination and Bias in Automated Systems. *Federal Trade Commission*, pp. 1-3. https://www.ftc.gov/system/files/ftc_gov/pdf/EEOC-CRT-FTC-CFPB-AI-Joint-Statement%28final%29.pdf.

CNIL. (2023). *Artificial Intelligence: The Action Plan of the CNIL*. CNIL. https://www.cnil.fr/en/artificial-intelligence-action-plan-cnil.

Constitution of the Italian Republic, December 22, 1947. https://www.cortecostituzionale.it/documenti/download/pdf/The_Constitution_of_the_Italian_Republic.pdf.

Council of Europe. (n.d.). *The Council of Europe & Artificial Intelligence.* Council of Europe. https://www.coe.int/en/web/artificial-intelligence.

Council of Ministers. (2024). *The Government of SPain Approves the Artificial Intelligence Strategy 2024.* La Moncloa. https://www.lamoncloa.gob.es/lang/en/gobierno/councilministers/paginas/2024/20240514-council-press-conference.aspx#:~:text=The%20Government%20of%20Spain%20approves%20the%20Artificial%20Intelligence%20Strategy%202024,-

Council%20of%20Ministers&text=The%20strategy%20enhances%20supercomputing%20and,languages%2C%20and%20incorporates%20ethical%20elements.

Crialese, E. (Director). (2006). *Nuovomondo* [Film]. Rai Cinema, Memento Films Production, Arte France Cinéma, Titti Film & Respiro.

Csernatoni, R. (2024). *Can Democracy Survive the Disruptive Power of AI?* Carnegie – Endowment for International Peace. https://carnegieendowment.org/research/2024/12/can-democracy-survive-the-disruptive-power-of-ai?lang=en.

Dal Bello, G., Hirsch-Hoefler, S. & Canetti, D. (2024). *AI Video Surveillance at the 2024 Paris Olympics.* The Loop - ECPR's Political Science Blog. https://theloop.ecpr.eu/ai-video-surveillance-at-the-2024-paris-olympics/#:~:text=Surveillance%20versus%20terrorism&text=While%20the%20eyes%20of%20the,thousands%20of%20spectators%20and%20athletes.

DDL n. 1146, Legislatura 19°, April 23, 2024. https://www.senato.it/japp/bgt/showdoc/19/DDLPRES/0/1418921/index.html?part=ddlpres_ddlpres1-articolato_articolato1.

DFKI AI – German Research Center for Artificial Intelligence. (2021). *AI vs. Cybercrime: Research Cooperation with BKA and LKA Launched.* DFKI AI. https://www.dfki.de/en/web/news/research-cooperation-bka-lka.

De La Cuesta, J.L. (2007). Anti-Terrorist Legislation and the Rule of Law: Spanish Experience. *Association Internationale De Droit Pénal.* https://www.penal.org/sites/default/files/files/JLDLCTerrorism.pdf.

Department of Justice. (2023). *Department of Justice's Review of Section 230 of the Communications Decency Act of 1996.* Archives – US Department of Justice. https://www.justice.gov/archives/ag/department-justice-s-review-section-230-communications-decency-act-1996.

DHS – Office of Intelligence and Analysis. (2024), *Homeland Threat Assessment.* Homeland Securit – DHS, 1-38. .https://www.dhs.gov/sites/default/files/2023-09/23_0913_ia_23-333-ia_u_homeland-threat-assessment-2024_508C_V6_13Sep23.pdf.

Dieu, O. & Montasari, R. (2022). How States' Recourse to Artificial Intelligence for National Security Purposes Threatens Our Most Fundamental Rights. In R. Montasari (Ed.), *Artificial Intelligence and National Security.* (pp. 19-45). Springer. https://doi.org/10.1007/978-3-031-06709-9.

Directive (EU) 2017/541 of the European Parliament and of the Council. (2017). Eur-Lex. https://eur-lex.europa.eu/eli/dir/2017/541/oj/eng.

ECHR - European Convention on Human Rights, November 4, 1950. https://www.echr.coe.int/documents/d/echr/convention_ENG.

