CYPRUSREVIEW

A Journal of Social Sciences

Spring 2026, Volume 38, Number 1

Published by the University of Nicosia

ISSN 1015-2881 (Print) | ISSN 2547-8974 (Online)

P.O. Box 24005
1700 Nicosia, Cyprus
T: 22-842301, 22-841500
E: cy_review@unic.ac.cy
www.cyprusreview.org

SUBSCRIPTION OFFICE:
The Cyprus Review
University of Nicosia
46 Makedonitissas Avenue
1700 Nicosia, Cyprus

Copyright: © 2026
University of Nicosia, Cyprus
ISSN 1015-2881 (Print), 2547-8974 (Online)

All rights reserved.
No restrictions on photo-copying.
Quotations from *The Cyprus Review*
are welcome, but acknowledgement
of the source must be given.

DISCLAIMER
The views expressed in the articles
and reviews published in this journal
are those of the authors and do
not necessarily represent the views
of the University of Nicosia,
the Editorial Board, or the Editors.

EDITORIAL TEAM

EDITORIAL BOARD

EDITORIAL TEAM

Editor-in-Chief:	Dr Christina Ioannou
Consulting Editor:	Prof. Achilles C. Emilianides
Managing Editor:	Dr Emilios A. Solomou
Publications Editor:	Andria Andreou
Publication Designer:	Thomas Costi
Linguistic Editors:	Lara Bastajian
	Elisavet Papageorgiou

EDITORIAL BOARD

Dr Constantinos Adamides	University of Nicosia
Dr Othon Anastasakis	University of Oxford
Prof. Panayiotis Angelides	University of Nicosia
Prof. George Christou	University of Warwick
Dr Odysseas Christou	University of Nicosia
Prof. Costas M. Constantinou	University of Cyprus
Prof. em. Van Coufoudakis	Indiana University–Purdue University Indianapolis
Prof. Alfred de Zayas	Geneva School of Diplomacy and International Relations
Prof. Thomas Diez	University of Tübingen
Prof. Dimitris Drikakis	University of Nicosia
Prof. Marios Evriviades	f. Panteion University
Prof. Hubert Faustmann	University of Nicosia
Prof. em. Kevin Featherstone	Hellenic Observatory Centre, London School of Economics
Prof. Vassilis Fouskas	University of East London
Dr Michael Given	University of Glasgow
Dr Christina Hajisoteriou	f. University of Nicosia
Prof. Evanthis Hatzivassiliou	National and Kapodistrian University of Athens
Prof. em. Robert Holland	University of London

Dr Sofia Iordanidou	Open University Cyprus
Prof. em. Andreas Kapardis	University of Cyprus
Dr Vassilis Kappis	University of Buckingham
Prof. Savvas Katsikides	University of Cyprus
Dr Erol Kaymak	Centre for Sustainable Peace and Democratic Development (SeeD)
Dr Michalis Kontos	University of Nicosia
Prof. em. Mary Koutselini	University of Cyprus
Prof. Petros Lois	University of Nicosia
Prof. Neophytos Loizides	University of Kent
Dr Diana Markides	Independent Researcher
Dr Yael Navarro	University of Cambridge
Prof. Phedon Nicolaides	University of Maastricht
Dr Petros Papapolyviou	University of Cyprus
Prof. em. Stelios Perrakis	Panteion University
Prof. Constantinos Phellas	University of Nicosia
Prof. Oliver Richmond	University of Manchester
Dr Soterios Rizas	Research Centre for the Study of Modern Greek History, Academy of Athens
Prof. Spyros Sakellaropoulos	Panteion University
Prof. Paul Sant Cassia	University of Malta
Dr Sertaç Sonan	Cyprus Academic Dialogue
Prof. Angelos Syrigos	Panteion University
Prof. Ioannis Tellidis	Kyung Hee University
Prof. em. Andreas Theophanous	University of Nicosia
Dr Alina Tryfonidou	University of Cyprus
Prof. Demetris Vrontis	University of Nicosia
Dr Craig Webster	Ball State University

CONTENTS

Contributors 10-15

Letter from the Editor-in-Chief 16-20

ARTICLES

ANTIGONI LOIZIDOU, NICHOLAS NICOLI, MARCOS KOMODROMOS
The Role of Political Branding in the 2016
Cyprus Parliamentary Elections 23-45

NICOLAS KYRIAKIDES AND GEORGE ISAIA
Watchdog Journalism:
A Tool to Strengthen Democracy? 47-103

SPECIAL SECTION: New Technologies in the World of Communication

IRINI STAMATOUDI
Copyright Protection for AI-Generated Works:
Is It Something We Want? 107-125

KONSTANTINOS DIMARELLIS
The Employment Status of Providers
in the Gig Economy 127-144

GIOVANNI CINÀ
No Content Without Context:
Regulating Hate Speech on Social Media 145-156

FILIPPO VIGLIONE
Rethinking Remedies for Hate Speech:
The Case for Apologies in Italian and Comparative Law 157-182

VAGIA POLYZOIDOU
Internet Service Providers:
Key Players in the New e-Evidence Paradigm? 183-198

DIMITRIOS I. KOUKIADIS
Artificial Intelligence, Fundamental Rights, and Digital
Constitutionalism from a Comparative Law Perspective:
The EU, US, and China Regulatory Models 199-236

OPINION PAPER

LOUKIS LOUCAIDES
Freedom of Expression and the Right to Respect Private Life 239-251

BOOK REVIEWS

ACHILLES C. EMILIANIDES
Free in Captivity. The Ayios Kassianos Women's March in 1989
[Ελεύθερες στην Αιχμαλωσία. Η Πορεία των Γυναικών
στον Άγιο Κασσιανό το 1989]
by Tasos Hadjianastasiou 255-257

ACHILLES C. EMILIANIDES
Accumulation of Treasures, Material Gleaned
from Cypriot Newspapers (1879-1962)
[Θησαύρισμα - Στοιχεία Σταχυολογηθέντα
από Κυπριακές Εφημερίδες (1879-1962)]
by Nicos Panayiotou 258-260

ACHILLES C. EMILIANIDES
The British Sovereign Base Areas in the Island of Cyprus
by Kyriacos Demetriades 261-263

ACHILLES C. EMILIANIDES
Horizontal Effect of Fundamental Rights
and its Reception in the Cypriot Legal Order
[Η Οριζόντια Επίδραση των Θεμελιωδών Δικαιωμάτων
και η Υποδοχή της στην Κυπριακή Έννομη Τάξη]
by Petros Constantinides 264-266

ACHILLES C. EMILIANIDES
Cyprus and the First World War
by Hubert Faustmann & Georgios Kazamias (eds) 267-272

PARSTUĞ COŞKUN
De-ethnicising the Ethnography of Cyprus: Political and Social Conflict
between Turkish Cypriots and Settlers from Turkey
by Yael Navaro-Yashin 273-276

ACHILLES C. EMILIANIDES
The Protection of the National Identity of the Member States
and the Principle of Supremacy in EU Law
[Η Προστασία της Εθνικής Ταυτότητας των Κρατών Μελών
& Η Αρχή της Υπεροχής στο Ευρωπαϊκό Δίκαιο]
by Theognosia Kouspi 276-277

CONTRIBUTORS

Giovanni Cinà is a Research Fellow in Comparative Law at the University of Padua, where he teaches 'Private Law and Economics' and 'Law, Information Technology, and Society'. In 2017, he obtained a PhD in Comparative Law from the same University, where in the following years he also held several fellowships. His main research interests include law and technology, fundamental rights, comparative legal systems, law and economics, and the capacity of vulnerable people. He is the author of several publications in peer-reviewed journals and of a book on the optimal level of private regulation in the multilevel governance systems of Italy and the United States.

Parstuğ Coşkun is an academic researcher currently pursuing an MA in Sociology at Ibn Haldun University, where he also completed his undergraduate studies. His scholarly focus encompasses the Middle East, the Balkans, and the Eastern Mediterranean, with specialized expertise in the evolution of nationalist ideologies and the intersection of religion and politics. Notably, his research investigates Turkish, Arab, and Cypriot nationalisms, alongside Ottoman history and Pan-Islamism. A recipient of the Türkiye Scholarships, he was also awarded a Golden Medal by the Philosophical Society of Turkey. He possesses multilingual proficiency in English, Turkish, Ottoman Turkish, and French.

Konstantinos Dimarellis was born in Thessaloniki. He holds a PhD degree in Labour Law of the Aristotle University of Thessaloniki. He graduated from the Law School of the Aristotle University of Thessaloniki and holds a Master of Law at the Department of Business and Labour Law of the Democritus University of Thrace and has also studied at Radboud Universiteit Nijmegen, The Netherlands. He teaches as an Assistant Professor at the University of Nicosia Labour Law, Social Security Law and Synthesis of Commercial Law. He has also given lectures as a guest Lecturer at the Universities of Athens (Greece), Stirling (Scotland), Limerick (Ireland), Verona (Italy), Padova (Italy) and Hasselt (Belgium). He works as a Lawyer at the Supreme Court of Greece, member of Thessaloniki Bar Association since 2008. He is the author of the monographs "Fundamental principles of Disciplinary Law" and "Lifting of corporate veil especially in Labour Law" and he is also co-author of the book "Labour Law in Greece" in co-operation with V. Douka and V. Koniaris. He has also published plenty of papers related to Labour Law issues. He is a mediator of the Organization of

Mediation and Arbitration (OMED). He has participated as a speaker at conferences and workshops, has published articles in newspapers and has given TV and radio interviews. He speaks English and German.

Achilles C. Emilianides is a practicing advocate, Professor of Law and Dean of the School of Law of the University of Nicosia. He holds a PhD in Law from the Aristotle University of Thessaloniki. He has taught and practiced in all fields of civil, public and commercial law. He is a Founding Member and President of the Cyprus Academy of Sciences, Letters, and Arts. He has also been elected as an Ordinary Member of the European Academy of Sciences and Arts and as a Titular Member of the International Academy of Comparative Law (IACL), and is a member of ALLEA (All European Academies) representing the Cyprus Academy. He is also the President of the Royal Commonwealth Society Cyprus Branch, and of the Center of Scientific Dialogue and Research.

George Isaia is the Executive Director of Oxygen for Democracy, focusing on participatory democracy, institutional resilience, and good governance. He also serves as Special Advisor to the Commissioner for the Citizen and as an independent governance advisor to Zenox Public Affairs.

Dimitrios Koukiadis is an Assistant Professor of Law at the University of Nicosia, and Adjunct Professor of Law at Goethe University Frankfurt/Main. He studied at Aristotle University of Thessaloniki, School of Law (LL.B., summa cum laude), at Northwestern University, School of Law/Kellogg School of Business (LL.M., Honorably/Dean's list), at Harvard University/Kennedy School of Government in the fields of public organization and leadership (EX.EDU), at the University of Chicago, Internet and Media Platforms, at the University of Edinburgh, Internet, Artificial Intelligence Planning and Philosophy of Sciences, and at Oxford University/Said Business School, Artificial Intelligence. He holds a Doctorate in Laws (LL.D., summa cum laude) from Goethe University Frankfurt/Main, School of Law. He is member of the Thessaloniki Bar Association and the Frankfurt Bar Association, as well as Certified Mediator/Arbitrator in Germany, USA and Greece specialized in Labor-, Commercial-, Energy-, Medical-, Family mediation and Labor/Commercial/Energy Arbitration. He is also an Arbitrator, member of the Greek Organisation of Mediation and Arbitration for labor/employment conflict resolution (O.ME.D.).

Nicolas Kyriakides is a lawyer, an academic and a lobbyist. He studied Law in Athens. He received postgraduate degrees from University College London and New

York University and a PhD from Oxford University. He was also a visiting researcher at Harvard. He is president of the Reform and Civil Procedure Committee of the Cyprus Bar Association and teaches at the University of Nicosia. His main research area is on procedural law and on private and public dispute resolution mechanisms as well as criminal law (civil, criminal and administrative procedural law). He is also a founding member of the non-profit organisation Oxygen for Democracy and its projects Nomoplatform and the Cyprus Forum.

Antigoni Loizidou is a Research and Development Officer and Project Manager at RESET in Limassol, Cyprus. Her academic foundation is anchored by a PhD in Digital Media and Political Sciences from the University of Nicosia, where her doctoral thesis explored the role of political branding in parliamentary elections. Her interdisciplinary scientific background is further supported by an MBA from the University of Cyprus, alongside Master's degrees in Media, Public Relations, and Social Networks (University of Nicosia) and Historical Religion (Aristotle University of Thessaloniki). Dr. Loizidou brings extensive experience in applied scientific research and public communication. From 2010 to 2017, she served as a Scientific Research collaborator for the National Committee of Children's Environmental Health, where she managed committee operations, organized academic seminars, and coordinated scientific press conferences. Currently, her research applies academic insights to European social empowerment and educational policy reform. She actively manages and contributes to high-impact Erasmus+ research projects, notably including SEMELI, which focuses on media literacy for socially excluded youth, and RECOMMIT, which develops cooperation partnerships in adult education.

Loukis Loucaides studied law in London, with post-graduate studies in Administrative Law in Paris. He worked as a lawyer in Nicosia between 1958-1964, and subsequently worked as Counsel (1964-1969), and Senior Counsel of the Republic (1969-1975). In 1975, he was appointed as Deputy Attorney-General of the Republic, a position he held until 1998. He represented the Republic of Cyprus in all recourses against Turkey before the European Commission of Human Rights, as well as the Human Rights Committee of the United Nations. Between 1989-1998, he served as a member of the European Commission of Human Rights. In 1998, he was elected as a Judge of the newly established European Court of Human Rights, where he served until 2008. Moreover, since 1987, he served as the President of the International Association for the Protection of Human Rights. The European Court of Human Rights published a book, with his separate opinions during his tenure as a Judge.

Nicholas Nicoli is Professor of Communication at the University of Nicosia and Associate Head of the Department of Communications. He is also the Coordinator of the Doctoral Programme in Media and Communication and a Member of the University of Nicosia Senate. Nicholas holds a PhD from City St. George's University of London's Department of Sociology. His research focuses on communication policy, digital platforms and the platformisation of the internet. He has conducted studies on the BBC, CYBC, Facebook, Tripadvisor and other similar entities. He is committed to the creation of a deliberative digital democratic landscape that prioritises the public interest. His current research investigates digital platform power and its consequences for news media.

Vagia Polyzoidou is an Associate Professor in Criminal Law, Criminal Procedure and Criminal Justice, working in University of Nicosia since 2015. She also participates as a lecturer in joint Master "Criminal Law and Addictions" (provided by AUTH and UNIC in collaboration with EOPAE" as well as in MA Criminology (School of Humanities, UNIC) and in brand new MA Regulation Technology (School of Law, UNIC). She completed her studies in Law School of Aristotle University of Thessaloniki (LLB, LLM in Criminal Law and Criminology/Degree: Excellent and a PhD from the same University/Degree: Excellent). Her thesis ("The Criminalization of Child Pornography") -for which she was awarded with a scholarship was published by Nomiki Vivliothiki. She is an attorney at law at the Greek Supreme Court and a member of several "criminal" associations and organizations as well as the co-director of the Laboratory for "Law and Literature" in UNIC. Her scientific research interests come from criminal law field, having dealt with topics related to Sexual Offenses, Juvenile Delinquency, Paternalism, the Principle of Proportionality, Financial Crime, Racism, Gender Based Violence, e-Evidence and Cybercrime.

Irini Stamatoudi is a professor and the director of the Athens Branch of the School of Law of the University of Nicosia (UNIC Athens). She is also an attorney-at-law. From 2007 to 2018, she served as the General Director of the Greek National Copyright Office (Greek Ministry of Culture and Sports). She has worked as a copyright expert on projects in Georgia, Abu Dhabi, Turkey, Egypt, Azerbaijan, and other locations. She has taught academic courses in the UK, Greece, Cyprus, Italy, France, and beyond. She co-authored the World Intellectual Property Organization (WIPO) Academy DL specialization course on copyright and has taught WIPO distance-learning courses since their inception in 2001. She has published extensively in Greece, Cyprus, the UK, and the US, and some of her writings are considered works of reference.

Filippo Viglione is a Full Professor of Comparative Private Law at the University of Padova, where he teaches Comparative Legal Systems, Anglo-American Law, and Privacy and AI Regulation. His research focuses on comparative methodology, family and succession law, contract law, new technology law. He is the Coordinator of the PhD Programme in International Law Private Law and Labour Law and former President of the Degree Programme in Law and Economics.

LETTER FROM THE EDITOR-IN-CHIEF

Dear Readers,

The spring 2026 issue of *The Cyprus Review* appears at a time of significant global uncertainty. The international system is facing a convergence of challenges, ranging from geopolitical instability and economic pressures to the profound transformations brought about by digital technologies. The continuing wars and armed conflicts that have emerged or intensified in recent years have underscored the fragility of international peace and the profound human, political, and economic consequences of military confrontation. Beyond their immediate humanitarian impact, such conflicts also reshape political discourse, influence democratic processes, and place additional strain on international cooperation and institutional trust.

At the same time, societies are navigating the far-reaching implications of technological change, particularly in the realms of digital communication, artificial intelligence, and data governance. These developments are transforming political communication, media ecosystems, labour relations, and legal frameworks. For countries such as Cyprus – situated at the crossroads of Europe and the Eastern Mediterranean and historically shaped by regional tensions – the interplay between geopolitical developments, democratic resilience, and technological transformation is especially salient. In this context, scholarly research plays a crucial role in fostering critical reflection, informed debate, and evidence-based policymaking.

The present issue of *The Cyprus Review* brings together a diverse set of contributions that engage with these themes from different perspectives. The first two articles focus on the functioning of democratic processes and the role of communication and media in shaping political life.

In their article, Antigoni Loizidou (Research and Education in Social Empowerment and Transformation – RESET), Nicholas Nicoli (University of Nicosia), and Marcos Komodromos (University of Nicosia) examine the role of political branding in the 2016 Cyprus Parliamentary Elections. The authors explore how political actors used branding strategies in an effort to reconnect with voters during a period marked by economic hardship, corruption scandals, and declining political trust. Drawing on qualitative research, including interviews with candidates and

communication professionals, the study analyses how branding techniques were employed to shape candidate identities and communicate political messages.

The second article, by Nicolas Kyriakides (University of Nicosia) and George Isaia (Oxygen for Democracy), turns to examine the evolution, impact, and global landscape of watchdog journalism in modern democracies. The authors trace the historical development of watchdog journalism and examine initiatives such as Nomoplatform in Cyprus, Vouliwatch in Greece, RKW in Poland, FactCheck.org in the United States, Lupa in Brazil, and the global "platform beat," highlighting the importance of independent scrutiny in strengthening democratic institutions. The authors argue that watchdog journalism remains indispensable in promoting transparency and civic engagement, while also emphasising the need for legal protections for journalists, stronger media literacy, and active civil society participation.

This issue also features an opinion paper by former Deputy Attorney-General of the Republic of Cyprus and former Judge of the European Court of Human Rights, Loukis Loukaides. In his contribution on freedom of expression and the right to respect private life, Loukaides engages with the topic of freedom of speech and the right of privacy on the basis of the provisions of the European Convention on Human Rights and the relevant case-law of the European Court of Human Rights.

This is followed by a special section devoted to *New Technologies in the World of Communication*, which brings together six contributions that explore the legal and societal implications of digital transformation. The articles in this section originate from a conference jointly organised by the School of Law of the University of Nicosia, the School of Law of the University of Padua, and *The Cyprus Review,* that took place on 31 May 2024 at the University of Nicosia.

In the first article of this section, Irini Stamatoudi (University of Nicosia) examines whether works generated through artificial intelligence should qualify for copyright protection and explores the extent to which works produced through artificial intelligence should qualify for copyright protection. By focusing on the role of human creativity in the concept of originality, the author argues that AI-generated works lacking meaningful human involvement do not meet the threshold required for copyright protection and therefore should not fall within its scope.

Giovanni Cina (University of Padua) next examines the regulation of hate speech on social media through a comparative and context-sensitive lens. The article critiques moderation approaches that rely predominantly on automated tools and

content-based assessments, arguing that they often fail to capture the social and cultural contexts that shape the meaning and harm of speech. The author proposes a model of context-aware governance that combines algorithmic tools with human expertise and locally grounded decision-making.

A complementary perspective is offered by Filippo Viglione (University of Padua), who explores private law remedies against hate speech. The article focuses on the potential of civil law remedies – particularly compelled apologies – to address the symbolic and relational harm caused by hate speech. Through an examination of Italian legal practice and anti-discrimination law, the author suggests that apologies may serve as an important instrument of corrective justice by acknowledging harm and reaffirming social norms.

In his contribution, Konstantinos Dimarellis (University of Nicosia) examines the employment status of the providers in the Gig Economy and addresses the growing significance of digital labour platforms. As the gig economy expands and digital platforms increasingly mediate the provision of services, questions regarding the legal classification and protection of workers become central. The author explores the criteria used to distinguish between employees and independent contractors and discusses the regulatory challenges associated with ensuring fair working conditions in platform-mediated labour.

Vagia Polyzoidou (University of Nicosia) next analyses the EU's e-Evidence package – a Directive and Regulation on European Production and Preservation Orders for electronic evidence. She examines its impact on cross-border criminal investigations and prosecutions, highlighting how it reshapes EU criminal procedure by expanding the role of private actors, particularly internet service providers, in handling evidence. The author considers how the measures raise concerns about safeguarding fundamental human rights.

The section concludes with Dimitris Koukiadis' (University of Nicosia) article on Artificial Intelligence, fundamental rights, and digital constitutionalism. The article explores how emerging technologies challenge traditional legal frameworks and discusses the concept of digital constitutionalism as a developing normative approach to regulating digital environments. Through a comparative analysis of regulatory models in the EU, the US and China, the author highlights the need for adaptable legal frameworks that balance technological innovation with the protection of fundamental rights.

As in every issue, *The Cyprus Review* also includes a dedicated book review section, featuring seven reviews of recent scholarly publications. The section continues the Journal's commitment to engaging critically with contemporary Cyprological academic literature and to fostering broader scholarly discussion.

The issue concludes with a Call for Papers for a forthcoming Special Section entitled *One Hundred Years since the Founding of the Communist Party of Cyprus, 1926–2026*. Marking the centenary of the founding of the Communist Party of Cyprus, the special section will invite contributions examining the historical, political, and social significance of this milestone within the broader context of Cypriot political history.

Together, the contributions herein reflect the wide range of topics that continue to shape contemporary academic inquiry—from political communication and democratic accountability to the complex legal challenges posed by digital technologies. By bringing together perspectives from different disciplines and methodological approaches, we aim to always contribute to ongoing discussions about governance, rights, and the evolving relationship between technology, society, and democracy.

Christina Ioannou
Editor-in-Chief

ARTICLES

The Role of Political Branding in the 2016 Cyprus Parliamentary Elections

Antigoni Loizidou,[1] **Nicholas Nicoli,**[2] **Marcos Komodromos**[3]

Abstract

This paper explores the role of political branding in the 2016 Cyprus parliamentary elections, focusing on how political actors used branding strategies to engage the electorate. In a socio-political landscape marked by corruption, economic hardship, and widespread voter apathy, political branding was strategically employed to reconnect with voters and influence electoral outcomes. Drawing on qualitative research, including semi-structured interviews with candidates and communication professionals, the study analyses how political branding strategies were implemented and assesses their effectiveness. The findings indicate that while political branding helped shape candidate identities and facilitated communication with fragmented voter groups, its overall success was limited by incomplete execution and inconsistent messaging. This study contributes to the understanding of branding in politics, particularly in small, fragmented electoral systems like Cyprus, and offers insights for future campaigns.

Keywords: political branding; political marketing; Cyprus parliamentary elections; voter apathy; social media

Introduction

Political branding,[4] a strategic communication tool rooted in communication and marketing principles, has increasingly become a focal point in modern electoral cam-

[1] [PhD. In Political Communication, Research and Education o Social Empowerment and Transformation RESET]

[2] Professor and Associate Head of the Department of Communications, University of Nicosia

[3] Professor, Department of Communications, University of Nicosia

[4] We use 'political branding' to denote the set of associations that voters hold about a party or candidate, co-produced by organisational identity, communications, and voter interpretation; see Margaret Scammell, *Consumer Democracy: The Marketing of Politics* (Cambridge, UK: Cambridge University Press, 2014); Jennifer Lees-Marshment & al., *Political Marketing: Principles and Applications* (3rd edn., Abingdon, UK: Routledge, 2019); Christopher Pich, Gudrun Armannsdottir & Deborah Dean, 'Exploring the Process of Creating and Managing Personal Political Brand Identities in Non-Party Environments: The Case of the Bailiwick of Guernsey' *Journal of Political Marketing*, (2020) 19(4) 414–434.

paigns. In essence, political branding involves crafting a distinct image for candidates or political parties and is designed to resonate with specific voter segments. As politics becomes more mediated and voters more fragmented, the ability to create and maintain a compelling political brand has gained new significance. This shift reflects broader trends in political communication, where the dynamics of trust, emotional engagement,[5] and identity play pivotal roles in shaping voter behaviour.[6] Branding, as used in elections, extends beyond logos or slogans; it is about fostering emotional connections and creating narratives that reflect the values, aspirations, and frustrations of the electorate.[7]

The 2016 Cyprus parliamentary elections offer a unique opportunity to examine the power and limitations of political branding within a small, politically charged society. The elections were held against a backdrop of severe voter dissatisfaction. Years of economic instability, rampant corruption scandals, and a political system heavily influenced by long-standing party loyalty had eroded public trust. As populist movements swept through Europe, Cyprus too experienced a shift in its political landscape. Voter disillusionment with traditional political elites and scepticism toward their promises became central themes of the elections, reflecting broader frustrations with the political system.[8]

In the 2016 Cyprus parliamentary elections, approximately 66.7% of registered voters cast their ballots, resulting in a record-high abstention rate of 33.3%.[9] This significant level of voter abstention highlighted a growing disengagement and loss of faith in the political process among the electorate,[10] and posed a critical challenge for

[5] 'Emotional engagement' here means affective responses (e.g., enthusiasm, hope, anger, cynicism) that shape attention, information processing, recall, and persuasion in political communication; see George E. Marcus, W. Russell Neuman & Michael MacKuen, *Affective Intelligence and Political Judgment* (Chicago: University of Chicago Press, 2000); Ted Brader, *Campaigning for Hearts and Minds: How Emotional Appeals in Political Ads Work* (Chicago: University of Chicago Press, 2006); Robin L. Nabi, 'A Cognitive–Functional Model for the Effects of Discrete Negative Emotions on Information Processing, Attitude Change, and Recall' (1999) 9(3) *Communication Theory* 292–320.

[6] Geoffrey Baym, 'Beyond Bias: Conservative Media, Documentary Form, and the Politics of Hysteria' (2022) 137(4) *Political Science Quarterly* 787–788.

[7] Lees-Marshment & al. (no 4) 22.

[8] Christos Kassimeris, *The Politics of Football: Political Crises and Political Branding* (Basingstoke, UK: Palgrave Macmillan, 2009) 134–135.

[9] Republic of Cyprus, Ministry of Interior, 'Elections Results (Official Portal)' (n.d.) available at https://www.gov.cy/moi/en/ministry/departments/civil-registry-section/central-electoral-service/4350-2/

[10] IPU Parline, 'Vouli Antiprosopon (May 2016): Election Results—Cyprus' (2016) available at https://data.ipu.org/parliament/CY/CY-LC01/election/CY-LC01-E20160522/.

political actors. Political branding emerged as a potential solution, offering a means to reconnect with voters through targeted messaging, digital platforms, and strategic positioning.[11]

This study investigates the role of political branding in these elections, exploring how political candidates and their communication teams employed branding strategies to navigate the complexities of Cypriot politics. At the core of this research are several critical questions: How did political actors target and segment the Cypriot electorate in 2016? What were the dominant thoughts and attitudes among candidates and communication consultants regarding branding's ability to influence voter behaviour? Moreover, what role did digital media and communication play in shaping these branding strategies, and to what extent did political branding ultimately affect the election outcomes?[12]

The research questions are as follows:

RQ1. How did political parties and candidates articulate and communicate their brand identities in the 2016 Cyprus parliamentary elections?

RQ2. In what ways did perceived authenticity and consistency shape voter evaluations of parties and candidates?

RQ3. How did economic hardship, perceptions of corruption, and political apathy condition branding strategies and voter reception?

RQ4. What roles did digital and traditional media play in constructing, amplifying, or undermining these brand identities?

In seeking answers to these questions, this study draws on a qualitative research design, including semi-structured interviews with 14 parliamentary candidates and 5 communication professionals. Through these interviews, the study delves into the various tools and techniques used in branding, focusing on how candidates attempted to appeal to specific voter groups and the socio-political themes that shaped their messaging. From analysing how politicians crafted their public images to exploring the role of social media as a branding tool, this paper offers a comprehensive look at how branding shaped the 2016 parliamentary elections in Cyprus.[13]

---

[11] Rachel Gibson & Andrea Römmele, *Political Communication in the Online World* (Abingdon, UK: Routledge, 2015) 41–45.

[12] W. Lance Bennett & Robert M. Entman, *Mediated Politics: Communication in the Future of Democracy* (Cambridge, UK: Cambridge University Press, 2001) 223.

[13] Andrew Chadwick, *The Hybrid Media System: Politics and Power* (Oxford, UK: Oxford University Press, 2013) 90.

Political branding in Cyprus, however, must be viewed within the broader context of global trends in political communication. In an era where politics is increasingly mediated by digital technology, the lines between commercial and political branding have blurred.[14] Political actors now operate in a highly competitive marketplace of ideas, where voters are both consumers of political messages and participants in an evolving digital public sphere. The rise of social media has given politicians new platforms to reach voters directly, bypassing traditional media channels. However, this direct engagement has also introduced new challenges, particularly around message consistency and authenticity[15] and the risks of misinformation.[16]

This paper aims to contribute to the understanding of political branding in such a dynamic environment. It not only sheds light on the specific case of Cyprus but also offers broader insights into how political actors in small, fragmented electoral systems can leverage branding strategies to navigate voter apathy and scepticism. In doing so, it also highlights the limitations of these strategies, particularly when they are inconsistently executed or when they fail to address the underlying causes of voter disengagement.[17]

Ultimately, the 2016 Cyprus parliamentary elections underscore both the potential and the limitations of political branding as a tool for electoral success. While branding offers candidates a way to differentiate themselves and build emotional connections with voters, it is not a panacea. In the absence of meaningful political reform and genuine voter engagement, branding alone is unlikely to resolve the deep-seated issues of voter apathy and political disillusionment that characterised the 2016 elections.[18] Abstention rose from 9% (2006) to 21% (2011) and 33.3% (2016), indicating a marked loss of trust (see Figure 1).

[14] Gianpietro Mazzoleni, *The Media and Neo-Populism: A Contemporary Comparative Analysis* (Westport, Praeger 2003) 12.

[15] 'Authenticity' refers to the perceived congruence between brand claims and observed behaviour, including consistency across channels and over time (see Pich, Armannsdottir & Dean (no 4)).

[16] Matthew A Baum & Philip BK Potter, *War and Democratic Constraint: How the Public Influences Foreign Policy* (Princeton, NJ: Princeton University Press, 2015) 210.

[17] Pippa Norris, *A Virtuous Circle: Political Communications in Post-Industrial Societies* (Cambridge, UK: Cambridge University Press, 2000) 34.

[18] Rachel K. Gibson, T*he Internet and Democracy in Global Perspective: Voters, Candidates, Parties, and Social Movements (*Cham, Switzerland: Springer, 2016) 112–114.

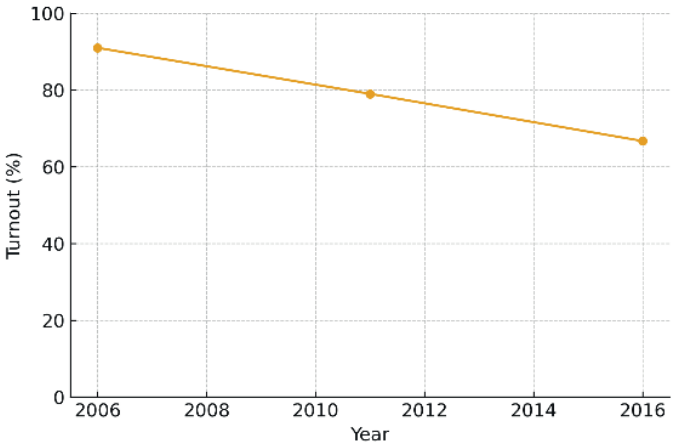

Figure 1. Parliamentary election turnout in Cyprus, 2006–2016.
Source: Official results / IPU Parline.

Literature Review

Concepts and Definitions

Political branding connects organisational identity, leader image, issue ownership, and channel strategy to the associations voters hold about a party or candidate. We adopt the working definitions of 'political branding', 'authenticity', and 'emotional engagement' as set out in the Introduction. Building on those definitions, this review positions the study within the political marketing and branding literature and sets up the small-State, fragmented-system context examined in Cyprus.

Political branding has become an essential aspect of electoral campaigns worldwide, merging principles of marketing with political communication to create compelling candidate identities. The process involves the strategic crafting of a politician's image and message to resonate with specific voter demographics, much like how corporations brand products to appeal to target consumers.[19] This approach recognises that voters, inundated with information and choices, often rely on emotional and psychological cues rather than solely on policy positions when making electoral decisions.[20]

[19] Margaret Scammell, 'Political Brands and Consumer Citizens: The Rebranding of Tony Blair' (2007) 611(1) *The Annals of the American Academy of Political and Social Science* 176–192.

[20] Gareth Smith & Alan French, 'The Political Brand: A Consumer Perspective' (2009) 9(2) *Marketing Theory* 209–226.

The significance of political branding has grown alongside advancements in communication technology and the fragmentation of media channels. In an era where traditional party loyalties are declining, and voters are more sceptical of political institutions, branding provides a means for candidates to establish trust and differentiate themselves.[21] Effective political branding goes beyond slogans and logos; it encompasses the values, narratives, and personal attributes that a candidate embodies and communicates consistently across all platforms.[22]

The Evolution of Political Branding

The integration of marketing techniques into politics is not a new phenomenon, but it has evolved significantly over the past few decades. Early instances of political marketing can be traced back to the use of advertising and public relations strategies in mid-20th-century campaigns.[23] However, the advent of television and later the internet revolutionised political communication, making image and media management central to campaign strategies.[24]

In recent years, there has been a shift towards more sophisticated branding techniques, with campaigns employing data analytics, behavioural science, and micro-targeting voter segments.[25] This personalisation aims to create stronger emotional connections with voters, addressing their unique concerns and aspirations.[26] The rise of social media platforms has further amplified this trend, allowing candidates to engage directly with voters and cultivate their brands in real time.[27]

Political branding has played a pivotal role in elections around the world. In the US, Barack Obama's 2008 presidential campaign is often cited as a landmark in mod-

[21] Stephan C.M. Henneberg & Nicholas J O'Shaughnessy, 'Political Relationship Marketing: Some Micro/Macro Thoughts' (2007) 23(9–10) *Journal of Marketing Management* 7–29.

[22] Christopher Pich & Klement Podnar, 'Political Brand Identity: An Examination of the Complexities of Conservative Brand and Internal Market Engagement During the 2010 UK General Election Campaign' (2018) 24(1) *Journal of Marketing Communications* 100–117.

[23] Lees-Marshment & al. (no 4).

[24] Scammell, Margaret, 'The Internet and Civic Engagement: The Age of the Citizen-Consumer' (2000) 17(4) *Political Communication* 351–355.

[25] Daniel Kreiss, *Prototype Politics: Technology-Intensive Campaigning and the Data of Democracy* (Oxford, UK: Oxford University Press, 2016).

[26] Daisuke Harada, 'Micro-Targeting and the Future of Political Communication' (2017) 14(2) *Journal of Information Technology & Politics* 98–111.

[27] Sanne Kruikemeier & al., 'Political Campaigning on Social Media: Politicians, Audiences, and the Mediation of Political Communication on Facebook and Twitter' (2013) 30(1) *Political Communication* 10–28.

ern political branding.[28] The campaign's emphasis on 'hope' and 'change', coupled with a strong visual identity and adept use of social media, created a powerful brand that resonated with a broad electorate.[29] Similarly, the 2016 and 2020 US elections saw candidates leveraging branding to galvanise support, with Donald Trump's 'Make America Great Again' slogan becoming a central element of his political identity.[30]

In Europe, political branding has also become integral to campaign strategies. Emmanuel Macron's 2017 presidential campaign in France, for example, successfully established a centrist brand that positioned him as a fresh alternative to traditional parties.[31] His movement, 'En Marche', was branded to symbolise progress and forward momentum, appealing to voters disillusioned with the status quo.[32]

Comparative Touchpoints

Two widely cited benchmarks illustrate how branding mechanisms operate in different contexts. In the US, Barack Obama's 2008 campaign built a coherent brand narrative around 'hope' and 'change', supported by a distinctive visual identity and an early, strategic use of digital mobilisation. In Europe, Emmanuel Macron's 2017 campaign positioned 'En Marche!' as a centrist, future-oriented alternative to traditional parties, using brand coherence to signal novelty and competence.

In smaller or fragmented party systems, branding typically relies less on grand ideological repositioning and more on credibility, consistency, and local responsiveness. For example, in Greece (2010s), party competition under proportional representation intensified the emphasis on leader credibility and problem-solving cues. In Malta (2010s), a small electorate and high political visibility increased the reputational costs of inconsistency, making cross-channel coherence particularly important. In Ireland (2010s), PR-STV reinforced candidate-centred competition, where

[28] Jennifer Stromer-Galley, *Presidential Campaigning in the Internet Age* (2nd edn, Oxford, UK: Oxford University Press, 2019).

[29] Gunn Enli, 'Twitter as Arena for the Authentic Outsider: Exploring the Social Media Campaigns of Trump and Clinton in the 2016 US Presidential Election' (2017) 32(1) *European Journal of Communication* 50–61.

[30] Brian L Ott, 'The Age of Twitter: Donald J Trump and the Politics of Debasement' (2017) 34(1) *Critical Studies in Media Communication* 59–68.

[31] Christian Lundberg, 'Political Branding in the Modern Age: Effective Strategies, Tools, and Techniques' (2019) 18(2) *Journal of Political Marketing* 105–110.

[32] Benjamin Moffitt, *The Global Rise of Populism: Performance, Political Style, and Representation* (Stanford, CT: Stanford University Press, 2016).

local service reputations and personal reliability were key brand assets, and digital communication mainly amplified on-the-ground standing.

These brief touchpoints are used here as contextual comparators to frame Cyprus as a small, fragmented system where branding is similarly constrained by structure and trust dynamics.

Digital Media, Platforms, and Disinformation

Digital media has revolutionised political branding by providing new channels for direct voter engagement. Social media platforms have become essential tools for candidates to build and disseminate their brands.[33] These platforms allow for immediate communication, feedback loops, and the ability to reach a wide audience at a relatively low cost.[34]

In the context of the 2016 Cyprus elections, social media played a crucial role in candidates' branding efforts. Platforms like Facebook and Twitter were used to share policy positions, respond to voter inquiries, and present the more personal side of the candidates.[35] Younger voters, who are typically more active on social media and less engaged with traditional media, became a key target demographic.[36]

However, the use of digital media also presented challenges, many of which are not unique to digital platforms but were amplified by their rapid and far-reaching nature. While the need for consistent and authentic messaging has always been crucial, digital media's ability to quickly spread discrepancies or missteps made candidates more vulnerable to scrutiny.[37] Additionally, the prevalence of disinformation and the echo chamber effect posed risks to the effectiveness of branding strategies on social media.[38] Research by Iosifidis and Nicoli highlights the challenges posed by digital platforms in political branding, particularly with the spread of disinformation and

[33] Jason Gainous & Kevin M Wagner, *Tweeting to Power: The Social Media Revolution in American Politics* (Oxford, UK: Oxford University Press, 2014).

[34] Andreas M. Kaplan & Michael Haenlein, 'Users of the World, Unite! The Challenges and Opportunities of Social Media' (2010) 53(1) *Business Horizons* 59–68.

[35] Athina Karatzogianni & al., 'Social Media, Political Branding, and the 2016 Cypriot Parliamentary Elections' (2018) 10(3*) International Journal of Electronic Governance* 223–238.

[36] Yannis Theocharis & Will Lowe, 'Does Facebook Increase Political Participation? Evidence from a Field Experiment' (2016) 19(10) *Information, Communication & Society* 1465–1486.

[37] Cristian Vaccari & al., 'Social Media and Political Communication: A Survey of Uses and Effects' (2015) 9 *International Journal of Communication* 2046–2066.

[38] Hunt Allcott & Matthew Gentzkow, 'Social Media and Fake News in the 2016 Election' (2017) 31(2) J*ournal of Economic Perspectives* 211–236.

the erosion of trust in political figures.[39] [40] This dynamic was observed in the Cyprus elections, where candidates had to navigate a digital landscape fraught with misleading information that could undermine their carefully crafted brands.[41]

Small-State and Fragmented Party Systems

While political branding offers numerous advantages, it is not without its limitations. One significant challenge is ensuring that branding efforts are authentic and reflect genuine commitment rather than superficial marketing ploys.[42] Voters are increasingly adept at recognising inauthenticity, which can lead to cynicism and further disengagement.[43] Moreover, branding alone cannot address underlying structural issues that contribute to voter apathy and distrust. In Cyprus, factors such as corruption, economic instability, and unresolved political disputes require substantive policy solutions and institutional reforms. Without addressing these core issues, branding efforts may have limited impact on long-term voter engagement.

Additionally, Nicoli and Komodromos argue that public relations play a crucial role in reinforcing the credibility of political brands.[44] In the case of Cyprus, PR strategies were instrumental in managing crises and communicating a candidate's core values across various platforms. Consistency in messaging and transparency were key factors in maintaining voter trust, particularly in a context where scepticism toward political elites was high.

The literature underscores the growing importance of political branding in modern electoral politics. From global examples to the specific case of Cyprus, it is evident that branding strategies have become integral to how candidates communicate with voters. The rise of digital media has further transformed the landscape, offering new opportunities and challenges for political branding.

---

[39] Petros Iosifidis & Nicholas Nicoli, 'The Battle to End Fake News: A Qualitative Content Analysis of Facebook Announcements on How It Combats Disinformation' (2020) 82(1) *International Communication Gazette* 60–81.

[40] Petros Iosifidis & Nicholas Nicoli, *Digital Democracy, Social Media and Disinformation* (London, UK: Routledge, 2020).

[41] Nicholas Nicoli & Marcos Komodromos, *Principles of Public Relations* (Nicosia: University of Nicosia Press, 2013).

[42] Alex Marland & Thierry Giasson, 'From Brokerage to Boutique Politics: Political Marketing and the Changing Nature of Party Politics in Canada' (2013) 46(4) *Canadian Journal of Political Science* 961–972.

[43] Stephen Coleman, *How Voters Feel* (Cambridge, UK: Cambridge University Press, 2013).

[44] Nicoli & Komodromos (no 41).

In small electorates and fragmented party systems, branding effects are often bounded by structure: turnout dynamics, party-system legacies, and institutional trust condition both strategy and reception. Candidate-centred appeals and community-level responsiveness tend to matter more where personal contact and reputational cues travel quickly, yet brand strength remains constrained when economic strain and distrust are high. In Cyprus, these pressures were salient in 2016 (see Introduction for country statistics and sources), shaping both how brands were constructed and how they were received.

This review highlights the need for authenticity, consistency, and a deep understanding of voter concerns in crafting effective political brands. As the subsequent sections of this paper will explore, the 2016 Cyprus parliamentary elections provide valuable insights into the practical application of these principles and the outcomes of such strategies in a complex political environment.

Methodology

This study employs a qualitative research approach to explore the role of political branding in the 2016 Cyprus parliamentary elections. A qualitative methodology was selected to capture in-depth insights into the strategies and perceptions of political actors and communication professionals. Given the complexity of political branding, involving subjective experiences, symbolic meanings, and emotional engagement, qualitative methods are well-suited for this type of research.[45]

Research Design

The research design centred on semi-structured interviews with key stakeholders in the 2016 elections. Semi-structured interviews allow for flexibility in exploring various aspects of political branding while maintaining a consistent framework for comparison across participants. The interview guide included questions on branding strategies, voter engagement, messaging techniques, and the role of digital media.[46] The open-ended nature of the questions enabled participants to offer rich, detailed insights into their branding efforts, including the challenges they faced and the strategies they perceived as most effective.

[45] Scammell (no 19).

[46] Gibson & Römmele (no 11).

Rationale for Qualitative Approach

A qualitative approach was deemed appropriate for several reasons. First, political branding is inherently subjective, involving the creation and management of symbolic meanings, emotional engagement, and voter perceptions. As such, capturing the nuances of branding strategies requires an approach that allows for a deep exploration of these meanings and their impact. Qualitative research enables researchers to gather rich, detailed data on how political actors view their own branding strategies and how these strategies align with the expectations of the electorate.[47] Furthermore, qualitative methods are well-suited to understanding the personal and social contexts that influence both the development and reception of political brands.

Sampling

A purposive sampling strategy was employed to ensure the inclusion of a diverse range of participants. The sample consisted of 14 parliamentary candidates from a range of political parties, representing a cross-section of Cyprus' fragmented political landscape. This ensured that the sample captured the experiences of both well-established parties and smaller, emerging political movements.[48] Additionally, five communication professionals who worked closely with these candidates were interviewed to provide insights into the strategic thinking behind branding efforts.

The inclusion of communication professionals is particularly important in understanding the operationalisation of branding strategies. These individuals play a crucial role in shaping a candidate's image, managing media relations, and crafting messages that align with the candidate's brand identity.[49] Their perspectives are valuable for understanding how political branding is translated into actionable communication tactics.

Participants and Coding

To preserve anonymity while enabling analytic transparency, interviewees are distinguished by non-identifying codes: PC1–PC14 for party candidates and CP1–CP5 for communication professionals. These codes are used only where a specific interview-derived statement, paraphrase, or quotation is reported, and are placed immediately after the relevant material.

---

[47] John W. Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (Sage, 2014) 22–23.

[48] Smith & French (no 20).

[49] Nicoli & Komodromos (no 41).

Data Collection

Data collection took place over a three-month period through a combination of in-person and virtual interviews. Each interview lasted between 45 and 90 minutes, allowing for an in-depth exploration of the participants' views and experiences. Semi-structured interviews were chosen for their flexibility, as they allow interviewees to express their thoughts freely while adhering to a structured set of guiding questions.[50] The interview questions focused on participants' branding strategies, how they differentiated themselves from opponents, and how they adapted their messaging to address key voter concerns. Specific questions were also included to explore how digital media platforms were utilised in political branding efforts.

In addition to interviews, campaign materials such as flyers, social media posts, speeches, and press releases were analysed. This document analysis provided a supplementary data source, allowing for a comparison between the candidates' self-reported branding strategies and the public-facing messages they produced during the campaign.[51] By analysing interviews and materials, the study was able to identify consistencies and discrepancies between branding strategies and their execution.

Ethical Considerations

Given the sensitive nature of political campaigning, this study adhered to strict ethical guidelines to ensure participant confidentiality and anonymity. All participants provided informed consent prior to their involvement in the research, with an understanding of how the data would be used and the steps taken to protect their identities.[52] For example, interview recordings were encrypted, and pseudonyms were used in all written reports to maintain anonymity.

Participants were also given the option to withdraw from the study at any point without providing a reason.[53] This ensured that the research process remained transparent and voluntary, which is particularly important when conducting interviews with political actors who may be concerned about reputational risks.

---

[50] Virginia Braun & Victoria Clarke, 'Using Thematic Analysis in Psychology' (2006) 3(2) *Qualitative Research in Psychology* 77–101.

[51] Sarah J. Tracy, *Qualitative Research Methods: Collecting Evidence, Crafting Analysis, Communicating Impact* (Wiley-Blackwell, 2019).

[52] Udo Kuckartz, *Qualitative Text Analysis: A Guide to Methods, Practice, and Using Software* (London, UK: Sage, 2014).

[53] Braun & Clarke (no 50).

Data Analysis

Thematic analysis was employed to identify recurring patterns and key themes within the data. This method was chosen for its flexibility and ability to provide a clear framework for analysing qualitative data. Thematic analysis involves coding the data and organising it into broader themes that are relevant to the research questions.[54]

The data analysis process unfolded in several stages. First, the interview transcripts and campaign materials were thoroughly reviewed to identify initial codes. These codes were based on recurring topics and themes, such as personal branding, voter engagement strategies, the role of digital media, and the challenges of engaging the electorate.[55] In the second stage, the codes were refined and grouped into broader categories that reflected the main objectives of the study. For instance, personal branding strategies were broken down into sub-themes such as authenticity, transparency, and the use of storytelling to build a connection with voters.[56] Finally, the themes were interpreted in the context of the broader literature on political branding to identify areas of convergence and divergence between theory and practice.[57]

Reliability and Validity

To enhance the reliability and validity of the findings, multiple strategies were employed. First, the use of document analysis alongside interviews allowed for the triangulation of data, ensuring that conclusions were not solely based on participant accounts but also on publicly available materials.[58] Second, member checking was conducted, wherein participants were given the opportunity to review and comment on the findings before they were finalised. This process helped to ensure that the interpretations accurately reflected the participants' views.[59]

Additionally, a reflexive approach was adopted throughout the research process. The researcher maintained a reflexive journal to document personal reflections, biases, and potential influences on the data analysis. This helped to mitigate the risk of researcher bias and ensured that the data was analysed as objectively as possible.[60]

---

[54] Lees-Marshment & al. (no 4).