Editorial Team. (2024). *Enhancing Counterterrorism Through Private Sector Collaboration.* Total Military Insight. https://totalmilitaryinsight.com/counter-terrorism-and-private-sector/.

Eckes, C. (2020). The Autonomy of the EU Legal Order. *Europe and the World: A Law Review, 19*, pp. 1-19. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3647153.

Elysee. (2025). *Make France an AI Powerhouse.* Elysee. https://www.elysee.fr/admin/upload/default/0001/17/d9c1462e7337d353f918aac7d654b896b77c5349.pdf.

Elliott, B. & Silverman, K. (2024). *Artificial Intelligence Law.* Lexology. https://www.lw.com/admin/upload/SiteAttachments/Lexology-In-Depth-Artificial-Intelligence-Law-France.pdf.

English Bill of Rights. (1689). *Yale Law School – The Avalon Project.* https://avalon.law.yale.edu/17th_Century/england.asp.

Erdos, D. (2019). The Development of European Human Rights and Freedom of Expression Law. In *European Data Protection Regulation, Journalism, and Traditional Publishers: Balancing on a Tightrope?* (pp. 19-34). Oxford Data Protection & Privacy Law. https://doi.org/10.1093/oso/9780198841982.003.0002.

Esmailzadeh, Y. & Motaghi, E. (2024). International Terrorism and Social Threats of Artificial Intelligence. *Journal of Globalization Studies, 15(1),* pp. 168-179. https://doi.org/10.30884/jogs/2024.01.09.

EU Charter - Charter of Fundamental Rights of the European Union, December 7, 2000, https://www.europarl.europa.eu/charter/pdf/text_en.pdf.

European Center for Not-For-Profit Law. (2024). *Packed with Loopholes: Why the AI Act Fails to Protect Civil Space and the Rule of Law.* European Center for Not-For-Profit Law. https://ecnl.org/news/packed-loopholes-why-ai-act-fails-protect-civic-space-and-rule-law.

European Center for Not-For-Profit Law. (2022). *EU AI Act Needs Clear Safeguards for AI Systems for Military and National Security Purposes.* European Center for Not-For-Profit Law. https://ecnl.org/news/eu-ai-act-needs-clear-safeguards-ai-systems-military-and-national-security-purposes.

European Commission. (2021). *Germany AI Strategy Report.* AI Watch - European Commission. https://ai-watch.ec.europa.eu/countries/germany/germany-ai-strategy-report_en.

European Commission. (2025a). *AI Act.* Shaping Europe's Digital Future - European Commission. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai.

European Commission. (2025b). *Governance and Enforcement of the AI Act.* Shaping Europe's Digital Future - European Commission. https://digital-strategy.ec.europa.eu/en/policies/ai-act-governance-and-enforcement.

European Commission. (2025c). *The EU Invests in Artificial Intelligence only 4% of What the US Spends.* Newsroom – European Commission. https://ec.europa.eu/newsroom/eismea/items/864247/en.

European Commission. (n.d.a). *Fundamental Rights.* Migration and Home Affairs. https://home-affairs.ec.europa.eu/networks/european-migration-network-emn/emn-asylum-and-migration-glossary/glossary/fundamental-rights_en#:~:text=An%20absolute%20right%20is%20a,a%20declared%20state%20of%20emergency.

European Commission. (n.d.b). *The EU Code of Conduct on Countering Illegal Hate Speech Online.* Racism and Xenophobia - European Commission. https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en.

European Group on Ethics in Science and New Technologies. (2018). *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems.* European Commission. https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1#.

Europol. (2024). *AI and Policing - The Benefits and Challenges of Artificial Intelligence for Law Enforcement.* Europol. https://www.europol.europa.eu/cms/sites/default/files/documents/AI-and-policing.pdf.

Exec. Order No. 14179 on Removing Barriers to American Leadership in Artificial Intelligence. (2025). *The White House.* https://www.whitehouse.gov/presidential-actions/2025/01/removing-barriers-to-american-leadership-in-artificial-intelligence/.

Exec. Order No. 14110 on Safe, Secure and Trustworthy Development and Use of Artificial Intelligence, 3 C.F.R. (2023). *Federal Register*. https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence.

Fabbrini, F. (2016). The Principle of Subsidiarity. In T. Tridimas & R. Schütze (Eds.), Oxford Principles of EU Law (OUP 2016*), iCourts Working Paper Series No. 66,* pp. 1-26. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2781845.

Farrow, R. (2022). *How Democracies Spy on Their Citizens.* The New Yorker. https://www.newyorker.com/magazine/2022/04/25/how-democracies-spy-on-their-citizens.

Feldman, S.M. (2017). Postmodern Free Expression: Philosophical Rationale for the Digital Age. *Marquette Law Review, 100(4),* 1123-1192. https://heinonline.org/HOL/P?h=hein.journals/marqlr100&i=1155.

15 U.S.C. 9401 – Definitions. *US Code.* https://uscode.house.gov/view.xhtml?req=(title:15%20section:9401%20edition:prelim).

Floridi, L. & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review, 1(1),* pp. 1-14. https://doi.org/10.1162/99608f92.8cd550d1.

47 US Code §230 – Protection for Private Blocking and Screening of Offensive Material. *Cornell Law School.* https://www.law.cornell.edu/uscode/text/47/230.

Funk, A., Shahbaz, A. & Vesteinsson, K. (2023). *The Repressive Power of Artificial Intelligence.* Freedom House. https://freedomhouse.org/report/freedom-net/2023/repressive-power-artificial-intelligence.

Ganapini, M.B., Lanteigne, C. & Gupta, A. (2020). Report prepared by the Montreal AI Ethics Institute for the Santa Clara Principles for Content Moderation. *Montreal AI Ethics Institute.* https://arxiv.org/pdf/2007.00700.

Ganor, B. (2018). Artificial or Human: A New Era of Counterterrorism Intelligence? *Routledge - Taylor & Francis Group.* https://www.tandfonline.com/doi/epdf/10.1080/1057610X.2019.1568815?needAccess=true.

GIFCT. (2023). *Advances in Hashing for Counterterrorism.* GIFCT. https://gifct.org/2023/03/29/advances-in-hashing-for-counterterrorism/.

GIFCT. (n.d.). *Preventing Terrorists and Violent Extremists from Exploiting Digital Platforms.* GIFCT. https://gifct.org/.

Global Institute for National Capability. (2024). *Germany's National AI Strategy.* Global Institute for National Capability. https://www.ginc.org/germanys-national-ai-strategy/.

Gobierno de España. (2020). *National Artificial Intelligence Estrategy (ENIA).* Ministerio de Economía y Empresa. https://portal.mineco.gob.es/RecursosArticulo/mineco/ministerio/ficheros/National-Strategy-on-AI.pdf.

Gobierno de España. (2021). *Carta Derechos Digitales.* Ministerio de Economía y Empresa. https://www.lamoncloa.gob.es/presidente/actividades/Documents/2021/140721-Carta_Derechos_Digitales_RedEs.pdf.

Gonzalez v. Google LLC, 598 US (2023). *Global Freedom of Expression – Columbia University.* https://globalfreedomofexpression.columbia.edu/cases/gonzalez-v-google-scotus-2023/.

Govaere, I. (2009). The Importance of International Developments in the Case Law of the European Court of Justice: Kadi and the Autonomy of the EU Legal Order. *European Legal Studies*, pp. 1-19. https://aei.pitt.edu/44312/1/research_paper_1_2009_govaere.pdf.

GPDP. (2022). *Riconoscimento Facciale: Il Garante Privacy Sanziona Clearview per 20 Milioni di euro. Vietato l'Uso dei Dati Biometrici e il Monitoraggio Degli Italiani.* GPDP - Garante per la Protezione dei Dati Personali. https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9751323.

Gruberg, M. (2009). *Voltaire.* Free Speech Center. https://firstamendment.mtsu.edu/article/voltaire/#:~:text=Voltaire%E2%80%99s%20prolific%20biting%20satire%20and,1906.