[55] Braun & Clarke (no 50).

[56] Tracy (no 51).

[57] Scammell (no 19).

[58] Kuckartz (no 52).

[59] Margaret M. MacQueen, *Team-based Qualitative Research*, (Altamira Press, 2005).

[60] Linda Finlay, 'Negotiating the Swamp: The Opportunity and Challenge of Reflexivity in Research Practice' (2002) 2(2) *Qualitative Research* 209–230.

Findings

The thematic analysis of the interview data revealed several key themes related to the use of political branding in the 2016 Cyprus parliamentary elections. These themes reflect the challenges of personal branding, the importance of digital media, and the difficulties of engaging the Cypriot electorate.

Personal Branding: Crafting Candidate Identities (Addresses RQ1 and RQ2)

In the lead-up to the 2016 parliamentary elections in Cyprus, political candidates turned to branding strategies to reconnect with voters. Seeking to differentiate themselves from the traditional political landscape, candidates crafted personal brands that highlighted their commitment to integrity, transparency, and a genuine concern for the issues faced by ordinary citizens. This approach marked a departure from conventional campaigning, as candidates increasingly relied on innovative communication tactics to resonate with an electorate that was growing more sceptical of traditional political parties. The most salient themes were authenticity, local problem-solving, social media engagement, leader image, and issue ownership (see Figure 2).

Theme → / RQ ↓	RQ1 Brand articulation	RQ2 Authenticity/ constistency	RQ3 Context conditioning	RQ4 Media roles
Authenticity	✓	✓		
Local problem-solving	✓		✓	
Social media engagement				✓
Leader image	✓	✓		
Issue ownership	✓			

Figure 2. Mapping of core themes to the study's research questions (qualitative coverage; no frequencies).

A significant theme that emerged from the interviews was the importance of personal branding. Candidates recognised that voters had grown sceptical of traditional political parties and were more interested in the individual qualities of political figures than in party platforms.[61] This shift required candidates to differentiate them-

[61] Smith & French (no 20).

selves from their competitors not just by their policies but by their personalities and values (PC3).

Many candidates described the process of creating a distinct identity that would resonate with voters on an emotional level. One candidate noted, 'I had to be authentic, transparent, and relatable because voters were tired of hearing the same scripted promises' (PC1). This emphasis on authenticity recurred across interviews (PC2–PC4). Candidates who could effectively communicate their personal story and values were more successful in creating a brand that appealed to disillusioned voters (PC3).

Personal branding also extended beyond merely showcasing personal achievements. Candidates sought to humanise themselves by sharing personal struggles and demonstrating empathy for the challenges faced by voters (PC1).[62] For example, several candidates mentioned that they emphasised their background as 'ordinary citizens' (PC4) who understood the hardships brought on by economic instability. This narrative helped candidates build trust and appear more relatable to an electorate that had grown increasingly disillusioned with political elites.

However, the emphasis on personal branding also posed challenges. Some candidates reported difficulty balancing their individual identity with the collective identity of their party. In many cases, candidates had to navigate the tension between promoting their personal brand while maintaining party unity. One communication professional explained, 'It was a delicate balance because you want to stand out, but at the same time, you don't want to alienate your party's base' (CP1).

The Role of Digital Media in Shaping Political Brands (Addresses RQ4)

The second key theme identified in the data was the crucial role of digital media in political branding. Social media platforms such as Facebook, Twitter, and Instagram provided candidates with the tools to engage directly with voters, bypassing traditional media gatekeepers and presenting a more personal, unfiltered version of themselves (PC8).[63] This shift toward digital communication was particularly important for reaching younger voters, who were less likely to engage with traditional campaign materials (PC3).

Many candidates and communication professionals highlighted the strategic use of social media to enhance their brand's visibility and relatability. One candidate commented, 'Social media allowed me to show who I really was, beyond the speeches

[62] Lees-Marshment & al. (no 4) 85–87.

[63] Kaplan & Haenlein (no 34).

and policy talks. I could share my day-to-day life, my thoughts on issues as they happened, and respond to people in real-time' (CP2). This sustained engagement helped candidates build a sense of immediacy and accessibility, which was critical in an environment where political elites were seen as distant and disconnected.

The use of digital media also allowed candidates to adapt their branding strategies on the fly. As one communication professional described: 'We were able to monitor voter responses to posts and adjust the messaging accordingly. If something didn't resonate, we could pivot quickly' (CP1).[64] However, the reliance on social media also introduced challenges. Candidates had to navigate the complexities of maintaining consistency across platforms while dealing with negative comments, misinformation, and the echo chamber effect of social media algorithms.[65] Some candidates expressed frustration with the rapid spread of false information, which often required them to respond quickly to prevent damage to their brand (PC2–PC4).[66]

Despite these challenges, the strategic use of digital media was seen as indispensable in modern political branding. Candidates who embraced social media as a branding tool were better able to control their public image and connect with voters in a more dynamic, interactive way.[67]

Challenges of Engaging a Disillusioned Electorate (Addresses RQ3)

A recurring theme in the interviews was the difficulty of engaging the electorate. Voter apathy and political cynicism were at unprecedented levels during the 2016 elections, largely due to economic challenges, corruption scandals, and widespread dissatisfaction with the political establishment. Candidates and communication professionals alike perceived this sense of disillusionment as creating significant barriers to political branding efforts. One candidate explained: 'We were fighting against a tide of distrust. People didn't believe in politicians anymore, so no matter how strong your brand was, it was hard to get through to them' (PC7). This widespread scepticism often undermined even the most well-crafted branding strategies. Several candidates mentioned that their focus on transparency and honesty was met with indifference or suspicion from voters (PC8). As another candidate put it, 'No matter how transparent we were, people just assumed we were hiding something' (PC9).

[64] Gibson & Römmele (no 11).

[65] Ibid.

[66] Iosifidis & Nicoli (no 40), *Digital Democracy, Social Media and Disinformation* (Abingdon: Routledge, 2020).

[67] Braun & Clarke (no 50).

The data also revealed that the most successful candidates were those who connected with voters on a more personal level. These candidates often focused on local, community-based issues, rather than abstract national debates. As one communication professional noted, 'We found that the only way to break through the distrust was to focus on very specific, local issues that mattered to people in their daily lives' (CP2). By emphasising their connection to the community, these candidates positioned themselves as problem-solvers rather than distant politicians. Another challenge was the persistent political disengagement among younger voters.[68] Although youth were more likely to engage with candidates through social media, they were still less likely to vote than older generations. One candidate reflected, 'Social media helped us reach younger voters, but it didn't always translate into votes. They might like your post, but that doesn't mean they're going to show up at the polls' (PC4).[69] This highlights the limitations of digital branding in addressing deeper structural issues such as voter apathy and disconnection from the political process—a broader disengagement that was already visible in turnout patterns (see Figure 1).

The Limitations of Branding in Addressing Systemic Issues

A critical theme that emerged from the analysis was the limitation of political branding in addressing broader systemic issues. While candidates invested heavily in crafting personal brands that were authentic, transparent, and relatable, these efforts were often overshadowed by the structural problems that plagued the Cypriot political system. Economic instability, corruption, and the unresolved Cyprus dispute loomed large over the elections, and many candidates expressed frustration that their branding efforts were not enough to overcome these challenges.[70]

Several participants noted that political branding could help build a connection with voters, but it could not provide the substantive solutions required to address the root causes of voter dissatisfaction. One communication professional explained, 'You can have the best brand in the world, but if voters don't believe that real change is possible, it won't matter'(CP2).[71] This sentiment was echoed by multiple candidates,

[68] Cyprus Mail, 'Our View: Today's Youth Are Justified in Abstaining from Politics' (19 March 2016), available at https://archive.cyprus-mail.com/2016/03/19/view-todays-youth-justified-abstaining-politics/; European Commission, 'Standard Eurobarometer 85: First Results (Spring 2016)', available at https://eaca.eu/wp-content/uploads/2016/08/Eurobarometer85_EN.pdf.

[69] Lees-Marshment & al. (no 4) 103-104.

[70] Ibid., 105–106.

[71] Ibid., 106.

who expressed the view that while branding could help win votes in the short term, it could not fundamentally alter the dynamics of the Cypriot political landscape.

Discussion

The findings from this study provide important insights into the role of political branding in the 2016 Cyprus parliamentary elections. These insights help illuminate the broader context of political branding in contemporary elections, particularly in environments where voters are increasingly disillusioned with traditional political systems. The prominence of authenticity and local problem-solving in the coded material (see Figure 2) supports this interpretation.

Personal Branding and Authenticity

One of the central findings of this study is the critical role of authenticity in personal branding. As highlighted by both candidates and communication professionals, voters in Cyprus were disillusioned with political parties, which necessitated a shift towards personalising political campaigns. Candidates who succeeded in crafting an authentic, transparent image that resonated emotionally with voters were more effective in building trust.[72]

The importance of storytelling also emerged as a key tool for candidates to build an emotional connection with the electorate. By emphasising personal struggles and their understanding of voter challenges, candidates humanised themselves and created a more relatable brand.[73] This approach is consistent with the literature, which argues that storytelling is an essential element in effective political branding.[74]

However, the study also revealed that personal branding has its limits. While some candidates were able to differentiate themselves from their competitors through personal stories and values, others struggled to balance their personal identity with the collective identity of their political party.[75] This tension between individual and party branding is a recurring challenge in multi-party systems, where candidates must align their personal narratives with broader party strategies.

[72] Lees-Marshment & al. (no 4) 85–87.
[73] Smith & French (no 20).
[74] Scammell (no 19).
[75] Ibid., 187.

Digital Media as a Branding Tool

The role of digital media in political branding was another prominent theme in the findings. Social media platforms provided candidates with a direct communication channel to engage with voters, allowing for greater visibility and interaction. Candidates who used platforms such as Facebook and Twitter to share their thoughts, daily activities, and responses to current events boosted their visibility.[76] Digital media allowed candidates to present a more authentic version of themselves, bypassing traditional media filters.

This finding aligns with research that highlights how digital media has become an essential tool for political marketing, especially among younger, more tech-savvy voters.[77] However, the study also highlighted the challenges posed by digital media. Candidates faced difficulties in maintaining consistency across platforms, dealing with misinformation, and managing negative comments.[78] These challenges underscore the importance of strategic social media management in political branding.

Engaging a Disillusioned Electorate

The study's findings also underscore the difficulty of engaging the electorate. While personal branding and digital media provided candidates with tools to reach voters, these efforts were not always enough to overcome deep-rooted cynicism and apathy.[79]

Candidates found it particularly difficult to engage younger voters, who were more likely to interact with campaigns on social media but less likely to translate that engagement into actual votes.[80] This finding highlights a key limitation in the power of branding: while it may help candidates reach a wider audience, it does not necessarily guarantee voter participation.

This mirrors findings from other research, which suggests that while political branding can influence voter perceptions, it is not a panacea for broader issues such as voter apathy, disconnection from the political process, and distrust in political institutions.[81]

[76] Enli (no 29).
[77] Kaplan & Haenlein (no 34).
[78] Gibson & Römmele (no 11).
[79] Stromer-Galley (no 28).
[80] Ibid., 115–117.
[81] Creswell (no 47) 199–201.

Limitations of Branding in Addressing Systemic Issues

Our findings emphasise that political branding in Cyprus 2016 hinged less on novel slogans than on perceived authenticity under conditions of economic strain and distrust. Voters rewarded consistency across channels and over time, while punishing perceived incongruence. Branding travelled through a hybrid media system in which digital cues amplified organisational credibility and local constituency work. In small and fragmented systems, brand strength is bounded by structural factors such as turnout dynamics, party-system legacies, and institutional trust.

Conclusion

This study has provided a comprehensive analysis of political branding strategies during the 2016 Cyprus parliamentary elections, offering insights into the evolving nature of political communication in a digital age. The findings underscore the critical role that personal branding and digital media play in modern political campaigns, but they also highlight the inherent limitations of branding in addressing broader systemic challenges.

Overall, the findings suggest that branding can help candidates differentiate and communicate effectively, but it remains bounded by structural constraints such as institutional distrust, economic strain and unresolved systemic issues. Practical implications for campaigns in small, fragmented systems include prioritising cross-channel consistency, grounding messaging in local problem-solving, and investing in rapid response capacity for online reputational threats.

Reframing the 2016 contest through a branding lens shows how authenticity claims and media hybridity intersected with structural constraints in Cyprus. Future research should compare adjacent small States (e.g., Malta, Greece, and Ireland) to test where branding can offset (rather than merely reflect) low trust and participation.

Political branding, while useful in election campaigns, cannot substitute substantive policy solutions. This study suggests that political actors in Cyprus and beyond must complement their branding strategies with concrete policy proposals that address the real concerns of voters. Only by doing so can they hope to rebuild trust and re-engage an increasingly disillusioned electorate.[82]

[82] Ibid., 193.

References

Allcott Hunt & Matthew Gentzkow, 'Social Media and Fake News in the 2016 Election' (2017) 31(2) *Journal of Economic Perspectives* 211–236.

Baum M.A. & Philip BK Potter, *War and Democratic Constraint: How the Public Influences Foreign Policy* (Princeton, NJ: Princeton University Press, 2015).

Baym G., 'Beyond Bias: Conservative Media, Documentary Form, and the Politics of Hysteria' (2022) 137(4) *Political Science Quarterly* 787–788.

Brader T., *Campaigning for Hearts and Minds: How Emotional Appeals in Political Ads Work* (University of Chicago Press, 2006).

Braun Virginia & Victoria Clarke, 'Using Thematic Analysis in Psychology' (2006) 3(2) *Qualitative Research in Psychology* 77–101.

Chadwick A., *The Hybrid Media System: Politics and Power* (Oxford, UK: Oxford University Press, 2013).

Coleman Stephen, *How Voters Feel* (Cambridge, UK: Cambridge University Press, 2013).

Creswell John W., *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches* (Sage, 2014).

Cyprus Mail, 'Our View: Today's Youth Are Justified in Abstaining from Politics' (19 March 2016), available at https://archive.cyprus-mail.com/2016/03/19/view-todays-youth-justified-abstaining-politics.

Enli G., 'Twitter as Arena for the Authentic Outsider: Exploring the Social Media Campaigns of Trump and Clinton in the 2016 US Presidential Election' (2017) 32(1) *European Journal of Communication* 50–61.

European Commission, 'Standard Eurobarometer 85: First Results (Spring 2016)', available at https://eaca.eu/wp-content/uploads/2016/08/Eurobarometer85_EN.pdf.

Finlay Linda, 'Negotiating the Swamp: The Opportunity and Challenge of Reflexivity in Research Practice' (2002) 2(2) *Qualitative Research* 209–230.

Gainous J. & Kevin M. Wagner, *Tweeting to Power: The Social Media Revolution in American Politics* (Oxford, UK: Oxford University Press, 2014).

Gibson Rachel K., *The Internet and Democracy in Global Perspective: Voters, Candidates, Parties, and Social Movements* (Cham, Switzerland: Springer, 2016).

Gibson Rachel & Andrea Römmele, *Political Communication in the Online World* (Abingdon, UK: Routledge, 2015).

Harada D., 'Micro-Targeting and the Future of Political Communication' (2017) 14(2) *Journal of Information Technology & Politics* 98–111.

Henneberg Stephan C.M. & Nicholas J. O'Shaughnessy, 'Political Relationship Marketing: Some Micro/Macro Thoughts' (2007) 23(9–10) *Journal of Marketing Management* 7–29.

Iosifidis P. & N. Nicoli, *Digital Democracy, Social Media and Disinformation* (London, UK: Routledge 2020).

Iosifidis P. & N. Nicoli, 'The Battle to End Fake News: A Qualitative Content Analysis of Facebook Announcements on How It Combats Disinformation' (2020) 82(1) *International Communication Gazette* 60–81.

IPU Parline, *Vouli Antiprosopon (May 2016): Election Results—Cyprus* (2016) available at https://data.ipu.org/parliament/CY/CY-LC01/election/CY-LC01-E20160522.

Kaplan A.M. & Michael Haenlein, 'Users of the World, Unite! The Challenges and Opportunities of Social Media' (2010) 53(1) *Business Horizons* 59–68.

Karatzogianni A. & al., 'Social Media, Political Branding, and the 2016 Cypriot Parliamentary Elections' (2018) 10(3) *International Journal of Electronic Governance* 223–238.

Kassimeris Ch., *The Politics of Football: Political Crises and Political Branding* (Basingstoke, UK: Palgrave Macmillan, 2009) 134–135.

Kreiss D., *Prototype Politics: Technology-Intensive Campaigning and the Data of Democracy* (Oxford, UK: Oxford University Press, 2016).

Kruikemeier Sanne & al., 'Political Campaigning on Social Media: Politicians, Audiences, and the Mediation of Political Communication on Facebook and Twitter' (2013) 30(1) *Political Communication* 10–28.

Kuckartz U., *Qualitative Text Analysis: A Guide to Methods, Practice, and Using Software* (London, UK: Sage, 2014).

Lance Bennett W. & Robert M. Entman, *Mediated Politics: Communication in the Future of Democracy* (Cambridge, UK: Cambridge University Press, 2001).

Lees-Marshment J. & al., *Political Marketing: Principles and Applications* (3rd edn, Abingdon, UK: Routledge, 2019).

Lundberg Ch., 'Political Branding in the Modern Age: Effective Strategies, Tools, and Techniques' (2019) 18(2) *Journal of Political Marketing* 105–110.

MacQueen M.M., *Team-based Qualitative Research*, (Altamira Press, 2005).

Marcus G. E., W. R. Neuman & M. MacKuen, *Affective Intelligence and Political Judgment* (University of Chicago Press, 2000).

Marland A. & Thierry Giasson, 'From Brokerage to Boutique Politics: Political Marketing and the Changing Nature of Party Politics in Canada' (2013) 46(4) *Canadian Journal of Political Science* 961–972.

Mazzoleni Gianpietro, *The Media and Neo-Populism: A Contemporary Comparative Analysis* (Westport: Praeger, 2003).

Moffitt B., *The Global Rise of Populism: Performance, Political Style, and Representation* (Stanford, CT: Stanford University Press, 2016).

Nabi R. L., 'A Cognitive–Functional Model for the Effects of Discrete Negative Emotions on Information Processing, Attitude Change, and Recall' (1999) 9(3) *Communication Theory* 292–320.

Nicoli N. & M. Komodromos, *Principles of Public Relations* (Nicosia: University of Nicosia Press, 2013).

Norris Pippa, *A Virtuous Circle: Political Communications in Post-Industrial Societies* (Cambridge, UK: Cambridge University Press, 2000).

Ott B.L., 'The Age of Twitter: Donald J Trump and the Politics of Debasement' (2017) 34(1) *Critical Studies in Media Communication* 59–68.

Pich C., G. Armannsdottir & D. Dean, 'Exploring the Process of Creating and Managing Personal Political Brand Identities in Non-Party Environments: The Case of the Bailiwick of Guernsey' (2020) 19(4) *Journal of Political Marketing* 414–434.

Pich C. & K. Podnar, 'Political Brand Identity: An Examination of the Complexities of Conservative Brand and Internal Market Engagement During the 2010 UK General Election Campaign' (2018) 24(1) *Journal of Marketing Communications* 100–117.

Republic of Cyprus, Ministry of Interior, *Elections Results (Official Portal)* (n.d.), available at https://www.gov.cy/moi/en/ministry/departments/civil-registry-section/central-electoral-service/4350-2.

Scammell M., 'The Internet and Civic Engagement: The Age of the Citizen-Consumer' (2000) 17(4) *Political Communication* 351–355.

Scammell M., 'Political Brands and Consumer Citizens: The Rebranding of Tony Blair' (2007) 611(1) *The Annals of the American Academy of Political and Social Science* 176–192.

Scammell M., *Consumer Democracy: The Marketing of Politics* (Cambridge, UK: Cambridge University Press, 2014).

Smith Gareth & Alan French, 'The Political Brand: A Consumer Perspective' (2009) 9(2) *Marketing Theory* 209–226.

Stromer-Galley J., *Presidential Campaigning in the Internet Age* (2nd edn, Oxford, UK: Oxford University Press, 2019).

Theocharis Y. & Will Lowe, 'Does Facebook Increase Political Participation? Evidence from a Field Experiment' (2016) 19(10) *Information, Communication & Society* 1465–1486.

Tracy Sarah J., *Qualitative Research Methods: Collecting Evidence, Crafting Analysis, Communicating Impact* (Wiley-Blackwell, 2019).

Vaccari C. & al., 'Social Media and Political Communication: A Survey of Uses and Effects' (2015) 9 *International Journal of Communication* 2046–2066.

Watchdog Journalism: A Tool to Strengthen Democracy?

Nicolas Kyriakides[1] **and George Isaia**[2] [3]

Abstract

This article explores the evolution, impact, and global landscape of watchdog journalism in modern democracies. Several organisations, including FactCheck.org in the United States, Lupa in Brazil, Vouliwatch in Greece, RKW in Poland, Nomoplatform in Cyprus, and the 'platform beat' globally, are explored as a critical mechanism for enhancing transparency, accountability, and democratic integrity. From its origins in the Enlightenment to its key role in uncovering 20th century scandals like Watergate and exposing governmental misconduct globally, watchdog journalism has been pivotal in informing citizens and holding powerful entities accountable. Recommendations to strengthen this vital function include supporting legal protections for journalists, fostering media literacy, and advocating for robust civil society engagement. This article underscores the indispensable role of watchdog journalism in safeguarding democratic values and promoting informed civic participation.

Keywords: watchdog journalism; accountability; transparency; investigative reporting; democracy

Introduction

Watchdog journalism, also known as investigative journalism, refers to a specialised form of investigative reporting that aims to monitor and expose misconduct, corruption, or falsehoods in the public sphere. While it shares many similarities with traditional journalism—such as professional standards of reporting, research, and fact verification—watchdog journalism often focuses more intensely on holding powerful institutions and individuals accountable, helping citizens to better understand what is happening in their countries and make more informed decisions at the polls.[4] Put

[1] Assistant Professor, Director (Procedural Law Unit), School of Law, University of Nicosia

[2] Executive Director of Oxygen for Democracy

[3] The authors wish to thank Márton Tóth and Alexandra Akins for their valuable research assistance.

[4] LibertiesEU, 'What Is Watchdog Journalism: Definition, Examples, Effects On Democracy' (Liberties.eu, 17 January 2022), available at https://www.liberties.eu/en/stories/watchdog-journalism/43959 (last

differently, watchdog journalism follows liberal ideals, and seeks to operate in accordance with the liberal view of the press, which is that the press is there to transmit ideas and serve the public while defending freedom of expression, especially against the government, which, according to this theory, is the greatest threat to the right to free speech.[5]

There are various examples of watchdog journalism, such as the Watergate scandal, which caused President Nixon to resign, and the 2009 report that uncovered widespread misuse of expenses by members of parliament in the UK. Watchdog journalism is generally supported by people in countries with good democracies. In a 2013 UK poll, 53% of the public and 88% of journalists and other media opinion-formers reported believing that investigative journalism has a positive impact on the UK's democracy.[6] Research found that 73% of adults in the US think it's important for journalists to function as watchdogs over elected officials.[7]

With watchdog journalism's function in societies as a news outlet and information source for citizens, it is no surprise that authoritarian countries such as North Korea and Russia prioritise silencing, threatening, and prosecuting watchdog journalists. Globally, there is a need for stricter legislation to protect watchdog journalists from States who wish to take extreme measures to silence them, from frivolous lawsuits (known as 'SLAPPs'[8]) to imprisonment or even murder.

This article employs a descriptive and analytical approach rooted in qualitative research. It begins with a conceptual overview of watchdog journalism, tracing its evolution and examining terminology that intersects with that of investigative journalism. The study further analyses selected case studies to illustrate the practical manifestations and broader societal implications of watchdog reporting. In addition, the article engages in a critical examination of the role and implications of investigative journalism within the broader geographical landscape, concluding with a set of informed recommendations aimed at strengthening watchdog practices in contemporary media environments.

---

accessed 15 February 2024).

[5] N. Jebril, *Is Watchdog Journalism Satisfactory Journalism?* (Reuters Institute for the Study of Journalism, 2013).

[6] M. Kowalewski, 'Impact of Investigative Journalism' (YouGov, 2013) https://yougov.co.uk/consumer/articles/8012-impact-investigative-journalism (last accessed 15 February 2024).

[7] Ibid.

[8] European Parliamentary Research Service, *The Role of Investigative Journalism to Uncover Fraud and Corruption in Europe*, June 2022) https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/747083/IPOL_BRI(2023)747083_EN.pdf (last accessed 15 February 2024).