Guarrigues. (2015). *Organic Law 13/2015 Amending the Criminal Procedure Law to Strengthen Procedural Guarantees and Regulate Technology Related Investigations Measures.* Garrigues. https://www.garrigues.com/en_GB/new/organic-law-132015-amending-criminal-procedure-law-strengthen-procedural-guarantees-and-regulate.

Gunton, K. (2022). The Use of Artificial Intelligence in Content Moderation in Countering Violent Extremism on Social Media Platforms. In R. Montasari (Ed.), *Artificial Intelligence and National Security.* (pp. 69-79). Springer. https://doi.org/10.1007/978-3-031-06709-9.

Hainsdorf, C. & Liard, B. (2024). *AI Watch-Global Regulatory Tracker - France.* White & Case. https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-france.

Hasimi, L. & Poniszewska-Marańda, A. (2024). Detection of Disinformation and Content Filtering Using Machine Learning: Implications to Human Rights and Freedom of Speech. *ROMCIR@ECIR*, pp. 68-77. https://ceur-ws.org/Vol-3677/paper6.pdf.

Hickman, T. & Lorenz, S. (2024). *AI Watch: Global Regulatory Tracker - European Union.* White & Case. https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-european-union.

Hogefeld v. Germany, Application No. 35402/97. (2000). HUDOC – ECHR. https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-5103%22]}.

Holland, M. (2024). *A State-by-State Guide to AI Laws in the US.* TechTarget. https://www.techtarget.com/searchenterpriseai/feature/A-state-by-state-guide-to-AI-laws-in-the-US.

Homeland Security. (n.d.). *Using AI to Secure the Homeland.* DHS. https://www.dhs.gov/ai/using-ai-to-secure-the-homeland.

Howard, J.F. (2024). *Freedom of Speech*. Standford Encyclopedia of Philosophy. https://plato.stanford.edu/entries/freedom-speech/#FutuFreeSpeeTheoPlatEthi.

Howell, J. & Lind, J. (2009). Government-Civil Society Relations Post-9/11. In *Counterterrorism, aid and civil society: before and after the war on terror*, (pp. 46-74). Palgrave Macmillan, https://books.google.it/books?hl=it&lr=&id=nNl8DAAAQBAJ&oi=fnd&pg=PP1&dq=war+on+terror+effect+on+counterterrorism+&ots=BAof_F_sb6&sig=kmpq-pimAs49yxuxg0vRu7ZNgZU&redir_esc=y#v=onepage&q&f=false.

ICCPR - International Covenant on Civil and Political Rights, December 16, 1966, https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights.

IEEE – Institute of Electrical and Electronics Engineers Standards Association. (2017). *Ethically Aligned Design: A Vision for Prioritizing Human Well-Being with Autonomous and Intelligent Systems.* IEEE. https://sagroups.ieee.org/global-initiative/wp-content/uploads/sites/542/2023/01/ead1e.pdf.

Jerome Hall Law Library. (2019). *Legal Dissertation: Research and Writing Guide.* Maurer School of Law. https://law.indiana.libguides.com/dissertationguide#:~:text=Generally%2C%20a%20methodology%20section%20will,data%20collection%20tools%20were%20administered.

Judge, B., Nitzberg, M. & Russell, S. (2025). When Code Isn't Law: Rethinking Regulation for Artificial Intelligence. *Policy and Society, 44(1),* pp. 85-97. https://doi.org/10.1093/polsoc/puae020.

Kausik, A.K. & Rashid, A.B. (2024). AI Revolutionizing Industries Worldwide: A Comprehensive Overview of its Diverse Applications. *Hybrid Advances, 7.* https://doi.org/10.1016/j.hybadv.2024.100277.

Key, F.S. (1814). The Star-Spangled Banner [Song]. *The Lyrics.* https://amhistory.si.edu/starspangledbanner/the-lyrics.aspx.

Kilpatrick, J. (2020). *When a Temporary State of Emergency Becomes Permanent – France as a Case Study.* Transnational Institute, pp. 1-32. https://www.tni.org/en/publication/when-a-temporary-state-of-emergency-becomes-permanent.