The Evolution of Watchdog Journalism and its Relevance to Modern Democracies

The Dawn of Journalism

Journalism is the collection, preparation, and distribution of news and related commentary and feature materials through print and electronic media, such as newspapers, magazines, books, blogs, webcasts, podcasts, social networking, and social media sites.[9] Journalism has a long history: There is evidence even in ancient Rome (59 BCE)—a sort of news sheet called *Acta Diurna* existed alongside documents that contained information of public interest, such as public speeches.[10] As time passed, journalism became more varied, as did its contents.

Shane Eisenman says that media and journalism have four main tasks: to entertain, to regulate, to inform, and to persuade or present opinion.[11] Consequently, we can classify the genres into two groups according to their function: regulators and opinion-providers. The former is the essence of watchdog journalism. Alongside regulation, watchdog journalism also functions to present opinion, for example on political leaders or parties, and to incite those consuming this media to take action, such as doing their own research or voting. To an extent, one can argue that this type of journalism also fulfils the task of persuasion, albeit unintentionally. When informing the public about politics, it is almost impossible not to be subjective. It is true that opinion-based news is not uncommon, but a question many ask is when did the standard for news become more objective and regulative? At what point did people expect to turn on a television and be presented only with facts, with no additional viewpoint? Of course, this change happened differently and at different times in each country, but there are common features. Generally, objective journalism and watchdog journalism evolved from 'free journalism'[12] and freedom of expression due to this type of journalism being more critical and serving the public interest, usually against the government or other economic or social actors, as it draws attention to shortcom-

[9] Encyclopaedia Britannica, 'Journalism | Definition, History & Facts' (2024) available at https://www.britannica.com/topic/journalism (last accessed 16 February 2024).

[10] Ibid.

[11] Shane Eisenman, 'Watchdog Journalism: Function and Future' (Columbia University, 1996) https://www.ee.columbia.edu/~shane/words/watchdog.htm (last accessed 16 February 2024).

[12] In countries with free journalism, journalists are able to express any opinions they want, even if these criticise the government or other organisations. Cambridge Dictionary (2024) https://dictionary.cambridge.org/dictionary/english/free-press (last accessed 16 February 2024).

ings and problems. Although it often develops in a free speech environment,[13] it is important to note that watchdog journalism often emerges non-democratic contexts. Some suggest that the structural linkages between non-democratic political systems and limited press freedom are key determinants of the overall role of watchdogs and have a major impact on their emergence and functioning.[14] However, under these circumstances, it cannot develop and achieve its goals and it remains a partisan or guerilla style of journalism, which seeks to undermine and change the existing system.

The Origins of Watchdog Journalism

Watchdog journalism first took shape in the 17th and 18th centuries, during the Enlightenment, when the concept of journalism as a check on power, rather than just information, gained ground. During this period, pamphlets and articles were already openly criticising or mocking political leaders, institutions, rulers, and other public figures.[15] In many cases, watchdog journalism started in magazines where negative cases affecting society were reported and published.

Watchdog journalism remained in the realm of critical gossip magazines until the 19th century and the dawn of the 20th century, when critical and investigative journalism regained ground thanks to the 'muckraking'[16] movement.[17] Critical newspapers, often looked down upon and labelled as muckraking, expanded into new sectors and areas, and journalists began to address community issues such as social injustice, corruption, and corporate and power abuses, bringing about major reforms and changes in the journalist profession. This investigative writing was still often mixed

[13] R. Drüeke, 'Medien, Öffentlichkeit und Demokratie: Zur Watchdog-Funktion von Medien' ['Media, the Public and Democracy: On the Watchdog Function of the Media'] (2018) 31(3) *Forschungsjournal Soziale Bewegungen* 19–28.

[14] M. Márquez-Ramírez & al., 'Detached or Interventionist? Comparing the Performance of Watchdog Journalism in Transitional, Advanced and Non-Democratic Countries' (2019) 25(1) *The International Journal of Press/Politics* 53, 55.

[15] Pascal Verhoest, 'Seventeenth-Century Pamphlets as Constituents of a Public Communications Space: A Historical Critique of Public Sphere Theory' (2018) 36(1) *Theory, Culture & Society* 47, 50.

[16] The term 'muckraker' was attributed to American writers who identified with pre-World War I reform and exposé writing. Muckrakers provided detailed, accurate journalistic accounts of the political and economic corruption and social hardships caused by the power of big business in a rapidly industrialising United States.

[17] M. R. Neuzil, 'Muckraking and the Progressive Movement: The Interaction of Journalism and Society, 1902-1911' (PhD Dissertation, University of Minnesota, 1993).

with sensationalist tabloid news, but closer attention was paid to investigating and reporting information objectively.[18]

Later, investigative journalism increasingly took on social issues and brought them to the public's attention, while tackling a wider range of topics, from health and women's equality to conservation and crime. This activity instigated major social and political changes, leading to reforms in areas such as civil rights, women's rights, environmental and health reforms, regulation of economic and political power, and crime.[19] During this period, several influential articles were published, leading to social mobilisation, new reforms, and legal norms. This kind of journalism was often accused of being biased and newsy, so it progressively detached from the tabloids, seeing the creation and institutionalisation of investigative groups in the 1960s and 70s. Alongside the press, other social observer organisations emerged, and government agencies provided support to the investigative bodies, giving birth to modern watchdog journalism.[20]

Looking for a definable moment when journalism moved to objectivity, we can look at the Vietnam War, when journalism became critical of the war and the decisions of the government.[21] According to Hallin,[22] the media did not become more oppositional; rather, journalists remained committed to objectivity, publishing information from their informants, which was influenced by the will and needs of those in power. This meant that the emergence of negative attitudes towards the war created controversy and new anti-war opinions that broke the pro-war consensus among the elite, and the media therefore had to follow and report this objectively.[23]

Watchdog journalism and opinions toward it were forever changed in the 1970s with the Watergate scandal in the US. The investigative work of the *New York Times* journalists shed light on the abuse of power and marked a major turning point for journalism. The scandal led to the resignation of President Richard Nixon, signalling

[18] J. A. Grenier, Jr., 'The Origins and Nature of Progressive Muckraking' (PhD dissertation, University of California-Los Angeles, 1965).

[19] W. Lawrence Bennett & William Serrin, 'The Watchdog Role' in G. Overholser & K. H. Jamieson (eds), *The Press* (Oxford University Press, 2005) 170, 175.

[20] Ibid.

[21] L. Gershon, 'When Did the Media Become a "Watchdog?"' *JSTOR Daily* (2018) available at https://daily.jstor.org/when-did-the-media-become-a-watchdog/ (last accessed 16 February 2024).

[22] D. Hallin, 'The Media, the War in Vietnam, and Political Support: A Critique of the Thesis of an Oppositional Media' (1984) 46(1) *The Journal of Politics* 2, 5.

[23] Ibid. 2–23.

a new era for investigative journalism.[24] The affair demonstrated the role of the media as a watchdog in democracies, reinforcing the faith and trust of citizens in investigative journalism, and creating a demand among the public for this type of journalism.

The digital age, from the late 20th century to the present day, has again fundamentally changed journalism. With the advent of the internet and the growing importance of digital media, information reaches more users, more easily than ever before. The use of online platforms and modern digital tools has become ubiquitous, allowing anyone to access information at any time and even to do their own research and publish in real time.[25] This cross-border network allows for a continuous flow of information and news, as well as easier access and greater transparency, if used properly, not to mention the possibility of global collaboration. One has only to think of revelatory outlets like Wikileaks and Edward Snowden's website to see the power of digital media and digital journalism in exposing the secrets of the world's governments.[26]

Watchdog Journalism vs. Traditional Journalism and Other Forms of Fact-Checking

To better understand how watchdog journalism operates in practice, it is useful to examine the ecosystem of investigative and fact-checking organisations that support it. These entities form the structural backbone of accountability journalism, working both within and alongside mainstream media to verify claims, expose falsehoods, and uncover misconduct.

Key Investigative and Fact-Checking Organisations

According to Liu & al., there are five main categories of fact-checking entities: professional fact-checking organisations, mainstream news outlets, social media platforms, AI-based systems, and crowdsourced initiatives.[27] They share three core aspects: fact-checking, investigative efforts, and the goal of exposing malpractice and wrongdoing. Broadly speaking, they differ primarily in the scope of their investigations and

[24] D. A. Ritchie, 'Investigating the Watergate Scandal' (1998) 12(4) *OAH Magazine of History* 49.

[25] Boonchai Hongcharu, 'The changing roles of mass media amidst the growth of the digital media' (2024) 10(1) *Cogent Social Sciences*.

[26] Benedetta Brevini, 'WikiLeaks: Between Disclosure and Whistle□blowing in Digital Times' (2017) 11(3) *Sociology Compass*.

[27] Xingyu Liu & al., 'Checking the Fact-Checkers: The Role of Source Type, Perceived Credibility, and Individual Differences in Fact-Checking Effectiveness' (2023) 52(6) *Communication Research* 719–746.

in their degrees of autonomy and professionalism, which can impact the objectivity of their work.

Professional fact-checking organisations—also referred to as watchdog groups—typically employ seasoned journalists, editors, and researchers who rely on news archives, scientific studies, and statistical databases to assess the accuracy of claims. These organisations are often independent and not tied to daily news production, which allows them to act as informants or whistleblowers by supplying verified information to larger media outlets.[28]

Mainstream news outlets such as *The Washington Post*, *The Guardian*, and *The Financial Times* also engage in fact-checking as part of their journalistic practice. These organisations use professional journalists trained in verification techniques. However, the key difference lies in purpose: while mainstream journalism aims to inform the public through publishable stories, watchdog journalism and independent fact-checkers may work behind the scenes, not necessarily seeking publication but rather accuracy and accountability.

Social media platforms—such as Meta and X (formerly Twitter)—once played an active role in flagging potentially misleading content through third-party fact-checking programmes. However, in recent years, many of these programmes have been scaled back or discontinued, leaving misinformation detection to users themselves.[29] This shift has led to the rise of crowdsourced fact-checking, where platform users voluntarily flag or annotate misleading posts. An example of this is X's 'Birdwatch' programme, in which users contribute notes to clarify or contextualise questionable content.[30]

Furthermore, artificial intelligence (AI) has emerged as a powerful, though imperfect, tool in fact-checking. AI algorithms can process and analyse massive volumes of data rapidly, identifying inconsistencies and flagging potentially false claims. However, their limitations are significant—they often struggle with the subtleties of political language, irony, or context, making them prone to errors. As Liu & al. point out, AI-based fact-checking, while efficient, lacks the human judgment required to navigate nuanced and ambiguous claims effectively.[31]

As traditional media and watchdog organisations share overlapping functions, distinguishing between them remains complex. This complexity arises from their

[28] Ibid.

[29] *The Briefing*, Pew Research Center newsletter (10 January 2025).

[30] Liu & al. (no 27).

[31] Ibid.

shared commitment to holding power accountable, but divergence in methods, institutional affiliations, and operational frameworks. The next section will delve into these nuanced distinctions.

Traditional Journalism and Watchdog Journalism: A Comparative Overview

Traditional journalism, as a whole, is complicated to define because it is incredibly diverse and takes many forms. It is connected to a fundamental human right, that of freedom of expression. There are many ways to publish information, but some only allow for one-way information transmission, making participation impossible. This is typical in non-democratic governance. In democracies, journalism is one of the most important links between the people and the government.[32] In democracies, popular journalistic publications tend to promote accountability or justice, expose criminals, power-abusers, and corrupt actors, or contribute to change in laws and policies.[33]

Many journalists argue about the roles, goals, and responsibilities associated with traditional journalism. In general, the role of traditional journalism is to inform people about public affairs, provide information, communicate objective facts,[34] and establish a link between the government and the people. It is also essential that a free press creates a public space in a democracy and ensures the freedom of speech and expression while creating a community of shared interests and identity. Traditional journalism also contributes to the economy by providing a platform for advertisement and economic operation. Furthermore, traditional journalism aims to educate people, to a certain extent, by spreading knowledge and providing access to public affairs. It also aims to raise awareness, inform, share opinions, and last but not least, entertain by telling interesting stories. The influence of the mass media on the ethics of public life, characterised by its watchdog role in controlling the behaviour of government officials, is vital to democracy.[35] In this sense, the characteristics of traditional journalism are impartiality, accuracy, truth, independence, and objectivity, and its

[32] Josef Trappel & Tales Tomaz (eds), 'Success and Failure in News Media Performance: Comparative Analysis' in *Media for Democracy Monitor 2021* (Nordicom, 2022).

[33] Márquez-Ramírez & al. (no 14).

[34] Kendel Walters, 'Comparative and Critical Analysis: The Roles of Civic and Traditional Journalism' (unpublished PhD thesis, University of Nevada Las Vegas, 2011).

[35] W. Francke, 'The Evolving Watchdog: The Media's Role in Government Ethics' (1995) 537 *The Annals of the American Academy of Political and Social Science* 109.

main goal is to serve the public mostly by informing.[36] Accordingly, traditional journalism does not try to (morally) mobilise people nor intervene in public affairs, it only provides objective information.[37] It does not aim to bring social or political change, or promote certain agendas (political, social, economic, cultural, etc.) by highlighting them, but only to present them impartially. Although it links the people and the government, it does not do so with an aim to defend the interests of the public.

Watchdog journalism has several similarities with traditional journalism as they share many of the same basic principles (see Figure 1). First, watchdog journalism can be dispersed digitally or in print. Second, the role of the watchdog is often referred to as one of the most important functions of the press.[38] Third, both types of journalism are expected to answer the same types of questions: 'Who? What? When? Where? Why? How?'. Fourth, the goals of watchdog journalism have much in common with traditional journalism, like serving the public and informing people about public affairs, just like education and awareness-raising, and presenting information that is particularly important to the people.[39] Lastly, traditional journalism, including mainstream news outlets such as *The Washington Post*, *The Guardian*, and *The Financial Times*, engages in fact-checking as a fundamental part of its journalistic practice, much like watchdog organisations.[40] As Gina Rushton notes, 'fact-checking is as old as journalism itself',[41] and these established media institutions have a long-standing tradition of employing professional journalists trained in fact-checking and verification techniques.

However, the key difference lies in purpose; while mainstream journalism aims to inform the public through publishable stories, watchdog journalism and independent fact-checkers may work behind the scenes, not necessarily seeking publication

[36] M. R. Fancher, 'The 21st century journalist's creed' (2009) 63(3) *Nieman Reports* 35–38.

[37] J. Rosen & al., 'Questions and Answers about Public Journalism' (2000) 1(4) *Journalism Studies*.

[38] J. Serrin & W. Serrin, *Muckraking! The Journalism that Changed America* (The New Press, 2002); R. Islam, 'Into the Looking Glass: What the Media Tell and Why – An Overview' in R. Islam, S. Djankov & C. McLeish (eds), *The Right to Tell: The Role of Mass Media in Economic Development* (The World Bank 2002) 1–23; S. Djankov & al., 'Who Owns the Media?' (Working paper, Harvard University, 2002); A. Dyck & L. Zingales, 'The Bubble and the Media' in P. K. Cornelius & B. Kogut (eds), *Corporate Governance and Capital Flows in a Global Economy* (Oxford University Press 2002) 83–102; A. Dyck and L. Zingales, 'Media and Asset Prices' (Working paper, Harvard University, 2003).

[39] R. Rodriguez, 'A Vital Responsibility in Need of Support' (Spring 2008) *Nieman Reports* 62.

[40] Liu & al. (no 27).

[41] Gina Rushton, 'How Checking Facts Got Political' *ABC News (Australia)* (16 March 2025) https://www.abc.net.au/news/2025-03-16/fact-checking-partisan-media-meta-trump-politics/105046680 (last accessed 16 February 2024).

Watchdog Journalism vs. Traditional Journalism: Similarities and Differences

SIMILARITIES

- **Distribution Methods**
  Both use paper-based and modern media tools to disseminate information to the public.
- **Core Press Function**
  Watchdog role is recognized as a fundamental function of the press in both journalism forms.
- **Fundamental Questions**
  Both answer the essential journalistic questions: Who? What? When? Why? How.
- **Public Service Goals**
  Both serve the public interest and aim to inform citizens about important public affairs.
- **Fact-checking Practices**
  Both engage in verification and fact-checking as fundamental journalistic practices.

DIFFERENCES

	Traditional Journalism	Watchdog Journalism
Purpose	Inform public through publishable stories that meet editorial standards and news cycles.	Focus on accuracy and accountability, not necessarily resulting in immediate publication
Approach	Primarily reports on known issues and events that are already in the public domain.	Proactively investigates and uncovers hidden misconduct, corruption, or systemic problems.
Autonomy	Often constrained by media institutions, editorial policies, and commercial considerations.	Typically enjoys greater independence from editorial influence and institutional constraints.

Figure 1: Watchdog Journalism vs. Traditional Journalism: Similarities and Differences

Figure 1. A comparison of watchdog journalism vs traditional journalism

but rather accuracy and accountability. A 2023 briefing by the European Parliament notes that investigative journalism is distinguished not merely by the subjects it covers—such as fraud, corruption, or abuse of power—but by its proactive approach to uncovering these issues. Rather than (merely) reporting on known wrongdoing, investigative journalism seeks to expose previously hidden misconduct and bring it to light, making such practices not only visible to the public but also actionable by law enforcement and regulatory authorities.[42] The same report categorises investigative journalism into three distinct forms: a) uncovering infringements of law, corruption, or abuse; b) evaluating institutions and organisations in a manner akin to external audits; and c) documenting political, social, cultural, or economic trends. This typology underscores the broad societal role that investigative journalism plays beyond traditional reporting.

Additionally, traditional journalism—despite its commitment to objectivity and truth—is often constrained by the interests and limitations of the media institutions it serves.[43] In contrast, watchdog organisations typically operate with greater autonomy, free from the editorial influence and commercial pressures that shape mainstream media. As a result, they are widely seen as neutral, nonpartisan defenders of public transparency.

Watchdog Journalism: Intersecting Concepts and Empirical Examples

A Few 'Family Resemblances'

This section outlines a network of terms and concepts related to 'watchdog' and 'investigative' journalism that share common denominators. Paying homage to Wittgenstein's notion of family resemblance, which draws an analogy between similar linguistic conceptions and a family tree that binds relatives together,[44] the section articulates differences and points of overlap to better understand investigative journalism and its conceptual 'relatives'.

[42] European Parliamentary Research Service (no 8).

[43] Liu & al. (no 27).

[44] Ludwig Wittgenstein, *Philosophical Investigations* (G. E. M. Anscombe tr., Blackwell Publishing, 1953) §§66–71.

Whistleblowing and Investigative Journalism: Similarities and Differences

Figure 2. A comparison of whistleblowers and investigative journalists

While whistleblowing and watchdog/investigative journalism frequently work in tandem to expose wrongdoing, they remain fundamentally different in terms of their institutional frameworks and the level of protection afforded, with investigative journalists typically enjoying greater safeguards than whistleblowers (see Figure 2).

Investigative journalism is a critical, analytical, and thorough form of reporting. It goes beyond simply recounting daily events to uncover stories that are deliberately concealed and would not reach the public without journalistic intervention. Its scope includes exposing scandals, reviewing the functioning of governments, businesses, and other organisations, and bringing misconduct to light.[45] At its core lies a moral compass; investigative journalism is committed to seeking and communicating the truth.

Whistleblowers, on the other hand, are typically individuals—often insiders—rather than structured media organisations. They frequently initiate contact with journalists, providing sensitive information in the hope that their revelations will eventually be published. Unlike journalists, whistleblowers do not usually produce documented or systematic reporting of a case, and many seek anonymity due to the risks involved.[46] Although contact is often initiated by whistleblowers, the relation-

[45] Thomas Olesen, 'Whistleblowing and the Press: Complicating the Standard Account' (2023) 24(11) *Journalism* 2418.

[46] Ibid.

ship is not strictly one-directional; in some instances, journalists approach insiders to encourage disclosure. Whistleblowers frequently place their careers, reputations, and even personal safety at risk when exposing sensitive information.[47]

Together, whistleblowers and investigative journalists serve as complementary forces in the pursuit of truth. They share a common goal, defending human rights, exposing malpractice, and strengthening democracy. Their relationship is symbiotic—investigative journalists rely on whistleblowers for access to otherwise unattainable information, leads, and legitimacy, while whistleblowers depend on journalists for protection, credibility, and the amplification of their disclosures to a wider public.

Watchdog Journalism and Media Watchdogs

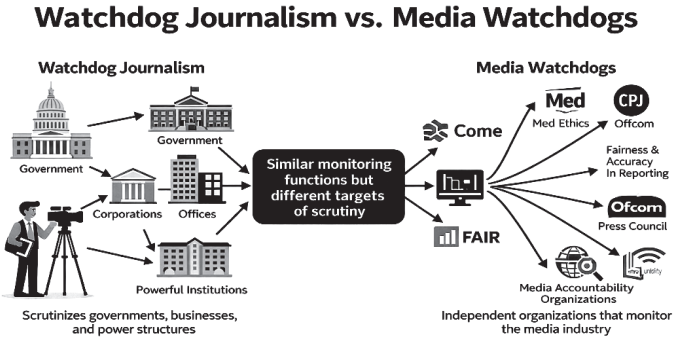

Figure 3. A comparison of watchdog journalism vs. media watchdogs

Watchdog or investigative journalism must be distinguished from media watchdogs. While the two concepts may sound similar and share certain points of overlap, they serve different functions (see Figure 3).

Both watchdog journalism and media watchdogs seek to expose wrongdoing and malpractice in the pursuit of transparency, integrity, and accountability. They embody the monitorial, detective, and watchdog roles; observing, investigating, and scrutinising to hold relevant actors accountable. However, their focus differs. Watchdog journalism primarily scrutinises governments, businesses, and other powerful

[47] UNESCO, 'Journalists and Whistleblowers: A Dynamic Relationship in the Fight against Corruption' (20 April 2023), available at https://www.unesco.org/en/articles/journalists-and-whistleblowers-dynamic-relationship-fight-corruption (last accessed 15 September 2025).

organisations, using investigation and reporting to inform the public. In contrast, media watchdogs oversee and examine the media industry itself. Independent from journalistic organisations, they work to ensure that the press upholds high standards of fairness, accuracy, and transparency.[48]

Media watchdogs are typically non-governmental organisations, advocacy groups, or regulatory bodies. Prominent examples include Ofcom in the United Kingdom, the Press Council of India, and the global Committee to Protect Journalists (CPJ).

Watchdog Journalism and Watchdog Media

Watchdog media refers to the broader media system or assembly that performs watchdog functions. It encompasses the ecosystem of journalists, news organisations, and media groups that—collectively, much like individual watchdog journalists—monitor, investigate, and scrutinise governments, businesses, and other powerful actors. Importantly, watchdog media not only carries out these functions but also enables, frames, and constrains them.[49] It shapes how watchdog activities such as monitoring, investigation, and scrutiny are funded, organised, and legitimised within the media landscape.

What Is the Overarching Purpose of Watchdog Journalism?

Watchdog journalism aims to increase accountability and transparency in every aspect of modern society, above all in politics. The term 'watchdog' refers to the behaviour of a journalist warning the public of certain problems through press reports, just as a watchdog warns of danger by barking.[50] In the same spirit, investigative journalism keeps an eye on powerful figures and institutions, promoting transparency, accountability, and democratic governance,[51] in line with moral and social norms and rules rooted in Enlightenment political ideals.[52]

---

[48] Susanne Fengler, 'Holding the News Media Accountable: A Study of Media Reporters and Media Critics in the United States' (2003) 80(4) *Journalism & Mass Communication Quarterly* 818.

[49] Rick Clifton Moore, 'Environmental Issues and the Watchdog Role of the Media: How Ellul's Theory Complicates Liberal Democracy' (2001) 21(5) *Bulletin of Science, Technology & Society* 325.

[50] M. Žuffová, 'Fit for Purpose? Exploring the Role of Freedom of Information Laws and Their Application for Watchdog Journalism' (2021) 28(1) *The International Journal of Press/Politics* 300–322.

[51] A. Carson, 'The Political Economy of the Print Media and the Decline of Corporate Investigative Journalism in Australia' (2014) 49(4) *Australian Journal of Political Science* 726; A. Carson, *Investigative Journalism, Democracy and the Digital Age* (1st edn, Routledge, 2019).

[52] Jeremy Bentham, 'Political Tactics' in Michael James, Cyprian Blamires & Catherine Pease-Watkin (eds), *Political Tactics* (Oxford University Press, 1999) 267.

By exposing hidden activities and scrutinising decision-makers, watchdog journalism pressures authorities to act responsibly and communicate openly with the public. Thus, it encourages those in power to act correctly and communicate their activities to the people, creating an opportunity for debate and cooperation. Watchdog journalism also gives people the chance to see the processes that govern society. It has a key role in highlighting the problems along which civil society can organise itself to work together and solve or at least counteract them.

Furthermore, the absence of watchdog journalism may indicate a lack of freedom of expression and a restricted or censored press. Such restriction of public space violates a fundamental principle of democracy. Journalism contributes to the functioning of democracy in a variety of ways: providing citizens with informative political news, offering political analysis and investigative reporting, presenting 'social empathy' stories that inform citizens about neighbours and groups they may not know or understand, providing a space for public debate, addressing the functioning of representative democracies, and mobilising citizens to participate in political life.[53]

The impact of watchdog journalism on democracy is thus significant. By ensuring transparency and accountability, journalists reveal the actions of government and parliament, hold those in power accountable for their decisions and actions, and reduce the risk of corruption or abuse of power. Watchdog journalists often bring to light issues that affect the proper functioning of a democracy, such as electoral fraud, corruption, attacks on civil liberties, and other threats.[54]

Informing the public remains a core function of watchdog journalism. Through in-depth investigations and research, journalists provide citizens the information they need to make responsible decisions on elections and other important policies. They also play an important role in monitoring companies to ensure that they are operating ethically, that they are not exploiting their sub-contractors, that their environmentally damaging activities are brought to light, and that their economic behaviours are transparent and accountable.[55] They also seek to provide a space for social justice, since investigative journalists frequently focus on social concerns, such as

[53] Michael Schudson, 'Reluctant Stewards: Journalism in a Democratic Society' (2013) 142(2) *Daedalus* 159; Michael Schudson, *Why Democracies Need an Unlovable Press* (Polity, 2008) 11–26.

[54] Nikki Usher & Sanghoon Kim-Leffingwell, 'How Loud Does the Watchdog Bark? A Reconsideration of Losing Local Journalism, News Nonprofits, and Political Corruption' (2023) 29(4) *The International Journal of Press/Politics* 960.

[55] Morten Skovsgaard, Lene Heiselberg & Kim Andersen, 'The Context-Dependent Demand for Watchdog Journalism: Dynamics in Audience Expectations for Journalists' Role Performance' (2024) 18(9) *Journalism Practice* 2454.

interference with or attacks on civil liberties, human rights abuses, inequality, discrimination, and abuses of power.[56] By bringing these issues into the open, they make them part of the public discourse in order to bring about social justice.

Examples of Watchdog Journalism

Other than offering alternative opinions, watchdog journalism also has a role as a regulator in modern democracies by investigating and publishing cases in the public interest linked to politicians, institutions, and economic and social actors, ensuring the transparency and accountability of those actors.[57] Thanks to the spread of democratic and liberal ideals, and the development of technology, watchdog journalism has become a key element of a modern democracy, which ensures the proper functioning of a democratic society.[58] It does this by putting information (and sometimes evidence) in the hands of the people and not allowing those in power to abuse their position.

The Watergate Scandal

One of the most famous cases of watchdog journalism in action is the Watergate scandal.[59] In 1972, five members of the Republican Party, allegedly under the directive of President Richard Nixon, broke into the headquarters of the Democratic National Committee in the Watergate Complex. Their objective was to install wiretaps and obtain sensitive information that could be used to gain an advantage in Nixon's re-election campaign.

Despite the White House's persistent denials, the scandal was eventually exposed thanks to the meticulous investigative efforts of *The Washington Post* journalists Carl Bernstein and Bob Woodward. Their dogged pursuit of the truth led to the discovery of a trail of corruption extending to the highest levels of government. In 1974, following a Supreme Court order to release the secretly recorded tapes, President Nixon resigned within weeks—the first US president to do so.[60]

---

[56] Ibid.

[57] William H. Dutton, 'The Fifth Estate Emerging through the Network of Networks' (2009) 27(1) *Prometheus*.

[58] Ibid.

[59] Ritchie (no 24).

[60] Harry Ransom Center, 'Inventory of the Woodward and Bernstein Watergate Papers' (The University of Texas at Austin, n.d.) available at https://norman.hrc.utexas.edu/fasearch/findingAid.cfm?eadid=00365 (last accessed 29 June 2025).

Although initially dismissed by some officials as a 'third-rate burglary',[61] the incident raised suspicions for Bernstein and Woodward, who were not satisfied with superficial explanations. They began by investigating separately but later joined forces, eventually linking the burglars directly to Nixon's administration. Their investigation culminated in the publication of two seminal works: *All the President's Men* and *The Final Days*, both offering comprehensive accounts of the Watergate scandal.[62]

The Watergate affair had a profound and lasting impact on American politics and governance. It reshaped the legal and political landscape by reinforcing the need for accountability and transparency among public officials.[63] The scandal led to several political reforms, including the establishment of special prosecutors, limits on presidential control over executive agencies, and new regulations on campaign financing.[64]

More broadly, Bernstein and Woodward's courage in confronting corruption inspired a new generation of journalists. Enrolment in journalism schools soared as their example highlighted the societal importance of investigative reporting.[65] Their efforts also condemned those in the press who merely echoed official statements without deeper analysis or questioning.[66]

WikiLeaks and Edward Snowden

WikiLeaks is a prominent whistleblowing platform that challenges traditional journalism. Founded in 2006, the non-profit organisation publishes leaked documents to expose misconduct and hold wrongdoers accountable. It has played a pivotal role in politics and international diplomacy by uncovering government corruption, unreported deaths, and sensitive classified information. WikiLeaks has helped usher in a new era of digital activism and inspired a new generation of social movements.[67]

In collaboration with major newspapers such as *The Guardian*, *The New York Times*, and *Der Spiegel*, WikiLeaks became internationally renowned for publishing the US war logs in 2010 (the Afghan War Logs and the Iraq War Logs). Containing

[61] Bob Woodward & Carl Bernstein, 'Watergate 40 Years Later: Nixon Waged Five Wars' (*The Fiscal Times*, 8 August 2014), available at https://www.thefiscaltimes.com/Articles/2012/06/11/Watergate-40-Years-Later-Nixon-Waged-Five-Wars (last accessed 11 February 2026).

[62] Harry Ransom Center (no 60).

[63] Stanley I. Kutler, 'In the Shadow of Watergate: Legal, Political, and Cultural Implications' (1994) 18(3) *Nova Law Review*.

[64] Ibid.

[65] Harry Ransom Center (no 60).

[66] Ibid.

[67] Brevini (no 26).

over 5,000 field reports and documents, these leaks provided a comprehensive account of both wars, revealing thousands of previously unreported deaths, including civilian casualties.[68] Another widely known case linked to WikiLeaks is the Panama Papers scandal, which exposed legal and financial practices associated with the Panama-based law firm Mossack Fonseca, revealing how elites used offshore companies to evade taxation.[69]

Another famous case is the Edward Snowden disclosures of 2013. Snowden, a former contractor for the US National Security Agency (NSA), revealed the illegal surveillance activities carried out by the NSA and its Five Eyes allies. By leaking sensitive information to journalists and the WikiLeaks platform, Snowden exposed the extent of government overreach. The revelations were so damaging to the NSA, the Obama administration, and the broader intelligence community that Snowden was charged with three felony counts by the US Department of Justice. He remains a divisive figure, embodying a traitor–hero dichotomy: celebrated by many for exposing the misuse of power and calling for reform, yet condemned by others as a threat to national security.[70] Alongside WikiLeaks founder Julian Assange, Snowden has since faced significant threats to his safety.[71]

Despite its accomplishments, critics argue that WikiLeaks has compromised its own effectiveness by straying from its core values. While it claims to resist oppressive power structures, some sceptics contend that WikiLeaks merely adopts the rhetoric of truth without upholding its standards, often prioritising sensationalism. Allegations of a 'one-leader cult' surrounding Assange have further fuelled claims that the organisation undermines journalistic integrity.[72] However, such criticisms could also be directed at many other media organisations and non-profits. What sets WikiLeaks apart is the scale of the opposition it faces: it remains one of the few whistleblowing platforms persistently targeted by governments, intelligence agencies, and traditional journalism. To this day, WikiLeaks endures legal challenges, cyberattacks, attempts

[68] Ibid.

[69] Ibid.

[70] Jessica Blusiewicz, 'The Case of Edward Snowden: A Different Path' (2014) 8(1) *Cornell International Affairs Review*.

[71] Luciana Amaral, 'Between Personal and Public Interests: A Look Back at the Impact of Snowden and WikiLeaks (Guest Blog)', *LSE Polis* (22 July 2014) available at https://blogs.lse.ac.uk/polis/2014/07/22/between-personal-and-public-interests-a-look-back-at-the-impact-of-snowden-and-wikileaks-guest-blog-polissummer/ (last accessed 15 September 2025).

[72] Danielle M. Logue & Stewart R. Clegg, 'WikiLeaks and *The News of the World*: The Political Circuitry of Labeling' (2015) 24(4) *Journal of Management Inquiry* 394, 396.

to dismantle its operations, and even reported threats to assassinate its leader, Julian Assange.[73]

The Covidgate Scandal

Another significant case study in watchdog journalism is the 2020 exposure of corruption in Zimbabwe's National Health Ministry by investigative journalist Hopewell Chin'ono. Dubbed 'Covidgate', the scandal involved Chin'ono uncovering fraudulent procurement deals with £47.5 million in funds meant for protective equipment for healthcare workers during the COVID-19 pandemic.[74] The exposé led to the dismissal and arrest of Zimbabwe's Health Minister, Obadiah Moyo.[75]

Despite this major journalistic achievement, Chin'ono faced State persecution and was arrested on charges widely seen as politically motivated. This action raised serious concerns about the state of press freedom in Zimbabwe.[76] Although the Constitution of the Republic of Zimbabwe guarantees freedom of expression and the press, the prosecution of journalists like Chin'ono suggests that these protections are often disregarded in practice.[77]

PEN International, an organisation dedicated to defending freedom of expression, issued a public statement urging citizens worldwide to appeal to Zimbabwean authorities to release Chin'ono and end the broader harassment of investigative journalists.[78] Similarly, the New York-based Committee to Protect Journalists (CPJ) condemned Chin'ono's arrest and advocated for his immediate release.[79]

Advancements in modern technology and tools mean that nearly every action is traceable and increasingly public. As information spreads more widely and rapidly, discrepancies become easier to detect—as demonstrated by the Covidgate scandal. However, detection alone is not enough. Meaningful support for journalists and the

[73] Brevini (no 26).

[74] G. Moyo & K. I. Phulu, 'The Weaponisation of the Coronavirus Crisis in Zimbabwe: Legal and Extra-Legal Instruments' (2021) 13(1) *iBusiness* 48, 50.

[75] BBC News, 'Hopewell Chin'ono: Whistle-blowing Zimbabwean journalist arrested' (20 July 2020) available at https://www.bbc.co.uk/news/world-africa-53477423 (last accessed 11 February 2026).

[76] Ibid.

[77] PEN International, 'Zimbabwe: Release Journalist Hopewell Chin'ono and Drop All Charges Against Him' available at https://www.pen-international.org/news/zimbabwe-release-journalist-hopewell-chinono-and-drop-all-charges-against-him (last accessed 29 June 2025).

[78] Ibid.

[79] Committee to Protect Journalists, 'Journalist Hopewell Chin'ono Arrested, Charged with Incitement in Zimbabwe' (20 July 2020) available at https://cpj.org/2020/07/journalist-hopewell-chinono-arrested-charged-with-incitement-in-zimbabwe/ (last accessed 11 February 2026).

exposure of corruption has largely come from the international community rather than from within Zimbabwe itself. The State's response suggests an intent to stifle future investigative efforts, reflecting a broader global trend of governments that intimidate or persecute investigative journalists and whistleblowers.

Overview of Watchdog Organisations Globally

A. Watchdog Organisations

In the previous sections, we discussed the various forms that watchdog journalism takes around the world, due to the different contexts in which it has developed and the diverse social, political, economic, and cultural circumstances in which it has been practiced. There can be significant differences between organisations in terms of methods and purpose. However, watchdog organisations tend to be grassroots (or civil) non-profit initiatives whose main objective is to monitor organisations, institutions, and other governmental and political institutions in the interest of society at large, with the involvement of the citizenry. Watchdog organisations strive to be independent to avoid the risk of bias and corruption that could jeopardise their objectives and credibility. Nevertheless, in trying to gain insight into the work of the targets they monitor, watchdog organisations risk being integrated or institutionalised.

These NGOs look beyond political and governmental institutions, as the public interest can be threatened from other areas as well. Many organisations have built-in observer organisations to ensure that they operate properly. There are also NGOs that act as external observers that try to bring as much information as possible to the public. Through critical monitoring, watchdog organisations and groups ensure that the work of all branches of government is free from corruption and abuse of power, and that the proper functioning of democracy is not threatened by waste, mismanagement of funds, and opaque legislation.[80] Ideally, the media should act as a fourth check on the three branches of power, recognising the flaws in the system and identifying problems, so that they can be communicated to the people and decision-makers and inform civil society and groups.[81]

Watchdog organisations can differ in many ways, and they evolve according to local needs and goals. This means that they do not always remain territorially bound.

[80] M. E. McCoy, *Scandal and Democracy: Media Politics in Indonesia* (Cornell University Press, 2019) 139.

[81] Ibid.

This has led to the emergence of many different types of watchdog organisations, such as consumer watchdogs, corporate watchdogs, charity watchdogs, advertising watchdogs, and government watchdogs.[82] Of these, government watchdog organisations play the most important role in political life, as they keep an eye on all three branches of government (executive, legislative, and judicial). As mentioned, they function as a sort-of fourth branch of power: 'by publicising corruption, scandal in high places, or the government's simple inattention to the needs of the people, the press could ensure that a nominally democratic government met its obligations to its constituents'.[83]

B. Watchdog Organisations Close to the EU and UN

There are many watchdog organisations in Europe, operating on the national, regional, and/or global levels. Some specialise in politics or decision-making, others in economic and social issues. There are also several monitoring organisations within the EU that ensure the EU functions properly, transparently, and without abuse.[84] They specialise in a wide range of issues, from the protection of peer groups with different legal statuses, to monitoring abuse and corruption, to ensuring accountability for any legal, economic, or political institution or actor. The objectives of these organisations are so diverse it is very difficult to group or define them. However, there are similarities in their methods and general characteristics.

Although the institutional structure of the EU is based on self-regulation and constant monitoring of other institutions through cooperation (mainly between the European Parliament, the Council of the European Union, and the European Commission), many organisations are not an integral part of the EU institutions, but nevertheless oversee their proper and transparent functioning.[85] Several organisations in the EU act as watchdogs against violations, such as the European Court of Human Rights, which ensures compliance with the European Convention on Human Rights.

---

[82] For more information on how and why watchdog organisations operate differently, see Trappel & Tomaz (no 32); Lars W. Nord, 'Investigative Journalism in Sweden' (2007) 8(5) *Journalism* 517; Lars W Nord & Tomas von Krogh, 'Sweden: Continuity and Change in a More Fragmented Media Landscape' in J. Trappel & T. Tomaz (eds), *The Media for Democracy Monitor 2021: How Leading News Media Survive Digital Transformation* (Vol 1, Nordicom, University of Gothenburg, 2021) 353.

[83] Mark Hampton, 'The Fourth Estate Ideal in Journalism History' in Stuart Allan (ed.), *The Routledge Companion to News and Journalism* (Routledge, 2009) 3–12.

[84] European Anti-Fraud Office (OLAF), 'OLAF Home' https://anti-fraud.ec.europa.eu/index_en (last accessed 15 September 2025).

[85] Ibid.

One internal control body is the European Anti-Fraud Office (OLAF), whose purpose is to investigate complaints, potential fraud concerning the EU budget, suspected corruption, and other abuses of power undermining EU institutions. OLAF investigates the risk of irregularities based even on anonymous reports and develops methods to help the European Commission remedy these problems. These institutionalised organisations, however, while fulfilling some of the functions of a watchdog, cannot be called watchdog organisations in the traditional sense, as they are not necessarily independent. Their main purpose is also not raising awareness, activating, and involving people, nor is it disseminating information.[86]

Beyond the EU, international organisations and institutions are also monitored by watchdog organisations. One such institution is the NGO UN Watch, which measures and monitors the performance of the UN, in accordance with its own statutes. It helps to achieve the goals of the UN more effectively by publicising their work, and by involving the public in putting positive pressure on decision-makers and giving people the opportunity to play a role in promoting the UN's work while representing their interests.[87]

Non-institutionalised organisations such as SDG Watch Europe also play an important role in EU compliance, covering four main areas to make the EU and its Member States (MS) work together more effectively. The focus of the organisation is advocacy and high-level policy coordination to enable the successful implementation and integration of the UN's Sustainable Development Goals (SDGs) (and the policy decisions and changes that go with them) at the MS level.[88] As a watchdog, it monitors and reports on the implementation of the SDGs by the EU and individual MS.[89] It checks whether MS have a strategy or implementation plan, how many targets are covered and to what extent, and whether they are implementing or institutionalising civil society involvement. Ongoing reporting allows the EU, the UN, as well as the public to monitor progress and potential problems. The organisation seeks to reach out to a wide range of civil society organisations inside and outside of the EU, involve as broad a range of civil society actors as possible in the implementation process of SDGs, and ensure that they are delivered and implemented as effectively as possible

[86] Ibid.

[87] UN Watch, 'About Us - UN Watch' available at https://unwatch.org/about-us/ (last accessed 15 September 2025).

[88] SDG Watch Europe, 'About Us' (n.d.) https://sdgwatcheurope.org/about-us/ (last accessed 15 September 2025).

[89] Ibid.

at all levels. By carrying out needs assessments among citizens, the NGO aims to ensure that the goals meet citizens' real needs and to bridge the gap between decision-makers and citizens by involving citizens in a bottom-up approach to legitimise decision-making.[90] The aim of the organisation is to give civil society organisations a greater voice and role in the implementation of SDGs and the related policy, preparation, decision-making, and advocacy process, and to enable citizens to monitor implementation and hold decision-makers to account.

Watchdog Networks

Watchdog organisations in Europe carry out a range of activities, from outreach and advocacy to lobbying. Often, smaller watchdog organisations may create a larger watchdog network to achieve a larger goal that they might not be able to individually.

The Better Regulation Watchdog initiative of Friends of the Earth Europe is one such network. It consists of European consumer, environmental, developmental, civil, and public health organisations, trade unions, and social justice organisations working together to promote and protect citizens' rights to high social, labour, environmental, consumer, and public health standards.[91] The network prioritises the protection of EU residents' rights. Important bases for the initiative are cooperation and the sharing of information and experience, public debates, the assessment of needs for more successful decision-making, as well advocacy on issues that affect citizens. With more than 50 organisations, the network raises awareness, puts pressure on decision-makers, and campaigns for the public interest.[92]

More recently, digital rights organisations are acting as a new breed of rights defenders. Experts, academics, NGOs, and public figures working against digital rights abuses has resulted in watchdog organisations to raise awareness of the new and unknown dangers posed by modern technology. The European Digital Rights (EDRi) network brings together precisely these actors to promote digital rights and thus create a Europe-wide digital rights movement.[93] They work to combat digital abuse by private and public actors to prevent them from using their power to manipulate the population through spreading misinformation or exploiting people's lack of infor-

[90] Ibid.

[91] BEUC, 'Better Regulation Watchdog Launches Website' available at https://www.beuc.eu/news/better-regulation-watchdog-launches-website (last accessed 15 September 2025).

[92] Ibid.

[93] European Digital Rights (EDRi), 'European Digital Rights (EDRi)' (27 September 2023) https://edri.org/ (last accessed 15 September 2025).

mation/understanding. They do so by raising public awareness, informing and educating the public, and enforcing rights. By mobilising the public, they help create an accountable technology market and a strong digital rights movement that can have a global impact.[94]

These networks go beyond national borders and focus on more than just policy. Arguably, they are not typical watchdog organisations since a defining feature is investigation and fact-finding, which is not necessarily the key tasks of the institutions mentioned above. In this sense, Amnesty International can be understood as a prominent actor within transnational advocacy networks rather than merely as an individual non-governmental organisation, as human rights NGOs operate through interconnected actors that mobilise information and coordinate action across borders.[95] Although formally constituted as a single entity, its influence has been linked to its embeddedness within broader advocacy structures that connect grassroots activists, national sections and international institutions. As a watchdog organisation, Amnesty conducts investigations and produces detailed reports on human rights violations, but the political impact of such activities depends upon their dissemination through networks that amplify claims and circulate information.[96] Through coordinated campaigns and the mobilisation of civic activists across its local groups and specialist networks, Amnesty functions as a key node within advocacy structures that seek to monitor power, promote accountability and reinforce international human rights norms.[97]

Such networks, which often operate at international or global level, are similar in objectives—to activate and involve the public in their activities and promote their causes by providing information. Their differences tend to be in their methods, their scope, and their relationship with the object of their observation, as some organisations are more independent and non-institutionalised.

The following section provides six examples of watchdog organisations to highlight similarities and differences around the world. Together, they provide insights into the promise, limitations, and potentials of watchdog journalism.

---

[94] Ibid.

[95] Amanda Murdie & Marc Polizzi, 'Human Rights and Transnational Advocacy Networks' in Jennifer Nicoll Victor, Alexander H. Montgomery & Mark Lubell (eds), *The Oxford Handbook of Political Networks* (Oxford University Press, 2016) 715–732.

[96] Stephen Hopgood, *Keepers of the Flame: Understanding Amnesty International* (1st edn, Cornell University Press, 2006) 105.

[97] European Digital Rights (EDRi) (no 93).

Watchdog Media Around the World

Watchdog organisations around the world monitor the work of branches of government and bring it to the public's attention. These organisations seek to instil a culture of active citizenship in all spheres of social life by giving citizens an insight into how these aspects work. These organisations play an important role within society particularly in the legislature, ensuring government accountability. These institutions are rapidly spreading around the world, and in Europe they have become commonplace, reporting on topics from politics to economics, consumer protection to industry.

Vouliwatch (Greece): Improving Everyday Accountability

In European countries with a democratic-corporatist tradition, watchdogs play an important role in the news and coverage of politics.[98] Founded in Athens, Greece in 2014, Vouliwatch is a non-profit watchdog organisation operating at a national level. Vouliwatch monitors the work of the Parliament and seeks to provide citizens with insights into legislature so as to hold elected representatives accountable at both national and international levels.[99] The aim of the organisation is integrating citizens into political life and decision-making processes by providing a platform for accountability of elected representatives. In doing so, it acts as a bridge between citizens and the politicians they elect, increasing transparency and encouraging active citizenship in political life.

Vouliwatch uses digital tools to make the work of parliamentarians more accessible and to encourage transparent communication between decision-makers and citizens. One tool that Vouliwatch runs and moderates is the Public Questioning platform, where citizens can put forth questions to their national and European members of parliament, who can post replies on the same site. Another digital tool is Votewatch, which allows users to track and receive information on bills, as well as monitor the voting behaviour of their MPs. Another important digital tool is the Policy Monitor, which informs citizens about key policies, allowing them to compare and comment on them and make new proposals.[100]

[98] Daniel Hallin & Paolo Mancini, *Comparing Media Systems: Three Models of Media and Politics* (Cambridge University Press 2004).

[99] Ibid.

[100] Ibid.

Since its inception, Vouliwatch has provided its services to the public free of charge; at the same time, it does not accept any monetary support from politicians, online advertisements, or companies—independence being a core value.[101]

Vouliwatch relies heavily on the possibilities and tools offered by modern technology to achieve its goals. The organisation constantly strives to improve its platforms and make them more user-friendly, while at the same time seeking to collaborate and expand its partnership network.

RKW (Poland): Improving Electoral Accountability

Similar to Vouliwatch, Ruch Kontroli Wyborów (RKW)[102] is a voluntary watchdog organisation that aims to hold presidential and parliamentary candidates accountable.[103] While Vouliwatch focuses on increasing citizen participation in everyday parliamentary decision-making by distributing information, RKW mostly focuses on monitoring electoral behaviours to improve the integrity of election outcomes.

RKW grew out of several major controversies in the November 2014 local government elections in Poland: a) These elections saw unexpectedly large proportions of spoiled ballots, and in some municipalities, over 50% of votes were invalid; b) the junior governing coalition partner Polish Peasant Party received unusually high support, which deviated substantially from exit polls; and c) the electoral commissions did not tabulate and report local election results in time, citing software failures. These factors generated mistrust amongst civil society groups towards the electoral process.[104]

To hold candidates to account, RKW created its own vote-counting system parallel to and independent from the National Electoral Commission to verify the official vote count. They formed a network of members in local electoral committees across administrative levels to improve transparency of information and enable cross-checking. They expanded the network by recruiting thousands of volunteers who chaired

[101] Vouliwatch, 'Empowering democracy', available at https://vouliwatch.gr/about/en (last accessed 15 September 2025).

[102] Stowarzyszenie Ruch Kontroli Wyborów https://stowarzyszenierkw.org/ (last accessed 11 February 2026).