Krüger, M. (2025). *German vs European AI Regulation: Comparison of the German Approach to AI Regulation with the European AI Act.* Linvelo. https://linvelo.com/german-vs-european-ai-regulation-comparison-of-the-german-approach-to-ai-regulation-with-the-european-ai-act/.

Leary, M.G. (2025). The Failed Experiment of Section 230 of the Communications Decency Act: How it Facilitates Exploitation and How it Must Be Reformed. *Villanova Law Review, 70(1),* pp. 48-113. https://digitalcommons.law.villanova.edu/cgi/viewcontent.cgi?article=3672&context=vlr.

Lenaerts, K. & Sijter, E. (2001). The Charter and the Rule of the European Courts. *Sage Journals, 8(1),* pp. 90-101. https://doi.org/10.1177/1023263X0100800107.

Leyla Şahin v. Turkey, Application No. 44774/98. (2005). HUDOC - ECHR. https://hudoc.echr.coe.int/fre#{%22itemid%22:[%22001-70956%22]}.

Lobel, O. (2023). The Law of AI for Good. *Florida Law Review, 75(6),* pp. 1073-1139. https://dx.doi.org/10.2139/ssrn.4338862.

Loi n° 55-385 du 3 avril 1955 relative à l'état d'urgence, (1955). Legifrance. https://www.legifrance.gouv.fr/loda/id/JORFTEXT000000695350.

Loi n° 2015-912 du 24 juillet 2015 relative au renseignement. (2015). Legifrance. https://www.legifrance.gouv.fr/loda/id/JORFTEXT000030931899.

Lorenz, S. (2024). *AI Watch: Global Regulatory Tracker - Germany.* White & Case. https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-germany.

Macklin, G. (2019). The Christchurch Attacks: Livestream Terror in Viral Video Age. *CTC Sentinel, 12(6),* pp. 18-29. https://ctc.westpoint.edu/christchurch-attacks-livestream-terror-viral-video-age/.

Madhani, A. & Adamson, T. (2025). *JD Vamce Rails Against 'Excessive' AI Regulation in a Rebuke to Europe at the Paris AI Summit.* AP News. https://apnews.com/article/paris-ai-summit-vance-1d7826affdcdb76c580c0558af8d68d2.

Meta. (2021). *Our New System to Help Tackle Harmful Content.* Meta. https://about.fb.com/news/2021/12/metas-new-ai-system-tackles-harmful-content/.

Meta. (n.d.). *Community Standards.* Meta. https://transparency.meta.com/en-us/policies/community-standards/.

Mills, J.S. (1859). *On Liberty*. The Project Gutemberg Ebook, released in 2011. https://www.gutenberg.org/files/34901/34901-h/34901-h.htm#Page_28.

Milmo, D. (2025). *Global Disunity, Energy Concerns and the Shadow of Musk: Key Takeaways from the Paris AI Summit.* The Guardian. https://www.theguardian.com/technology/2025/feb/14/global-disunity-energy-concerns-and-the-shadow-of-musk-key-takeaways-from-the-paris-ai-summit.

Montreal Declaration for a Responsible Development of Artificial Intelligence, December 4, 2018. https://declarationmontreal-iaresponsable.com/wp-content/uploads/2023/04/UdeM_Decl_IA-Resp_LA-Declaration-ENG_WEB_09-07-19.pdf.

Nordin, J., Giles, I. & Graves, P. (2023). *The Swedish Freedom of the Press Ordinance of 1766 – Background and Significance.* Kungliga Biblioteket.

https://lucris.lub.lu.se/ws/portalfiles/portal/159022995/The_Swedish_Freedom_of_the_Press_Ordinance_of_1766.pdf.

OHCHR. (2021). *Counter-Terrorism Measures Increasingly Diverging from Human Rights Standards, UN Expert Warns.* United Nations - Office of the High Commissioner. https://www.ohchr.org/en/press-releases/2021/10/counter-terrorism-measures-increasingly-diverging-human-rights-standards-un.

Oh, D. & Downey, J. (2024). Does Algorithmic Content Moderation Promote Democratic Discourse? Radical Democratic Critique of Toxic Language AI. *Information, Communication and Society,* pp. 1-20. https://doi.org/10.1080/1369118X.2024.2346531.

Ortutay, B. (2025). *Meta Rolls Back Hate Speech Rules and Zuckerberg Cites 'Recent Elections' as a Catalyst.* AP News. https://apnews.com/article/meta-facebook-hate-speech-trump-immigrant-transgender-41191638cd7c720b950c05f9395a2b49.

OSCE. (2024). *Artificial Intelligence in the Context of Preventing and Countering Violent Extremism and Terrorism: Challenges, Risks, and Opportunities.* https://www.osce.org/files/f/documents/4/f/575877.pdf.

Overview of All AI Act National Implementation Plans. (2025). EU Artificial Intelligence Act. https://artificialintelligenceact.eu/national-implementation-plans/.

Palombino, G. (2022). [Review of the Book *La Constitución del Algoritmo,* by F. Balaguer Callejón]. *Nomos, 3,* pp. 1-4. https://www.nomos-leattualitaneldiritto.it/wp-content/uploads/2023/02/8-RECENSIONE-PALOMBINO.pdf.

Pashentsev, E. (Ed.). 2023. *The Palgrave Handbook of Malicious Use of AI and Psychological Security.* Palgrave Macmillan. https://link.springer.com/book/10.1007/978-3-031-22552-9.

Patel, F. & Toomey, P.C. (2023). *An Oversight Model for AI in National Security: The Privacy and Civil Liberties Oversight Board.* Just Security. https://www.justsecurity.org/94999/an-oversight-model-for-ai-in-national-security-the-privacy-and-civil-liberties-oversight-board/.

Paust, J.J. (2013). Basic Forms of International Law and Monist, Dualist, and Realist Perspectives. In M. Novakovic (Ed.), *Basic Concepts of Public International Law – Monism & Dualism.* (pp. 244-265). University of Houston Law Center No. 2013-A-11, SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2293188.

Raban, O. (2018). Some Observations on the First Amendment and the War on Terror. *Tulsa Law Review, 53(2),* pp. 141-158. https://heinonline.org/HOL/P?h=hein.journals/tlj53&i=162.

Rassler, D. (2021). Commentary: Data, AI, and the Future of US Counterterrorism: Building An Action Plan. *CTS Sentinel, 14(8),* pp. 31-44. https://ctc.westpoint.edu/wp-content/uploads/2021/10/CTC-SENTINEL-082021.pdf.

Rees, M. (2015). *Intelligence Bill: All the Black Spots Denounced by the CNIL - Towards a Chain Reaction of Surveillance.* Next. https://next.ink/18857/93509-projet-loi-sur-renseignement-tous-points-noirs-denonces-par-cnil/.

Regulation (EU) 2019/1020 of the European Parliament and of the Council. (2019). Eur-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32019R1020.

Regulation (EU) 2016/679 of the European Parliament and of the Council. (2016). Eur-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679.

Regulation (EU) 2021/784 of the European Parliament and of the Council. (2021). Eur-Lex. https://eur-lex.europa.eu/legal-

content/EN/TXT/HTML/?uri=CELEX%3A32021R0784#:~:text=This%20Regulation%20aims%20to%20ensure,public%20security%20across%20the%20Union.

Regulation (EU) 2022/2065 of the European Parliament and of the Council. (2022). Eur-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R2065.

Regulation (EU) 2024/1689 of the European Parliament and of the Council. (2024). Eur-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689.

Şahin v. Turkey, No. 44774/98. (2005). *Global Freedom of Expression – Columbia University.* https://globalfreedomofexpression.columbia.edu/cases/sahin-v-turkey/.

Salo-Pöntinen, H. & Saariluoma, P. (2022). Reflections on the Human Role in AI Policy Formulation: How do National AI Strategies View People?. *Discover Artificial Intelligence, 2(1),* pp. 1-24. https://doi.org/10.1007/s44163-022-00019-3.

Samson, M. (n.d.). *Stratton Oakmont, Inc. et al. v. Prodigy Services Company, et al., No. 18-506 (1995).* Internet Library of Law and Court Decisions. http://www.internetlibrary.com/cases/lib_case80.cfm.