[103] J. Fałkowski & P. J. Kurek, 'The Power of Social Mobilisation: The Impact of Monitoring the 2015 Presidential Elections in Poland' (2021) 49(1) *Journal of Comparative Economics* 38.

[104] Ibid.

electoral committees, scrutinised committee behaviours, and photographed district electoral commissions' statements.[105]

Although the exact scale of RKW's impact is difficult to measure due to the nature of their endeavour, nevertheless, as Fałkowski and Kurek found in their research, after the founding of RKW, electoral support for the incumbent candidate was substantially reduced compared to the outcome of the previous election in 2010. This indicates that RKW may have had a disciplining effect on local elections which improved electoral integrity.[106]

Nomoplatform (Cyprus): Keeping an Eye on Policymakers

Nomoplatform[107] is a watchdog journalism platform as well as a news portal and a database; it operates a website with a searchable archive and provides news on the Cyprus Parliament's activities and events. It also publishes summaries of the most important events in print via *Nomopress*.[108] Nomoplatform can therefore be seen as a hybrid initiative combining the functions and methods of investigative journalism with the role and techniques of traditional journalism in providing news and general reporting and the purpose of a database.

Nomoplatform is a politically independent digital legislative watcher that aims to increase transparency and encourage social participation in the political life of Cyprus by allowing residents to directly follow ('and on an ongoing basis'[109]) the submission, examination, and voting of bills, to receive parliamentary news, and to access information on individual Members of Parliament (MPs). The initiative operates on the national level.

Nomoplatform allows users to access the biographical data of all MPs, providing background on elected officials. A key feature of the platform is that the public can access the process of draft legislation from its submission, through its scrutiny, to its vote. The platform provides access to an archive of the data already collected, and it is possible to review the biographical data of MEPs.[110] On this platform, residents can

[105] Ibid.

[106] Ibid.

[107] Nomoplatform, https://www.nomoplatform.cy/ , accessed 11 February 2026

[108] Polcypress Publications, https://policypress.cy/ekdoseis-policypress/ (last accessed 11 February 2026).

[109] Nomoplatform, 'About - Nomoplatform' available at https://www.nomoplatform.cy/about/ (last accessed 15 September 2025).

[110] Nomoplatform, 'NGO Releases Nomoplatform™ Electronic Legislative Observatory' (2021) Nomoplatform available at https://www.nomoplatform.cy/ngo-releases-nomoplatform-electronic-legislative-observatory/ (last accessed 15 September 2025).

learn about the work of the Parliament, including the functioning of its committees and plenary sessions, and all related activities, as well as draft laws submitted to the parliament and their amendments and other provisions. For the initiative, it is of paramount importance that citizens gain insight into the inflexible legislative processes in the House of Representatives and have more information on the development of procedures.[111] An important activity of the initiative is the *Nomopress*, an independent political magazine that aims to engaged young people and provide high-quality information on the legislature.[112]

FactCheck.org (United States): Accountability or Polarisation?

Founded in 2003 by the Annenberg Public Policy Center of the University of Pennsylvania, FactCheck.org scrutinises government officials' declarations and highlights false statements as one of the pioneer watchdog organisations in the US. It describes itself as a 'non-partisan, non-profit consumer advocate for voters [aiming to] reduce the level of deception and confusion in U.S. politics'.[113]

As Amazeen argued, the organisation emerged and prospered amidst the 'decline among leading journalistic institutions', just as consumer reform movements emerged from a loss of trust in big businesses, serving as an alternative channel for grassroot actors to hold politicians accountable.[114] This, alongside advancements in information technology, initiated a new age of watchdog journalism in the US.

While FactCheck.org is generally regarded as more neutral and non-partisan than many mainstream media outlets, it has also been subject to criticism. A case study by Pham Phuong suggests that FactCheck.org may not be as politically neutral as widely assumed. In the lead-up to the 2020 US presidential election, the platform was observed to scrutinise statements made by Donald Trump more than twice as often as those made by Joe Biden,[115] suggesting a political agenda. While this apparent

[111] Ibid.

[112] 'NomoPress – by Nomoplatform' (n.d.) available at https://nomopress.cy/ (last accessed 15 September 2025); Nomoplatform, 'NomoPress Archives - Nomoplatform' (n.d.) available at https://www.nomoplatform.cy/category/nomopress/ (last accessed 15 September 2025).

[113] FactCheck.org, 'Our Mission', available at https://www.factcheck.org/about/our-mission/ (last accessed 15 September 2025).

[114] M. A. Amazeen, 'Journalistic Interventions: The Structural Factors Affecting the Global Emergence of Fact-Checking' (2017) 21(1) *Journalism* 95, 111.

[115] U. D. Pham Phuong, 'Check the checks: A comparison of fact-checking practices between newspapers and independent organizations during 2020 U.S. election presidential debates' (2024) 46(1) *News Research Journal* 71–95.

imbalance can be attributed to the fact that then-President Trump's statements more frequently featured false or misleading claims than those of Joe Biden, the disproportionate scrutiny of one politician over another may inadvertently fuel greater public scepticism toward watchdog organisations.

Mistrust risks spilling over into the broader political landscape. One limitation of watchdog journalism is its relative inability to hold politicians accountable or foster public trust in elected officials. Indeed, as Ferracioli & al. pointed out, on FactCheck.org, an overwhelming majority of statements scrutinised were labelled 'False' or 'Partially False', with barely any earning the label of 'True'.[116] Put differently, fact-checking organisations like FactCheck.org focus more heavily on debunking falsehoods rather than confirming accurate information. This pattern suggests that even in liberal democracies with active civil societies committed to accountability, media practices may unintentionally reinforce a culture of mistrust rather than alleviate it.

Overall, the rise of watchdog journalism since the founding of FactCheck.org could be read in one of two ways: on the one hand, FactCheck.org maybe be viewed as a pioneering force for good; on the other, the growing demand for watchdog journalism two decades later may reflect a deepening culture of polarisation and mistrust.[117]

Lupa (Brazil): Reinforcing the Status Quo?

In Brazil, journalism has traditionally steered clear of overt political engagement, yet the emergence of investigative and fact-checking ventures such as Agência Lupa signals a decisive break from that norm. Founded in 2015 by reporters who left legacy newsrooms, Lupa—together with Aos Fatos[118]—became one of the country's first independent fact-checking organisations. Operating outside the major media conglomerates gives these outlets a measure of autonomy, freeing them from routine alignment with powerful corporate or political interests.[119]

---

[116] P. Ferracioli & al., 'The Watchdog Role of Fact-Checkers in Different Media Systems' (2022) 10(5) *Digital Journalism* 717, 737.

[117] See section 2A 'The Dawn of Journalism'.

[118] Meta, 'Facebook launches news verification product in Brazil in partnership with Aos Fatos and Agência Lupa' (*Sobre a Meta (Facebook news site))*, 2018) available at https://about.fb.com/br/news/2018/05/facebook-lanca-produto-de-verificacao-de-noticias-no-brasil-em-parceria-com-aos-fatos-e-agencia-lupa (last accessed 11 February 2026).

[119] Francesco Zaffarano, 'Attack: Brazil's "Capitol"—Lupa Was Already Debunking Lies on TikTok' (*Reuters Institute*, 10 January 2023), available at https://reutersinstitute.politics.ox.ac.uk/news/attack-brazils-capitol-lupa-was-already-debunking-lies-tiktok (last accessed 1 July 2025).

Lupa quickly distinguished itself as a digital innovator, using social media channels to promote media literacy and expose misinformation. Under President Lula's administration (2023–), the group deepened its online presence, tailoring explainers and debunks for social media platforms to better connect with youth. Notably, it was the first Brazilian fact-checker to open a TikTok account, underscoring its commitment to public engagement through accessible, digestible content.[120]

The practical value of these efforts became particularly evident during the 2018 and 2022 presidential elections, when Brazil's Electoral Justice partnered with Lupa and similar outlets to counter an unprecedented wave of false claims circulating on social media.[121] Such collaboration highlights the growing recognition of fact-checkers as frontline actors in safeguarding electoral integrity.

Yet independence brings its own constraints. As Thales Lelo observes, organisations like Lupa must still navigate an ecosystem shaped by the expectations of institutional and philanthropic funders.[122] To remain financially viable, they often adopt editorial routines and professional norms set by those very backers—dynamics the fact-checkers were created to scrutinise in the first place.

Lupa's trajectory also mirrors the broader limits of watchdog journalism revealed in cases such as the US site FactCheck.org. Although Lupa operates separately from Brazil's traditional news outlets, it now collaborates regularly with them and relies on professional journalists to verify claims made by politicians, celebrities, and influencers across every major medium.[123] Quantitative analysis by Ferracioli & al. shows, however, that presidents and other senior office-holders are rarely the primary subjects of its checks; nearly half of the scrutinised statements came from political actors outside formal office.[124] Compounding this, Lupa's verdicts skew more often toward 'True' than those of comparable fact-checkers abroad—an outcome that may hint at excessive caution or complacency, though the evidence is not conclusive.[125]

In short, Lupa exemplifies both the promise and the persistent challenges of contemporary fact-checking. Its digital agility and institutional partnerships demonstrate how watchdog journalism can shore up democratic processes. At the same

[120] Ibid.

[121] Thales Lelo, 'The Rise of the Brazilian Fact-checking Movement: Between Economic Sustainability and Editorial Independence' (2022) 23(9) *Journalism Studies* 1077.

[122] Ibid.

[123] P. Biancovilli, L. Makszin & C. Jurberg, 'Misinformation on Social Networks During the Novel Coronavirus Pandemic: A Quali-Quantitative Case Study of Brazil' (2021) 21(1) *BMC Public Health* 1200.

[124] P. Ferracioli & al. (no 116).

[125] Ibid.

time, the pressures of funding, limited scrutiny of top officials, and a potential tilt toward less confrontational verdicts reveal the structural hurdles that continue to constrain its capacity to speak truth to power.

The 'Platform Beat' (Global): Scrutinising Digital Platforms

Unlike the examples mentioned above, the 'platform beat' (PB) is not an organisation in a strict sense, nor is it located in any specific nation-State; rather, it is the overarching term for the various platforms and news organisations across the world that scrutinise journalistic performance on major digital platforms such as Facebook and Google.[126]

The largest digital platforms, as some scholars pointed out, often function as 'nation-states in their own right', since they manage their own digital 'territories' in the same way that a government manages its actual state territory.[127] However, unlike traditional governments, these platforms are not frequently held accountable for their governing behaviours, partly because they are not subjected to an electorate, and partly because scrutinising their algorithms demands far greater expertise than many civil society actors possess. This creates a large gap that can be detrimental to democracy.

Consequently, the PB was born. It has two inter-related components. On the one hand, it monitors and scrutinises digital platforms' handling of user data, curating of algorithms, and administering of content moderation systems. On the other hand, it monitors the very content and journalistic behaviours on digital platforms, focusing particularly on fake news and systematic dissemination of hate speech. Together, these tasks allow the organisations under the umbrella term of the PB to hold platforms accountable as distributors of news and information, or as governors of the digital realm.[128]

Because digital governance is a rapidly evolving field, the dynamics between platforms and the PB are constantly shifting, and so the impact of the PB is difficult to measure. However, as Napoli correctly noted, PB is currently the 'most viable and

[126] N. Diakopoulos, 'Algorithmic Accountability Reporting: On the Investigation of Black Boxes' (Report from the Tow Center for Digital Journalism, December 3, 2014) available at https://www.cjr.org/tow_center_reports/algorithmic_accountability_on_the_investigation_of_black_boxes.php (last accessed 15 September 2025).

[127] P. M. Napoli, 'The Platform Beat: Algorithmic Watchdogs in the Disinformation Age' (2021) 36(4) *European Journal of Communication* 377.

[128] Ibid., 378.

important line of defence' against widespread misinformation and political harm, because national government oversight of platforms risk restoration of traditional tyranny, whilst a total lack of oversight risks digital tyranny.[129] Therefore, even if the PB's impact is hard to measure, it still deserves attention and support for the simple reason that it is tangible.

The Implications of Watchdog Journalism

A. The Influence of Watchdog Media

Aided by civil society organisations, watchdog journalist initiatives play an important role in modern liberal democratic life. Thanks to their bottom-up structure and independence, these organisations can reach and mobilise the population and make real and fundamental changes. They function differently in every culture, country, and political context, and they operate according to social circumstances and needs. However, the role of the press as a watchdog has been a consistent marker of stable democracy.

These organisations highlight a broad range of societal issues for the public, sometimes leading to real solutions. Democracy faces many challenges, including lack of transparency in the legislature and a low level of civic participation.[130] The watchdog media aims to draw the attention of society and lawmakers to these problems and to provide decision-makers with appropriate proposals so that they can make decisions and improve the situation in the best interests of the population. Many believe that the media is an indispensable and integral part of political and public life, with a key role in the functioning of democracy.[131] Media can aim to increase the quality, speed, and efficiency of decision-makers work by mobilising the public and making the work of representatives transparent and thus easier to assess.

B. The Role of Digital Platforms: Caveats and Optimised Solutions

Journalism can struggle in the digital age—the rise of echo chambers and the proliferation of misinformation have significantly eroded journalistic integrity. Social media platforms, with their personalised algorithms, often place users in isolated informa-

[129] Ibid., 385.

[130] T. Meyer, *Media Democracy: How the Media Colonize Politics* (Polity Press, 2002); J. Spring, *Conflict of Interests: The Politics of American Education* (McGraw-Hill, 2004); D. Trend, *Cultural Democracy: Politics, Media, New Technology* (State University of New York Press, 1997).

[131] Ibid.

tional bubbles where they are exposed only to views that confirm their existing beliefs.[132] This phenomenon reinforces polarised opinions and creates a fertile ground for misinformation to spread unchecked.

This shift has created a semantic crisis, where words like 'truth', 'integrity', and 'justice' have become politicised and subjective. These terms, once universally understood, are now often reframed as tools to support particular ideological narratives.[133] The rise of terms like 'mainstream media' and 'fake news' only further deepens these divisions, casting legitimate news outlets as enemies of the people and fuelling distrust in the media.

As a result, traditional journalism has increasingly aligned itself with ideologically driven audiences, leading to media outlets becoming less focused on the pursuit of truth and more on reinforcing partisan beliefs. In this context, the role of the journalist has shifted from an independent seeker of truth to, in many cases, an ideological vehicle. This is where watchdog journalism comes in.

With the rise of misinformation, echo chambers, and the proliferation of user-generated content—all phenomena facilitated by the emergence of digital technologies—investigative journalism has become indisputably valuable. At times, misinformation stems from poor fact-checking practices; in other instances, it is deliberately spread by 'malignant actors' aiming to manipulate public opinion.[134] The advent of AI technologies has complicated matters even further. Deepfakes—artificially manipulated images and videos that make it nearly impossible to distinguish between real and fabricated content—pose a significant challenge to verifying the truth and debunking falsehoods. This makes it increasingly difficult for journalists to prove the authenticity of information to the public.

Watchdog journalism, by its very nature, requires a continuous cycle of verification and fact-checking, similar to the operations of reputable 24-hour news outlets. The cases that expose malpractice or wrongdoing typically involve in-depth investigations, requiring comprehensive information-gathering that goes far beyond the superficial coverage often seen in typical news reporting. These investigations not

[132] A. Khan & al., 'The Impact of Social Media Echo Chambers on Semantic Integrity in Journalism: A Study of Language Change and Pragmatic Shift' (2024) 9(3) *Remittances Review* 1384–1409.

[133] Kim Makhlouf, 'Investigative Journalism: A Tool to Fight Misinformation and Fake News' (*Q Views, Northwestern Qatar*, n.d.) available at https://qviews.qatar.northwestern.edu/11204/uncategorized/investigative-journalism-a-tool-to-fight-misinformation-and-fake-news/ (last accessed 15 September 2025).

[134] Ibid.

only uncover the facts but also explore the underlying implications, challenges, and circumstances that enable and facilitate such malpractices and wrongdoing.

Fortunately, there are digital tools to help counter misinformation and disrupt echo chambers. Tools like Google Earth and Google Digital Global, along with Open-Source Intelligence (OSINT), allow journalists to gather information from a wide array of online sources, including search engines and cyberspace, or the deep web. By aggregating and analysing this data, these tools can help address specific 'intelligence requirements'.[135]

Additionally, there is an increasing demand for coding and programming skills to enhance the effectiveness of watchdog journalism in the digital age.[136] Investigative journalists must adapt to new technologies and methods, leveraging them to track down the truth in a landscape rife with digital manipulation and misinformation.

C. The Credibility of Watchdog Journalism (with a Closer Look at the Trump Administration)

A few important points must be made regarding the trustworthiness—or lack thereof—of fact-checking and watchdog organisations. When, if ever, can these organisations be considered reliable? Why should the public trust them? These questions are particularly pressing today, as former President Donald Trump and his administration have repeatedly framed fact-checkers and watchdog groups as biased, unreliable, and even as democratic threats. These accusations have often served as part of a broader strategy to delegitimise political opposition.

From an academic perspective, concerns about the reliability of fact-checking organisations are not new. Scholars have noted that different fact-checkers often arrive at divergent conclusions, thereby undermining confidence in their work as a whole. In one influential study, Marietta & al. demonstrated how three prominent fact-checking organisations—PolitiFact, The Fact Checker, and FactCheck.org—disagreed on issues such as the national debt and racism. Such discrepancies erode trust in fact-checking as a dependable method for evaluating truth claims.[137]

Trust in fact-checking organisations also depends significantly on their perceived credibility. Liu & al. found that expert sources are more likely to be trusted and, as

[135] Ibid.

[136] Ibid.

[137] Sian Lee & al., '"Fact-Checking" Fact Checkers: A Data-Driven Approach' (2023) *Harvard Kennedy School (HKS) Misinformation Review*, available at https://doi.org/10.37016/mr-2020-126 (last accessed 15 September 2025).

a result, are more effective at countering misinformation. For example, warnings issued by watchdog groups or academic institutions tend to be taken more seriously than similar alerts from general media outlets.[138] Moreover, highly credible sources are more successful in persuading individuals to revise false beliefs.[139] The perception that watchdog organisations and universities are relatively independent, less politically entangled, and less likely to intentionally deceive, contributes to their authority.

However, this trust is not uniformly distributed across the political spectrum. Other studies suggest that political affiliation plays a more decisive role than source credibility. Conservatives are generally more sceptical of fact-checkers than liberals, often viewing them as ideologically biased.[140] This implies that while fact-checking organisations and mainstream media maintain a reputation for rigor and factual accuracy, they may fall short in perceived fairness and impartiality.

Accusations of media bias—particularly from conservative commentators—have been ongoing, surfacing as early as 2015 and continuing throughout the Trump presidency.[141] Trump's administration, beginning in 2017, coincided with an explosion in fact-checking efforts. By 2023, the number of fact-checking organisations had grown to over 400 globally.[142] According to *The Washington Post* and other outlets, Trump made over 10,000 false or misleading claims by 2020, averaging 21 per day.[143]

Importantly, Trump's strategy may have been intentionally misleading. As he once put it, 'The Democrats don't matter... The real opposition is the media. And the way to deal with them is to flood the zone with shit'.[144] This tactic—overwhelming media outlets with falsehoods—forces journalists to spend time and resources on verification, ironically giving the misinformation more airtime through repetition.

Despite the rise of fact-checking, its effectiveness in changing public opinion remains questionable. In Trump's case, a substantial portion of his base continues to support him—either because they do not perceive his claims as false or because they

[138] Liu & al. (no 27).

[139] Ibid.

[140] Ibid.

[141] Ibid.

[142] Leon Barkho, 'A Critical Inquiry into US Media's Fact-Checking and Compendiums of Donald Trump's Falsehoods and "Lies"' in A. Akande (ed.), *The Perils of Populism* (Springer Nature Switzerland, 2023) 259–278.

[143] Ibid.

[144] Ibid.

are unbothered by his lack of credibility.[145] Arguably, Trump's sustained attacks on fact-checkers have helped to erode public trust in these organisations altogether, contributing to growing scepticism toward both watchdog journalism and institutional fact-checking.[146]

D. An Overview of Watchdog Journalism in Regimes Experiencing Autocracy or Self-Censorship

Different questions arise when the role of watchdog journalism and watchdog media is explored within authoritarian regimes. With State manipulation and censorship rife in autocratic countries and competitive autocracies, exposing wrongdoing and holding those responsible accountable becomes not only difficult but also life-threatening. To better understand the dynamics of censorship, this section briefly sketches the historical development of restrictions on the printed word before turning to their impact on watchdog journalism in contemporary times.

The rise of journalism prompted authoritarian States to employ increasingly sophisticated systems and mechanisms of censorship. The first organised effort to censor the press was initiated by the Catholic Church in the 16th century. This was followed by absolutist monarchies across Europe seeking to regulate the periodical press from the 17th century onward.[147] However, thanks to persistent activism against censorship, such systems were largely dismantled in Europe by the end of the 19th century. Yet the 20th century marked a major setback for press freedom with the emergence of totalitarian regimes after World War I. Fascist dictatorships in Germany, Italy, and Spain, communist States in Eastern Europe, and the apartheid regime in South Africa all imposed far-reaching restrictions on the press.[148]

In authoritarian regimes, the press is expected to be subservient to the State, serving as a mouthpiece for political ideologies while avoiding independent analysis or criticism. For instance, in the Soviet Union, the *Glavlit* was created in 1922 to oversee

[145] Briony Swire & al., 'Processing political misinformation: comprehending the Trump phenomenon' (2017) 4(3) *Royal Society Open Science* 160802.

[146] The Washington Post, 'Trump Fact-Checks Interviews and Debates' (14 October 2024) available at https://www.washingtonpost.com/politics/2024/10/14/trump-fact-checks-interviews-debates/ (last accessed 15 September 2025).

[147] Marius Dragomir, 'Authoritarian Societies and Journalism' (23 December 2019) *Oxford Research Encyclopedia of Communication*, available at https://doi.org/10.1093/acrefore/9780190228613.013.783 (last accessed 15 September 2025).

[148] Carl J. Friedrich & Zbigniew K. Brzezinski, *Totalitarian Dictatorship and Autocracy* (Harvard University Press, 1965) 165–167.

media-related matters and enforce propaganda dissemination. Similarly, Nazi journalists operated under strict rules and regulations imposed by the regime.[149] Such governments often imposed extreme penalties on journalists who deviated from the official line. Punishments ranged from execution and confinement in concentration camps during Stalin's harshest years to imprisonment in later decades. Despite such life-threatening risks, resistance movements emerged. Some Soviet and Polish journalists dared to speak the truth, challenging the State monopoly over information.[150]

The Soviet censorship model also influenced other States during the Cold War. China developed its own comprehensive system of media control, while Cuba nationalised television and radio under similar principles. Vietnam, North Korea, and Cambodia also adopted restrictive frameworks inspired by Soviet practices.[151] In such communist ideocracies, media is heavily centralised and owned by the State. In several Central Asian regimes, efforts are made to insulate the population from foreign media influence, while domestic journalism remains tightly controlled. In the case of China, for example, journalists are expected to avoid topics that are explicitly prohibited by State authorities, reflecting the media's role as a tool for political control rather than public accountability.[152]

In non-ideocratic autocracies, such as sultanistic regimes, elite enrichment and endemic corruption are central characteristics. Media outlets are frequently State-owned and lack institutional independence or alternative sources of funding.[153] The presence of a robust and objective media system capable of scrutinising power structures is typically considered undesirable, as it threatens to expose the systemic rot underpinning governance.

In pseudo-democratic or electoral autocracies, such as Venezuela, limited media freedom may be tolerated to sustain the appearance of democratic practice and satisfy a portion of the electorate. However, this permitted space is often carefully managed, with most media outlets remaining sympathetic to—or controlled by—the

[149] J. A. Cheibub, J. Gandhi & J. R. Vreeland, 'Democracy and dictatorship revisited' (2010) 143 *Public Choice* 67–101.

[150] Axel Hadenius & Jan Teorell, 'Pathways from Authoritarianism' (2007) 18(1) *Journal of Democracy* 143–156.

[151] Ibid.

[152] Sebastian Stier, 'Democracy, autocracy and the news: the impact of regime type on media freedom' (2015) 22(7) *Democratization* 1273.

[153] G. Egorov, S. Guriev & K. Sonin, 'Why Resource-poor Dictators Allow Freer Media: A Theory and Evidence from Panel Data' (2009) 103(4) *American Political Science Review* 645–668; Peter Lorentzen, 'China's Strategic Censorship' (2014) 58(2) *American Journal of Political Science* 402–414.

dominant ruling party.[154] Investigative journalism that challenges State narratives or reveals misconduct is thus discouraged or actively suppressed.

Even in one-party autocratic regimes, where some space for investigative journalism might be allowed—often in the interest of improving economic performance—media freedom is highly contingent.[155] Such liberties tend to contract during times of economic crisis or political instability, reaffirming the instrumental view of media within these regimes.