Sanoma Uitgevers B.V. v. the Netherlands, Application No. 38224/03. (2010). HUDOC – ECHR. https://hudoc.echr.coe.int/eng#{%22itemid%22:[%22001-100448%22]}.

Siagian, R., Siahaan, L. & Hamzah, M.I. (2023). Human Rights in the Digital Era: Online Privacy, Freedom of Speech, and Personal Data Protection. *Journal of Digital Learning and Distance Education, 2(1),* pp. 548-558. https://doi.org/10.56778/jdlde.v2i4.149.

Summerfield, C., Argyle, L., Bakkerm M., Collins, T., Durmus, E., Eloundou, T., Gabriel, I., Ganguli, D., Hackenburg, K., Hadfield, G., Hewitt, L., Huang, S., Landemore, H., Marchal, N., Ovadya, N., Procaccia, A., Risse, M., Schneier, B., Seger, E., Siddarth, D., Skaug

Sætra, H., Tessler, M.H. & Botvinick, M. (2024). How Will Advanced AI Systems Impact Democracy? *Cornell University*, pp. 1-25. https://arxiv.org/pdf/2409.06729.

TATE - Tech Against Terrorism Europe. (n.d.). *European Regulation on Terrorist Content Online (TCO) - Briefing.* TATE. https://www.techagainstterrorism.org/hubfs/TCO-Guide.pdf.

Tao, Y. (2023). *The Criminalization of Incitement to Terrorism from an International Perspective.* Springer. https://doi.org/10.1007/978-3-031-34370-4.

TEU – Consolidated Version of the Treaty on European Union, October 26, 2012. *Eur-Lex.* https://eur-lex.europa.eu/resource.html?uri=cellar:2bf140bf-a3f8-4ab2-b506-fd71826e6da6.0023.02/DOC_1&format=PDF.

TFEU – Consolidated Version of the Treaty of the Functioning of the European Union, October 26, 2012. *Eur-Lex.* https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:12012E/TXT:en:PDF.

The Constitution of the Fifth Republic. (1958). Élisée. https://www.elysee.fr/en/french-presidency/constitution-of-4-october-1958.

The Declaration of the Rights of Man and of the Citizen. (1789). Élisée. https://www.elysee.fr/en/french-presidency/the-declaration-of-the-rights-of-man-and-of-the-citizen.

The Federal Government. (2014). *Digital Agenda 2014-2017.* Federal Ministry of Interior and Community, pp. 1-36. https://www.bmi.bund.de/SharedDocs/downloads/EN/publikationen/2014/digital-agenda.html.

The Spanish Constitution, December 29, 1978 (As Amended on February 17, 2024). https://www.tribunalconstitucional.es/es/tribunal/normativa/normativa/constitucioningles.pdf.

The White House. (2022). Blueprint for an AI Bill of Rights – Making Automated Systems Work for the American People. *The White House*, pp. 1-73. https://bidenwhitehouse.archives.gov/ostp/ai-bill-of-rights/.

Thibout, C. (2018). *Villani's Report: Defence at the Age of AI.* IRIS - Institut de Relations Internationales et Stratégiques. https://www.iris-france.org/en/110108-villanis-report-defence-at-the-age-of-ai/.

Thormundsson, B. (2024). *Artificial Intelligence (AI) Market Size Worldwide from 2020 to 2030.* Statista. https://www.statista.com/forecasts/1474143/global-ai-market-size.

Trujillo, A., Fagni, T. & Cresci, S. (2025). The DSA Transparency Database: Auditing Self-Reported Moderation Actions by Social Media. *The 28th ACM Conference on Computer-Supported Cooperative Work and Social Computing, 9(2),* pp. 1-28. http://doi.org/10.1145/3711085.

UDHR – Universal Declaration of Human Rights., December 10, 1948. https://www.un.org/en/about-us/universal-declaration-of-human-rights.

United Nations. (n.d.). *Information and Communications Technologies.*UN - Security Council - Counter-Terrorism Committee (CTC). https://www.un.org/securitycouncil/ctc/content/information-and-communications-technologies#:~:text=In%20its%20resolution%202617%20.

United Nations Security Council. (1999). Resolution 1267. S/Res/1267 (1999). https://docs.un.org/en/S/RES/1267%20(1999).

United Nations Security Council. (2005). Resolution 1624. S/Res/1624 (2005). *United Nations Digital Library*. https://digitallibrary.un.org/record/556538?v=pdf.

United Nations Security Council. (2013). Resolution 2129. S/Res/2129 (2013). *United Nations Digital Library*. https://digitallibrary.un.org/record/762593?v=pdf.

United Nations Security Council. (2014). Resolution 2178. S/Res/2178 (2014). *United Nations Digital Library.* https://docs.un.org/en/S/RES/2178%20(2014).

United Nations Security Council. (2017). Resolution 2341. S/Res/2341 (2017). *United Nations Digital Library.* https://digitallibrary.un.org/record/858856?v=pdf.

United Nations Security Council. (2021). Resolution 2617. S/Res/2617(2021). *United Nations Digital Library.* https://docs.un.org/S/RES/2617(2021).

United Nations. (2020). *Secretary-General's Roadmap for Digital Cooperation.* United Nations. https://www.un.org/en/content/digital-cooperation-roadmap/.

US Commission on Civil Rights. (2024). The Civil Rights Implications of the Federal Use of Facial Recognition Technology. *USCCR.* https://www.usccr.gov/files/2024-09/civil-rights-implications-of-frt_0.pdf.

US Const. Amend. I. https://constitution.congress.gov/constitution/amendment-1/#:~:text=Congress%20shall%20make%20no%20law,for%20a%20redress%20of%20grievances.

US Const. Art. I, § 7. https://constitutioncenter.org/the-constitution.

US Department of State. (2016) *Country Reports on Terrorism 2015 - Italy.* Refworld. https://www.refworld.org/reference/annualreport/usdos/2016/en/110414.

Voigt, P. & Von Dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR) – A Practical Guide.* Springer. https://link.springer.com/content/pdf/10.1007/978-3-319-57959-7.pdf.

Volkov, M. (2025). The Root of Algocratic Illegitimacy. *Philosophy and Technology, 38(48)*, pp. 1-15. https://doi.org/10.1007/s13347-025-00879-4.

Webber, G. (2016). Proportionality and Absolute Rights. In V. Jackson & M. Tushnet (Eds.), *Proportionality: New Frontiers, New Challenges,* Cambridge University Press, 2016, LSE Legal Studies Working Paper No. 10/2016. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2776577.

Weimann, G., Pack, A.T., Sulciner, R., Scheinin, J., Rapaport, G. & Diaz, D. (2024). Generating Terror: The Risks of Generative AI Exploitation. *CTC Sentinel, 17(1),* pp. 17-24. https://ctc.westpoint.edu/wp-content/uploads/2024/01/CTC-SENTINEL-012024.pdf.

Wessel, R.A. (2011). Reconsidering the Relationship Between International and EU Law: Towards a Content-Based Approach? In Cannizzaro, E., Palchetti, P. & Wessel, R.A. (Eds.), *International Law as the Law of the European Union,* pp. 1-17. https://ris.utwente.nl/ws/portalfiles/portal/112667450/Wessel2011reconsidering.pdf.

White & Case. (2024). *AI Watch: Global Regulatory Tracker - Italy.* White & Case. https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-italy.

Works Constitution Act (Betriebsverfassungsgesetz – BetRVG). (2024). Federal Law Gazette 2024 I p. 248. Federal Ministry of Justice. https://www.gesetze-im-internet.de/englisch_betrvg/englisch_betrvg.html.

Yassin Abdullah Kadi and Al Barakaat International Foundation v Council of the European Union and Commission of the European Communities, C-402/05 P and C-415/05 P. (2008). Eur-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A62005CJ0402.

Zeran v. America Online, CA-96-1564-A, 129 F.2d 327 (1997). *Tech Law Journal.* http://www.techlawjournal.com/courts/zeran/Default.htm.