Japan is an interesting case as it provides a curious example of self-censorship even though journalists generally do not fear harsh retribution such as imprisonment or assassination.[156] Officially, the Japanese Constitution prohibits censorship, upholds democratic principles, and guarantees freedom of the press. Any overt violations of these rights often provoke public protest.[157] Nevertheless, self-censorship among journalists remains pervasive. In the 2024 World Press Freedom Index, Japan ranked 70th—last among the G7 nations.[158] Additionally, in 2017, UN Special Rapporteur David Kaye raised concerns about the *kisha* clubs, exclusive press clubs that reinforce close ties between media organisations and government agencies. These clubs grant privileged access to official information, but in return, journalists are expected to provide favourable coverage. As a result, *kisha* clubs obstruct investigative journalism and prevent the public from accessing information that authorities prefer to suppress.[159]

At the same time, some autocracies tolerate a limited degree of press freedom. Allowing constrained watchdog activity can improve bureaucratic accountability and transparency at minimal cost. Independent journalism, in this sense, is less expensive than maintaining an extensive secret police apparatus. This practice, sometimes referred to as 'strategic censorship',[160] preserves a facade of legitimacy. By granting the appearance of increased media freedom, regimes can placate both domestic and international critics while using limited transparency to mitigate corruption.

[154] E. E. Bracho-Polanco, 'Journalism practice, media and democracy in Venezuela (2000-2010)' (Doctoral dissertation, University of Westminster, 2014).

[155] Ibid.

[156] Jeff Kingston, 'Watchdog Journalism in Japan Rebounds but Still Compromised' (2018) 77(4) *The Journal of Asian Studies* 881, 893.

[157] Ibid.

[158] Yosuke Buchmeier, 'Media politics in Japan: News journalism between interdependence, integrity, and influence' (2024) 69 *Publizistik* 495–526.

[159] Kingston (no 156).

[160] Stier (no 152).

Recommendations to Strengthen and Support Watchdog Journalism

Watchdog organisations, as we can see, play a key role in the effective functioning of modern liberal democracies.[161] The existence of watchdog organisations is an important indicator of the strength of democracies because they ensure that freedom of expression and opinion and the rights that go with it are respected. Through their work, these organisations ensure civil society participation in political economy and social life by providing a platform for the flow of information and by raising public awareness on issues of social interest. Without these organisations, transparency in political and economic life would be a major challenge, and corruption and abuse of power would quickly flourish. These problems make it impossible for democracies to function properly and for the principles of the rule of law to be implemented, leading to a loss of trust in the system and a loss of motivation on the part of the citizens to participate actively in social and political life. Among the watchdog organisations, investigative journalism organisations and initiatives play a particularly important role, as they aim to shed light on political, economic, and social developments, especially those that would otherwise go undetected and undisclosed. The role of investigative journalists and whistleblowers is to warn the public and to ensure that democracy is protected.[162]

Why Strengthen Watchdog Organisations?

Watchdog journalism organisations have an important role to play in the proper functioning of democracies, as they aim to connect the public with political life, making the latter open and transparent, giving people the opportunity to hold decision-makers and all actors in political (and economic) life to account. It also provides an opportunity to monitor the political process and, through positive pressure, encourages decision-makers to do their job more effectively.

These organisations, of course, operate differently in different political systems and social contexts, and in many cases they take serious risks, since their work is often viewed unfavourably by political and economic actors whose status and power are threatened by investigative activity, ranging from personal intimidation and har-

[161] W. Donsbach & T. E. Patterson, 'Comparing Political Communication: Theories, Cases, and Challenges' in F. Esser & B. Pfetsch (eds), *Comparing Political Communication: Theories, Cases, and Challenges* (Cambridge University Press, 2004) 251, 270.

[162] James H. Dygert, *The Investigative Journalist* (Prentice Hall, 1976).

assment to imprisonment and killing. Investigative journalists are therefore in some cases putting themselves in peril. Even when not outright dangerous, their work can be challenging—information is often difficult to access, and support and cooperation on the subject under investigation are not always readily given.

The importance of watchdog journalism is reflected in public attitudes around the world. US surveys show the importance of media watchdogs to the public, although there are noticeable differences in opinion by party affiliation as to the extent to which the media should act as watchdogs and how well they perform this role. There is, nevertheless, a consensus that the majority considers the media's watchdog role to be one of paramount importance.[163] Research in the UK has shown similar results.[164]

Watchdog journalism is thus essential to strengthening accountability in all modern liberal democracies so that they can operate smoothly, effectively, and transparently, and so that the public can play a more active role across aspects and levels of political, economic, and social life. It is also important that they work to ensure that people have confidence in their elected leaders, rather than feel disillusioned towards the political system. They also promote the most important feature of participatory democracy by providing information to the public and thereby activating society to increase the proportion of people participating in political life and the decision-making process.

It is particularly important to strengthen and deepen the role of watchdog journalism as there are many social, economic, and political problems on which society needs to be consistently informed. The active involvement of the population and the exertion of pressure on decision-making also plays a major role in problem-solving, but this requires reliable and unbiased information, deep investigations, and critical analysis. However, to strengthen watchdog journalism, it is necessary to be aware of a country's political culture and traditions and the needs of its communities.

Soft Politics: Transforming Political Culture

To strengthen watchdog journalism, it would be first worthwhile to assess the needs of the population and what issues would be worth focusing on for investigative journalists. Such an assessment should, of course, be carried out independently of any

[163] Mark Jurkowitz & Amy Mitchell, 'Most say journalists should be watchdogs, but views of how well they fill this role vary by party, media diet' (Pew Research Center, 26 February 2020) available at https://www.pewresearch.org/journalism/2020/02/26/most-say-journalists-should-be-watchdogs-but-views-of-how-well-they-fill-this-role-vary-by-party-media-diet/ (last accessed 15 September 2025).

[164] Kowalewski (no 6).

political party. Conducting such a survey would probably be challenging due to political differences and the status quo.

Perhaps a significant proportion of society are very ill-informed about the political process and, in many cases, see no possibility of having a say in the decision-making—meaning many consider their participation useless. However, this mentality is harmful and weakens the most important elements of democracy: participation and the representation of interests. Unfortunately, this in turn leads to a lack of pressure on policymakers and a tendency for them to abuse their power and not represent the interests of the people who have elected them. Strengthening watchdog journalism, then, could be coupled with introducing civic education and active citizenship in schools, so that the population, from an early age, understands their role in society, the political process, and ways to assert their rights and demands. If the impact of watchdog journalism on democracies was successfully demonstrated and the education of young people for active social and political participation was effectively implemented in schools, the political culture of the population could undergo an eventual transformation that would radically change the attitudes and mentality of the people towards public affairs.

As watchdog journalism has a strong social role, this could be further enhanced by educating people to take an active role. In a sense, watchdog journalism needs to create a demand for itself by providing the public with a steady flow of useful, accurate, and independent information and by encouraging the public to take action to bring about concrete changes.

Watchdog journalism's mission is to show that bottom-up initiatives can be effective in bringing about change in the political, social, and economic spheres and that reliable, credible, and useful information is not just wanted, but needed. These organisations show that it is possible to achieve through active engagement, even if you are not a top-down organisation.

The promotion of watchdog journalism could be further strengthened by the organisation of cultural, political, economic, and other social forums where well-developed proposals could be discussed in public and put to a vote by decision-makers. These events could then be extended to include workshops related to these topics, especially for young people, to help them learn about the political decision-making process and how to participate in it.

It would also be worthwhile to create a forum or platform where watchdog organisations and other initiatives could summarise their activities, cooperate, and present

their results and work to the public. Such a forum or platform could also be extended to an international level in terms of the exchange of experience and further development of these initiatives and cooperation in specific areas.

The role of watchdog organisations could be further enhanced by the creation of monitoring committees to oversee the work of watchdog organisations and to enforce their proper functioning through accountability. These monitoring committees would continue to investigate and inform the public and would be empowered by the people directly to take action against abuse and corruption.

Of course, both sides should promote watchdog journalism and strengthen cooperation between these organisations to increase interdependence between the parties. However, as previously demonstrated through the examples of FactCheck.org and Lupa, watchdog organisations are more often subjected by, instead of strengthened through, the political process. Bolleyer's recent comparative study of legal regulations of civil society organisations in democracies suggests that, even in 'long-lasting democracies', countries develop divergent regulatory frameworks under the influence of different structural constraints and incentives, which watchdog organisations themselves are unable to alter.[165] Therefore, we must turn to international legal frameworks for further possibilities on strengthening watchdog organisations.

Hard politics: Creating and Enforcing the Legal Framework

For the above change to take place and for watchdog journalism to be strengthened and to function well, legislative changes are needed in addition to the tools of soft politics. Organisations, and the journalists who work for them, need better protection than they are currently granted.

International legal rules and principles can be adapted to strengthen watchdog journalism on a national level. These rules and principles are perhaps even more important outside the EU, as MS are already under an obligation to follow European legal frameworks. Beyond EU legal frameworks, international treaties to which the State is a party are also relevant.

As stated above, EU MS must adopt a number of international documents and fulfil national obligations in meeting the principles of freedom of expression and freedom of the press, fundamental, inalienable human rights that are enshrined in several important international documents; they must also comply with a host of other

[165] N. Bolleyer, *The State and Civil Society: Regulating Interest Groups, Parties, and Public Benefit Organizations in Contemporary Democracies* (1st edn, Oxford University Press, 2018).

related legal principles and treaties that form an essential part of the international legal system. In the EU, the most relevant are the Treaty on European Union, the Treaty on the Functioning of the European Union, the Charter of the European Union, the European Convention on Human Rights, the Universal Declaration of European Rights, the European Charter of Freedom of the Press, and the Venice Commission's list of criteria for the rule of law. These documents lay down the rules and principles of membership in the EU and shape its political culture, values, and norms. They are essential for the existence, functioning, and strengthening of watchdog journalism, as they ensure the international protection of human rights, which protects society and organisations against violations at the nation-State level and affirms the enforceability and realisation of freedom of expression, freedom of the press, and other relevant fundamental human rights and ensures that the principles of the rule of law are upheld.

The Treaty on the European Union (which lists the most fundamental principles of the EU) mentions the protection of human rights and the recognition of freedoms in the preamble. In the second article, respect for human rights and democracy as a fundamental value are also included among the principles and values of the Union:

> The Union is founded on the values of respect for human dignity, freedom, democracy, equality, the rule of law, and respect for human rights, including the rights of persons belonging to minorities. These values are common to the Member States in a society in which pluralism, non-discrimination, tolerance, justice, solidarity, and equality between women and men prevail.[166]

The Treaty on the Functioning of the European Union also contains numerous references to the protection of human rights and fundamental freedoms and related principles. MS must also comply with the Charter of Fundamental Rights of the European Union, which contains numerous references to the protection and promotion of fundamental human rights. The document includes in its preamble the fundamental and inalienable rights of citizens of the EU. Article 11 of the Charter is about freedom of expression and information, which includes two sub-articles:

> 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.

[166] Consolidated Version of the Treaty on the Functioning of the European Union (Treaty of Lisbon) C 326/01 available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:C2012/326/01 (last accessed 15 September 2025).

2. The freedom and pluralism of the media shall be respected.[167]

These are two points that MS must respect, and which they must guarantee for their citizens in their own national laws. Article 10 of the European Convention on Human Rights speaks of freedom of expression, stating that everyone has the right to hold and express opinions freely, to receive information and ideas freely, and to have them freely communicated without interference of any kind.[168] Article 19 of the Universal Declaration of European Rights also enshrines the principle that everyone has the inalienable right to freedom of expression, which is linked to the right not to be discriminated against or harassed for their freedom of opinion and expression, and to seek, receive, and impart information and ideas without interference or obstacles in any form.[169]

The European Charter on Freedom of the Press summarises the relevant documents on freedom of expression and freedom of the press adopted by MS, and which form the EU's system of values and standards on the subject.[170] In the Venice Commission's opinion papers on the criteria for the rule of law and on freedom of expression and the freedom to receive and impart information, we can also see how important a role freedom of expression plays in the functioning of modern democracies and how its violation is condemned by the international community.

These documents enshrine the EU's core values and emphasise that freedom of the mass media is essential for modern democracies, because true democracy cannot function without a free media to hold governments and powerful individuals accountable. Free media is seen as an essential pillar for the implementation of checks and balances that underpin the democratic order.

The European Commission's rule of law report also highlights free media as a pillar of the rule of law, addressing issues such as the independence of media regulators, transparency of media ownership, transparency and fairness in the distribution of public advertising, the safety of journalists, and access to information. To ensure freedom of the mass media, the EU has legal instruments such as the Directive against

[167] EU Charter of Fundamental Rights, available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT (last accessed 15 September 2025).

[168] European Convention on Human Rights, available at https://www.echr.coe.int/d/convention_eng (last accessed 16 September 2025).

[169] United Nations, Universal Declaration of Human Rights, available at https://www.un.org/en/about-us/universal-declaration-of-human-rights (last accessed 16 September 2025).

[170] PressFreedom.eu, available at https://www.pressfreedom.eu/en/index.php (last accessed 16 September 2025).

high-profile lawsuits that hinder public participation and the European Freedom of the Media Act.[171] The Directive, proposed in April 2022, protects public actors from possible unfair or abusive legal proceedings with various safeguards such as financial compensation, fast-track procedures, and protection against third-party judgments. The Directive has also been supplemented by proposals for MS to extend the safeguards beyond domestic and civil matters and to organise awareness-raising campaigns, judicial training, and aggregate data collection at the international level. The European Media Freedom Act, also proposed by the European Commission in 2022, builds on the Audiovisual Media Services Review Directive and includes mechanisms and rules to promote the independence and diversity of the media in the EU.[172]

The measures protect the independence of experts as well as the rights of media service providers, including effective protection of workers and their families, and safeguard against the use of spyware against media organs, journalists and their families. They also provide for adequate, stable, and transparent financing of public service media, and ensure the diversity of the mass media. The requirement is the assessment of the impact of media market concentration on the diversity of media outlets and editorial independence, transparent and non-discriminatory public advertising, safeguards against unjustified removal of online media content and a proposal for a new independent European Media Regulators Board to replace the European Audiovisual Media Regulators Group. The requirement also includes a proposal to facilitate the effective and consistent implementation of the Directive by strengthening cooperation between national regulators and national bodies.

Support for media freedom is reinforced through initiatives such as the Digital Services Package, the Media and Audiovisual Action Plan, and the Action Plan for European Democracies. These include measures to ensure freedom of the media, such as a) the introduction of a media ownership monitoring system to create a country-by-country database of media ownership information; b) the provision of grants to support innovation in local and regional media and to boost media diversity; c) a rapid response mechanism that can provide practical assistance to journalists in distress; d) an emergency support fund for investigative journalists and media organisations, and e) the implementation of monitoring of media pluralism to identify

[171] European Commission, 'European Media Freedom Act' available at https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/new-push-european-democracy/protecting-democracy/european-media-freedom-act_en (last accessed 16 September 2025).

[172] European Commission (no 171).

potential risks to media pluralism.[173] These laws, directives, proposals, and strategies can be applied at the national level to strengthen and promote watchdog journalism.

Beyond the EU, the Universal Declaration of Human Rights, promoted by the UN, states in its Article 19 that everyone has the right to freedom of expression and the right to receive, seek, and impart information.[174]

The principles and laws can only achieve their purpose if they are enforced at the State level. Another way to strengthen watchdog journalism is to make changes within the State itself. One example is the EU initiative 'Action against abusive litigation (SLAPP) targeting journalists and rights defenders',[175] which aims to protect journalists and human rights defenders in strategic lawsuits against public participation. SLAPPs are lawsuits to prevent people from informing the public and reporting on matters of public interest. The fact that this initiative was approved by the European Parliament (in February 2024) with an overwhelming majority proves that governments recognise the importance of journalism, and watchdog journalism, within democratic societies.[176]

In many countries, journalists are not in danger of receiving death threats, but they still face serious challenges that make their work extremely difficult, such as chronic understaffing, underpayment, and difficulty in accessing information. Underpayment makes the profession less attractive to young people, demotivates those already in the profession, and poses a serious livelihood issue.[177] A further problem

[173] European Media and Audiovisual Action Plan, available at https://digital-strategy.ec.europa.eu/en/policies/media-and-audiovisual-action-plan (last accessed 16 September 2025); European Commission, 'European Democracy Action Plan: making EU democracies stronger' (3 December 2020) available at https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2250 (last accessed 16 September 2025); Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Europe's Media in the Digital Decade: An Action Plan to Support Recovery and Transformation COM(2020)-784, 3 December 2020 available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0784 (last accessed 16 September 2025).

[174] United Nations (no 169).

[175] European Commission, 'EU Action Against Abusive Litigation (SLAPP) Targeting Journalists and Rights Defenders' available at https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13192-EU-action-against-abusive-litigation-SLAPP-targeting-journalists-and-rights-defenders_en (last accessed 16 September 2025).

[176] Nathalie Weatherald, 'Anti-SLAPP Law Gets Final Seal of Approval from EU Parliament' (Euractiv, 27 February 2024) available at https://www.euractiv.com/section/media/news/anti-slapp-law-gets-final-seal-of-approval-from-eu-parliament/ (last accessed 16 September 2025).

[177] B. Houston, 'The future of investigative journalism' (2010) 139(2) *Daedalus* 45–56; H. Kurtz, *Media circus: The trouble with America's newspapers* (Three Rivers Press, 1994); G. Lanosga & B. Houston,

is that journalists risk losing their jobs, thus their livelihoods, if their investigations uncover information of special interest groups. Work needs to be done so that the profession is more attractive, and journalists can enjoy the financial independence that is essential for their work. The situation would be much improved if journalists did not fear losing their jobs. It would also help if information were more accessible and organisations in the public and private sectors were more open and transparent in their dealings, working in partnership with journalists, not against them. These objectives should be promoted through a range of legislation with a sound legal basis.

As previously mentioned, with consistent technological advancements, international bodies should improve multilateral treaties and frameworks concerning the governance of digital space. Currently, both digital platforms and the PB function as political processes ungoverned by conventional political borders and often beyond conventional rules. This is unsustainable. Therefore, lawmakers should consider empowering PB through frameworks such as the GDPR or consult and collaborate with PB organisations to hold powerful actors in the digital realm to account.

D. Final Remarks

For watchdog journalism to be strengthened, countries must have comprehensive legal protection and freedom of the press. Safeguards include strong laws to protect freedom of expression and press freedom, as well as whistleblowers and journalists, and anti-SLAPP legislation.[178] Investigative journalism needs to be further encouraged by circumscribing defamation laws and preventing intimidation and threats against journalists so that investigative reporters can report without fear of negative consequences. The legal framework can also help to improve access to information, thus strengthening investigative work. Access can be facilitated by transparency laws that allow journalists to obtain government and corporate documents and data essential for investigative reporting. Ease of access can be further enhanced by open data initiatives providing journalists with structured data for analysis and reporting.

'Spotlight: Journalists assess investigative reporting and its status in society' (2017) 11(9) *Journalism Practice* 1101–1120; K. Li & C. Sparks, 'Chinese newspapers and investigative reporting in the new media age' (2018) 19(3) *Journalism Studies* 415–431.

[178] European Parliament, 'Anti-SLAPP: EU protection against legal actions that silence critical voices' (27 June 2023) available at https://www.europarl.eu.europa.eu/news/en/press-room/20230626IPR00818/anti-slapp-eu-protection-against-legal-actions-that-silence-critical-voices (last accessed 16 September 2025).

In addition to legal support, economic incentives are also essential. Grants and resources should be made available for investigative journalism projects, initiatives, subsidies, and programmes. This will encourage partnerships and collaborations between media organisations, NGOs, and government agencies to carry out investigative reporting. There is a need to diversify the sources of revenue to allow investigative work to continue independently. Diversification of revenue streams allows media organisations to become less dependent on advertisements and other external actors by including subscriptions, membership fees, crowdfunding, and other subsidies. Financial support promotes sustainable and quality work and ensures financial independence for journalists, which improves the credibility and quality of their work. It is also essential that there is continuous follow-up, and that capacity is expanded. This requires programmes providing training material for journalism students and introducing investigative techniques, data analysis, and digital security and ethical reporting practices. These training programmes would help prepare the next generation of journalists. In addition, workshops and seminars can encourage cooperation between journalists, experts, and civil society organisations, which would further strengthen watchdog journalism globally and promote the tackling of local issues and common problems.

However, while the recommendations proposed to strengthen a culture of investigative journalism may be broadly applicable, their feasibility must be critically examined, particularly in the context of autocratic regimes. In such environments, independent watchdog media are often absent, play only a marginal role, or are permitted to operate solely insofar as they align with the interests of the ruling authorities.

The arrival of the internet and social media has revolutionised what once seemed like impenetrable systems of control. Digital platforms have offered journalists new opportunities to reveal secrets and expose wrongdoing. This is especially vital in authoritarian regimes, where strict State-imposed censorship distorts the truth and forces the media to toe the party line. Social media platforms, with their vast reach and decentralised nature, allow sensitive information and inconvenient truths to occasionally slip through the cracks of censorship.[179]

In countries where institutionalised watchdog media organisations are absent or heavily restricted, the rise of digital technology has led to individual actors increasingly assuming the role of the watchdog. Activists, citizen journalists, and independent bloggers can bypass State-controlled narratives and offer alternative views, often

[179] Dragomir (no 147).

at great personal risk. This form of grassroots journalism, while imperfect and occasionally fragmented, helps to ensure a measure of accountability in places where traditional, independent journalism is either non-existent or severely compromised.

This enables both local populations and the international community to access critical insights that would otherwise remain suppressed. In China, for example, online spaces provide channels for journalists to engage with audiences and publish sensitive material overlooked by the mainstream press. Once reported widely on social media, such stories often gain traction in traditional outlets as well.[180]

However, authoritarian regimes have responded quickly to these challenges. In Kazakhstan, the State maintains control over radio and television, which remain the most consumed forms of media, insulated from foreign influence. More recently, the government has extended its presence into online platforms. Azerbaijan provides another striking example—the State not only dominates broadcast media but also uses digital platforms to target political opposition, human rights activists, and what little remains of independent journalism.[181]

Overall, a potentially fruitful avenue for strengthening investigative journalism in autocratic or pseudo-democratic settings lies in the affordances of digital technology. Digital platforms enable the publication of sensitive information by individuals or informal networks, circumventing traditional institutional media structures. However, the effectiveness of this alternative rests heavily on the degree of autonomy retained by technology companies, the presence of meaningful regulation, and the extent to which these platforms are insulated from State capture.

References

Amaral L., 'Between Personal and Public Interests: A Look Back at the Impact of Snowden and WikiLeaks (Guest Blog)', *LSE Polis* (22 July 2014) available at https://blogs.lse.ac.uk/polis/2014/07/22/between-personal-and-public-interests-a-look-back-at-the-impact-of-snowden-and-wikileaks-guest-blog-polissummer/ (last accessed 15 September 2025).

Amazeen M. A., 'Journalistic Interventions: The Structural Factors Affecting the Global Emergence of Fact-Checking' (2017) 21(1) *Journalism* 95–111.

Barkho L., 'A Critical Inquiry into US Media's Fact-Checking and Compendiums of Donald Trump's Falsehoods and "Lies"' in A. Akande (ed.), *The Perils of Populism* (Springer Nature Switzerland, 2023) 259–278.

[180] Ibid.

[181] Dragomir (no 147).

BBC News, 'Hopewell Chin'ono: Whistle-blowing Zimbabwean journalist arrested' (20 July 2020) available at https://www.bbc.co.uk/news/world-africa-53477423 (last accessed 11 February 2026).

Bennett W. L. & William Serrin, 'The Watchdog Role' in G. Overholser & K. H. Jamieson (eds), *The Press* (Oxford University Press, 2005) 170–188.

Bentham J., 'Political Tactics' in Michael James, Cyprian Blamires & Catherine Pease-Watkin (eds), *Political Tactics* (Oxford University Press, 1999) 267.

BEUC, 'Better Regulation Watchdog Launches Website' available at https://www.beuc.eu/news/better-regulation-watchdog-launches-website (last accessed 15 September 2025).

Biancovilli P., L. Makszin & C. Jurberg, 'Misinformation on Social Networks During the Novel Coronavirus Pandemic: A Quali-Quantitative Case Study of Brazil' (2021) 21(1) *BMC Public Health* 1200.

Blusiewicz J., 'The Case of Edward Snowden: A Different Path' (2014) 8(1) *Cornell International Affairs Review*.

Bolleyer N., *The State and Civil Society: Regulating Interest Groups, Parties, and Public Benefit Organizations in Contemporary Democracies* (1st edn, Oxford University Press, 2018).

Bracho-Polanco E. E., 'Journalism practice, media and democracy in Venezuela (2000-2010)' (Doctoral dissertation, University of Westminster, 2014).

Brevini B., 'WikiLeaks: Between Disclosure and Whistle-blowing in Digital Times' (2017) 11(3) *Sociology Compass*.

Buchmeier Y., 'Media politics in Japan: News journalism between interdependence, integrity, and influence' (2024) 69 *Publizistik* 495–526.

Carson A., 'The Political Economy of the Print Media and the Decline of Corporate Investigative Journalism in Australia' (2014) 49(4) *Australian Journal of Political Science* 726–742.

Carson A., *Investigative Journalism, Democracy and the Digital Age* (1st edn, Routledge, 2019).

Cheibub, J. Gandhi & J. R. Vreeland, 'Democracy and dictatorship revisited' (2010) 143 *Public Choice* 67–101.

Committee to Protect Journalists, 'Journalist Hopewell Chin'ono Arrested, Charged with Incitement in Zimbabwe' (20 July 2020) available at https://cpj.org/2020/07/journalist-hopewell-chinono-arrested-charged-with-incitement-in-zimbabwe/ (last accessed 11 February 2026).

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: Europe's Media in the Digital Decade: An Action Plan to Support Recovery and Transformation

COM(2020)-784 (3 December 2020) available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0784 (las accessed 16 September 2025).

Consolidated Version of the Treaty on the Functioning of the European Union (Treaty of Lisbon) C 326/01 available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex-:C2012/326/01 (last accessed 15 September 2025).

Diakopoulos N., 'Algorithmic Accountability Reporting: On the Investigation of Black Boxes' (Report from the Tow Center for Digital Journalism, December 3, 2014) available at https://www.cjr.org/tow_center_reports/algorithmic_accountability_on_the_investigation_of_black_boxes.php (last accessed 15 September 2025).

Djankov S. & al., 'Who Owns the Media?' (Working paper, Harvard University, 2002).

Donsbach W. & T. E. Patterson, 'Comparing Political Communication: Theories, Cases, and Challenges' in F. Esser & B. Pfetsch (eds), *Comparing Political Communication: Theories, Cases, and Challenges* (Cambridge University Press, 2004) 251–270.

Dragomir M., 'Authoritarian Societies and Journalism' (23 December 2019) *Oxford Research Encyclopedia of Communication*, available at https://doi.org/10.1093/acrefore/9780190228613.013.783 (last accessed 15 September 2025).

Drüeke R., 'Medien, Öffentlichkeit und Demokratie: Zur Watchdog-Funktion von Medien' ['Media, the Public and Democracy: On the Watchdog Function of the Media'] (2018) 31(3) *Forschungsjournal Soziale Bewegungen* 19–28.

Dutton, William H., 'The Fifth Estate Emerging through the Network of Networks' (2009) 27(1) *Prometheus* 49.

Dyck A. & L. Zingales, 'The Bubble and the Media' in P. K. Cornelius & B. Kogut (eds), *Corporate Governance and Capital Flows in a Global Economy* (Oxford University Press 2002) 83–102.

Dyck A. & L. Zingales, 'Media and Asset Prices' (Working paper, Harvard University, 2003).

Dygert J. H., *The Investigative Journalist* (Prentice Hall, 1976).

Egorov G., S. Guriev & K. Sonin, 'Why Resource-poor Dictators Allow Freer Media: A Theory and Evidence from Panel Data' (2009) 103(4) *American Political Science Review* 645–668.

Eisenman S., 'Watchdog Journalism: Function and Future' (Columbia University, 1996) available at https://www.ee.columbia.edu/~shane/words/watchdog.htm (last accessed 16 February 2024).

Encyclopaedia Britannica, 'Journalism | Definition, History, & Facts' (2024), available at https://www.britannica.com/topic/journalism (last accessed 15 February 2024).

EU Charter of Fundamental Rights, available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:12012P/TXT (last accessed 15 September 2025).

European Anti-Fraud Office (OLAF), 'OLAF Home', available at https://anti-fraud.ec.europa.eu/index_en (last accessed 15 September 2025).

European Commission, 'EU Action Against Abusive Litigation (SLAPP) Targeting Journalists and Rights Defenders', available at https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/13192-EU-action-against-abusive-litigation-SLAPP-targeting-journalists-and-rights-defenders_en (last accessed 16 September 2025).

European Commission, 'European Media Freedom Act' available at https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/new-push-european-democracy/protecting-democracy/european-media-freedom-act_en (last accessed 16 September 2025).

European Commission, 'European Democracy Action Plan: making EU democracies stronger' (3 December 2020) available at https://ec.europa.eu/commission/presscorner/detail/en/ip_20_2250 (last accessed 16 September 2025).

European Convention on Human Rights, available at https://www.echr.coe.int/d/convention_eng (last accessed 16 September 2025).

European Digital Rights (EDRi), 'European Digital Rights (EDRi)' (27 September 2023) available at https://edri.org/ (last accessed 15 September 2025).

European Media and Audiovisual Action Plan (n.d.) https://digital-strategy.ec.europa.eu/en/policies/media-and-audiovisual-action-plan accessed 16 September 2025.

European Parliament, 'Anti-SLAPP: EU protection against legal actions that silence critical voices' (27 June 2023) available at https://www.europarl.eu.europa.eu/news/en/press-room/20230626IPR00818/anti-slapp-eu-protection-against-legal-actions-that-silence-critical-voices (last accessed 16 September 2025).

European Parliamentary Research Service, *The Role of Investigative Journalism to Uncover Fraud and Corruption in Europe (June 2022)* available at https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/747083/IPOL_BRI(2023)747083_EN.pdf (last accessed 15 February 2024).

FactCheck.org, 'Our Mission', available at https://www.factcheck.org/about/our-mission/ (last accessed 15 September 2025).

Fałkowski J. & P. J. Kurek, 'The Power of Social Mobilisation: The Impact of Monitoring the 2015 Presidential Elections in Poland' (2021) 49(1) *Journal of Comparative Economics* 38.

Fancher M. R., 'The 21st century journalist's creed' (2009) 63(3) *Nieman Reports* 35–38.

Fengler Susanne, 'Holding the News Media Accountable: A Study of Media Reporters and Media Critics in the United States' (2003) 80(4) *Journalism & Mass Communication Quarterly* 818–832.

Ferracioli P. & al., 'The Watchdog Role of Fact-Checkers in Different Media Systems' (2022) 10(5) *Digital Journalism* 717–737.

Francke W, 'The Evolving Watchdog: The Media's Role in Government Ethics' (1995) 537 *The Annals of the American Academy of Political and Social Science* 109–121.

Friedrich C. J. & Z. K. Brzezinski, *Totalitarian Dictatorship and Autocracy* (Harvard University Press, 1965) 165–167.

Gershon L., 'When Did the Media Become a "Watchdog?"' *JSTOR Daily* (2018) available at https://daily.jstor.org/when-did-the-media-become-a-watchdog/ (last accessed 16 February 2024).

Grenier J. A. Jr., 'The Origins and Nature of Progressive Muckraking' (PhD dissertation, University of California-Los Angeles, 1965).

Hadenius Axel & Jan Teorell, 'Pathways from Authoritarianism' (2007) 18(1) *Journal of Democracy* 143–156.

Hallin D., 'The Media, the War in Vietnam, and Political Support: A Critique of the Thesis of an Oppositional Media' (1984) 46(1) *The Journal of Politics* 2–23.

Hallin D. & P. Mancini, *Comparing Media Systems: Three Models of Media and Politics* (Cambridge University Press, 2004).

Hampton M., 'The Fourth Estate Ideal in Journalism History' in Stuart Allan (ed.), *The Routledge Companion to News and Journalism* (Routledge 2009) 3–12.

Harry Ransom Center, 'Inventory of the Woodward and Bernstein Watergate Papers' (The University of Texas at Austin, n.d.) available at https://norman.hrc.utexas.edu/fasearch/findingAid.cfm?eadid=00365 (last accessed 29 June 2025).

Hongcharu B., 'The Changing Roles of Mass Media Amidst the Growth of the Digital Media' (2024) 10(1) *Cogent Social Sciences*.

Houston B., 'The future of investigative journalism' (2010) 139(2) *Daedalus* 45–56.

Islam R., 'Into the Looking Glass: What the Media Tell and Why – An Overview' in R. Islam, S. Djankov & C. McLeish (eds), *The Right to Tell: The Role of Mass Media in Economic Development* (The World Bank 2002) 1–23.

Jebril N., *Is Watchdog Journalism Satisfactory Journalism?* (Reuters Institute for the Study of Journalism, 2013).

Jurkowitz Mark & Amy Mitchell, 'Most say journalists should be watchdogs, but views of how well they fill this role vary by party, media diet' (*Pew Research Center*, 26 February 2020) available at https://www.pewresearch.org/journalism/2020/02/26/most-say-journalists-should-be-watchdogs-but-views-of-how-well-they-fill-this-role-vary-by-party-media-diet/ (last accessed 15 September 2025).

Khan A. & al., 'The Impact of Social Media Echo Chambers on Semantic Integrity in Journalism: A Study of Language Change and Pragmatic Shift' (2024) 9(3) *Remittances Review* 1384–1409.

Kingston J., 'Watchdog Journalism in Japan Rebounds but Still Compromised' (2018) 77(4) *The Journal of Asian Studies* 881–893.

Kowalewski Max, 'Impact of Investigative Journalism' (YouGov, 31 October 2013) available at https://yougov.co.uk/consumer/articles/8012-impact-investigative-journalism (last accessed 15 September 2025).

Kurtz H., *Media circus: The trouble with America's newspapers* (Three Rivers Press, 1994).

Kutler S.I., 'In the Shadow of Watergate: Legal, Political, and Cultural Implications' (1994) 18(3) *Nova Law Review* 1743–1763.

Lanosga, G. & B. Houston, 'Spotlight: Journalists assess investigative reporting and its status in society' (2017) 11(9) *Journalism Practice* 1101–1120.

Lee S. & al., '"Fact-Checking" Fact Checkers: A Data-Driven Approach' (2023) *Harvard Kennedy School (HKS) Misinformation Review*, available at https://doi.org/10.37016/mr-2020-126 (last accessed 15 September 2025).

Lelo T., 'The Rise of the Brazilian Fact-checking Movement: Between Economic Sustainability and Editorial Independence' (2022) 23(9) *Journalism Studies* 1077–1095.

Li K. & C. Sparks, 'Chinese newspapers and investigative reporting in the new media age' (2018) 19(3) *Journalism Studies* 415–431.

LibertiesEU, 'What Is Watchdog Journalism: Definition, Examples, Effects On Democracy' (Liberties.eu, 17 January 2022), available at https://www.liberties.eu/en/stories/watchdog-journalism/43959 (last accessed 15 February 2024).

Liu X. & al., 'Checking the Fact-Checkers: The Role of Source Type, Perceived Credibility, and Individual Differences in Fact-Checking Effectiveness' (2023) 52(6) *Communication Research* 719–746.

Logue D. M. & S. R. Clegg, 'WikiLeaks and *The News of the World*: The Political Circuitry of Labeling' (2015) 24(4) *Journal of Management Inquiry* 394–404.

Lorentzen P., 'China's Strategic Censorship' (2014) 58(2) *American Journal of Political Science* 402–414.

Makhlouf, K., 'Investigative Journalism: A Tool to Fight Misinformation and Fake News' (*Q Views, Northwestern Qatar*, n.d.) available at https://qviews.qatar.northwestern.edu/11204/uncategorized/investigative-journalism-a-tool-to-fight-misinformation-and-fake-news/ (last accessed 15 September 2025).

Márquez-Ramírez M. & al., 'Detached or Interventionist? Comparing the Performance of Watchdog Journalism in Transitional, Advanced and Non-Democratic Countries' (2019) 25(1) *The International Journal of Press/Politics* 53–75.

McCoy M. E., *Scandal and Democracy: Media Politics in Indonesia* (Cornell University Press, 2019) 139–160.

Meyer T., *Media Democracy: How the Media Colonize Politics* (Polity Press, 2002)

Moore R. Cl, 'Environmental Issues and the Watchdog Role of the Media: How Ellul's Theory Complicates Liberal Democracy' (2001) 21(5) *Bulletin of Science, Technology & Society* 325–333.

Moyo G. & Phulu K. I., 'The Weaponisation of the Coronavirus Crisis in Zimbabwe: Legal and Extra-Legal Instruments' (2021) 13(1) *iBusiness* 48–66.

Napoli P. M., 'The Platform Beat: Algorithmic Watchdogs in the Disinformation Age' (2021) 36(4) *European Journal of Communication* 377.

Neuzil M. R., 'Muckraking and the Progressive Movement: The Interaction of Journalism and Society, 1902-1911' (PhD Dissertation, University of Minnesota, 1993).

Nomoplatform, 'About - Nomoplatform' available at https://www.nomoplatform.cy/about/ (last accessed 15 September 2025).

Nomoplatform, 'NGO Releases Nomoplatform™ Electronic Legislative Observatory' (2021) Nomoplatform available at https://www.nomoplatform.cy/ngo-releases-nomoplatform-electronic-legislative-observatory/ (last accessed 15 September 2025).

Nomoplatform, 'NomoPress Archives - Nomoplatform' (n.d.) available at https://www.nomoplatform.cy/category/nomopress/ (last accessed 15 September 2025).

Nord L. W., 'Investigative Journalism in Sweden' (2007) 8(5) *Journalism* 517–521.

Nord L. W. & T. von Krogh, 'Sweden: Continuity and Change in a More Fragmented Media Landscape' in J. Trappel & T. Tomaz (eds), *The Media for Democracy Monitor 2021: How Leading News Media Survive Digital Transformation* (Vol 1, Nordicom, University of Gothenburg, 2021) 353–380.

Olesen T., 'Whistleblowing and the Press: Complicating the Standard Account' (2023) 24(11) *Journalism* 2418–2435.

PEN International, 'Zimbabwe: Release Journalist Hopewell Chin'ono and Drop All Charges Against Him' available at https://www.pen-international.org/news/zimbabwe-release-journalist-hopewell-chinono-and-drop-all-charges-against-him (last accessed 29 June 2025).

Pham Phuong U. D., 'Check the checks: A comparison of fact-checking practices between newspapers and independent organizations during 2020 U.S. election presidential debates' (2024) 46(1) *News Research Journal* 71–95.

PressFreedom.eu, available at https://www.pressfreedom.eu/en/index.php (last accessed 16 September 2025).

Ritchie, D. A., 'Investigating the Watergate Scandal' (1998) 12(4) *OAH Magazine of History* 49.

Rodriguez R., 'A Vital Responsibility in Need of Support' (Spring 2008) *Nieman Reports* 62–65.

Rosen J. & al., 'Questions and Answers about Public Journalism' (2000) 1(4) *Journalism Studies*.

Rushton G., 'How Checking Facts Got Political' ABC News (Australia) (16 March 2025) available at https://www.abc.net.au/news/2025-03-16/fact-checking-partisan-media-meta-trump-politics/105046680 (last accessed 16 February 2024).

Schudson Michael, *Why Democracies Need an Unlovable Press* (Polity, 2008).

Schudson Michael, 'Reluctant Stewards: Journalism in a Democratic Society' (2013) 142(2) *Daedalus* 159–176.

Serrin J. & W. Serrin, *Muckraking! The Journalism that Changed America* (The New Press, 2002).

Skovsgaard M., L. Heiselberg & K. Andersen, 'The Context-Dependent Demand for Watchdog Journalism: Dynamics in Audience Expectations for Journalists' Role Performance' (2024) 18(9) *Journalism Practice* 2454–2475.

Spring J., *Conflict of Interests: The Politics of American Education* (McGraw-Hill, 2004).

Stier S., 'Democracy, autocracy and the news: the impact of regime type on media freedom' (2015) 22(7) *Democratization* 1273–1295.

Swire B. & al., 'Processing political misinformation: comprehending the Trump phenomenon' (2017) 4(3) *Royal Society Open Science* 160802.

The Briefing, Pew Research Center newsletter (10 January 2025).

The Washington Post, 'Trump Fact-Checks Interviews and Debates' (14 October 2024) available at https://www.washingtonpost.com/politics/2024/10/14/trump-fact-checks-interviews-debates/ (last accessed 15 September 2025).

Trappel Josef & T. Tomaz (eds), 'Success and Failure in News Media Performance: Comparative Analysis' in *Media for Democracy Monitor 2021* (Nordicom, 2022).

Trend D., *Cultural Democracy: Politics, Media, New Technology* (State University of New York Press, 1997).

UN Watch, 'About Us - UN Watch' available at https://unwatch.org/about-us/ (last accessed 15 September 2025).

UNESCO, 'Journalists and Whistleblowers: A Dynamic Relationship in the Fight against Corruption' (20 April 2023), available at https://www.unesco.org/en/articles/journalists-and-whistleblowers-dynamic-relationship-fight-corruption (last accessed 15 September 2025).

United Nations, Universal Declaration of Human Rights, available at https://www.un.org/en/about-us/universal-declaration-of-human-rights (last accessed 16 September 2025).

Usher N. & S. Kim-Leffingwell, 'How Loud Does the Watchdog Bark? A Reconsideration of Losing Local Journalism, News Nonprofits, and Political Corruption' (2023) 29(4) *The International Journal of Press/Politics* 960–982.

Verhoest P., 'Seventeenth-Century Pamphlets as Constituents of a Public Communications Space: A Historical Critique of Public Sphere Theory' (2018) 36(1) *Theory, Culture & Society* 47–62.

Vouliwatch, 'Empowering democracy', available at https://vouliwatch.gr/about/en (last accessed 15 September 2025).

Walters K., 'Comparative and Critical Analysis: The Roles of Civic and Traditional Journalism' (unpublished PhD thesis, University of Nevada Las Vegas, 2011).

Weatherald Nathalie, 'Anti-SLAPP Law Gets Final Seal of Approval from EU Parliament' (Euractiv, 27 February 2024) available at https://www.euractiv.com/section/media/news/anti-slapp-law-gets-final-seal-of-approval-from-eu-parliament/ (last accessed 16 September 2025).

Wittgenstein L., *Philosophical Investigations* (G. E. M. Anscombe tr., Blackwell Publishing, 1953).

Zaffarano F., 'Attack: Brazil's "Capitol"—Lupa Was Already Debunking Lies on TikTok' (*Reuters Institute*, 10 January 2023) available at https://reutersinstitute.politics.ox.ac.uk/news/attack-brazils-capitol-lupa-was-already-debunking-lies-tiktok (last accessed 1 July 2025).

Žuffová M., 'Fit for Purpose? Exploring the Role of Freedom of Information Laws and Their Application for Watchdog Journalism' (2021) 28(1) *The International Journal of Press/Politics* 300–322.

SPECIAL SECTION

New Technologies in the World of Communication

Copyright Protection for AI-Generated Works: Is It Something We Want?

Irini Stamatoudi[1]

Abstract

This article explores the extent to which works generated by artificial intelligence (AI) are, or should be, eligible for copyright protection. It considers two key questions: (a) whether and to what extent human involvement is present in the creation of AI-generated works—an issue closely tied to the concept of originality; and (b) whether we should want such works to be covered by copyright, which grants creators exclusive rights and restricts use by others, even for a limited time. The article concludes that when AI-generated works lack meaningful human involvement, they do not meet the threshold of originality and therefore cannot—and should not—qualify for copyright protection.

Keywords: copyright; artificial intelligence; originality; AI-generated works; copyrighted works

Introduction

In recent years, the extent to which works generated by artificial intelligence (AI) are, or should be, protected by copyright has been much debated. The discussion considers two key questions: (a) whether and to what extent humans are involved in the creation of AI-generated works—an issue closely tied to the concept of originality; and (b) whether we should want such works to be covered by copyright, which grants creators exclusive rights and restricts use by others, even for a limited time. The latter case would require a change in copyright law to accommodate AI-generated works.

This paper is divided into three parts. The first part discusses what we mean by AI-generated works. The second addresses the basics of copyright in the EU and how the notions of human authorship and originality lie at the heart of this type of protection. The third part explores how the above principles apply to AI-generated works. It also references various countries' experiences (e.g., case law, soft law/guidelines). The conclusion sums up where we stand today, the way forward, and whether AI-generated works should be protected by copyright and to whose benefit.

[1] Professor, Faculty of Law, University of Nicosia, UNIC Athens.

AI and AI-Generated Works

Broadly defined, AI is technology that enables computers and machines to simulate human intelligence and problem-solving capabilities. Article 3(1) European Union's Artificial Intelligence Act (hereafter, the EU AI Act) defines it 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.[2]

Since the 1970s, computers have been used as tools to generate creative output. Their primary characteristic, however, is that their output has been directed and determined by programmers. AI systems depart from this model. The EU AI Act clarifies that we should 'distinguish [AI systems] from simpler traditional software systems or programming approaches'; and that AI systems 'should not cover systems that are based on the rules defined solely by natural persons to automatically execute operations'.[3] In the case of AI, computers are fed vast amounts of data. Machine learning, which is a subset of AI, uses algorithms trained on this data to produce models that can mimic the ways humans think and, on that basis, performs complex tasks such as reasoning, analysis, synthesis, 'creating', producing, and so on. Software is no longer a tool; it makes many of the decisions involved in the creative process without human intervention, at least as we traditionally know it.

The type of AI that is today used to 'create' or produce 'works' is called generative AI (GenAI). GenAI generates content, such as text, images, videos, sounds, code, 3D designs, etc. based on user orders or prompts. Machine learning—using built-in algorithms and data input—produces autonomous systems capable of learning and acting without being programmed by a human. While programmers can set parameters, the work is generated by the computer programme itself (the 'neural network') in a process similar to that of a human brain. The content produced is based on the data inserted by humans, i.e. the programmers—or programmed by humans to be found, scanned, and inputted into the system.[4] The data is processed by the machine using algorithms also designed and trained by programmers. The outcome is directed by humans (the

[2] Artificial Intelligence Act (Regulation (EU) 2024/1689), OJ 13 June 2024.

[3] Ibid, Recital 12.

[4] This means that humans have control as to the inputted data.

users) based on general guidelines/orders they input into the application, the prompts. In fact, prompts serve as orders that describe, in a limited number of words (ranging from a few to usually 400 words, depending on the application), the output that the user envisions. The user also has the option, instead of entering a single prompt into the system, to submit iterative or repeated prompts. However, even in the latter case, the outcome cannot be predicted by either the programmer or the user.

Human Authorship

Copyright protects original creations of the mind that have been expressed in any form in the field of art, literature, or science. The main requirement of a work being protected by copyright is to have originality or creativity.[5] Originality is a notion closely linked to human authorship in the sense that only humans can create[6] and therefore merit copyright protection[7] in so far their creation is original (and meets the requirements for protection as provided in the law).[8] In fact, authors/creators are afforded a set of absolute and exclusive legal rights (i.e. legal monopoly)—albeit limited in duration—which allows them to be compensated for their intellectual effort by exploiting their work exclusively (economic rights) and securing their personal interests in it (moral rights). This exclusive protection gives authors an incentive to

[5] This is also a principle enshrined in the 1886 Berne Convention (last amended in 1979). On the importance of maintaining human authorship in copyright law, see Sam Ricketson, 'People or Machines? The Berne Convention and the Changing Concept of Authorship' (1991) 16(1) Columbia Journal of Law and the Arts 1.

[6] Yet, authorship has also been described as a legal fiction used to allocate rights (Lionel & al., *Intellectual Property Law* 6th edn., OUP, 2022, 136). According to Daniela Simone's chapter titled 'Authorship' in a forthcoming volume (Paul Torremans & al. (eds), *Elgar Encyclopedia of Intellectual Property Law* (Cheltenham, UK: Edward Elgar, 2026)): 'this is most clearly the case in relation to entrepreneurial copyright, which is a fixation only form of protection that does not rely on originality as a condition of subsistence. In the case of entrepreneurial works, the person responsible for the work might be deemed to be its author. This person is usually the instigator of a work, and their contribution is likely to be primarily organisational or financial rather than creative. Thus, for example, a sound recording might be authored by its producer; and a broadcast might be authored by the person making it (CDPA, s9(2)). Similarly, authorship operates as a legal fiction in the context of the US provisions on work made for hire, where an employer might be deemed the author of works made by their employees in certain circumstances (Copyright Act 1976 (US), s201(b))'.

[7] According to the principle of truth, only human beings can exercise creativity. Animals (*Naruto v Slater* No. 16-15469 (9th Cir. 2018)) or spirits (*Cummins v Bond* (1926) 1 Ch 167) are not entitled to copyright protection.

[8] Sam Ricketson & Jane Ginsburg, *International Copyright and Neighbouring Rights: The Berne Convention and Beyond* (2nd ed, OUP, 2006) 358–359.

create further. In this way, copyright protection benefits society by contributing to the enhancement of knowledge, culture, and the economy.

New types of works have been included in copyright over the years, primarily due to the evolution of technology, such as photographs, films, software, databases, and 3D printing. The law has adapted to the novelties, either through new provisions, or, in the case of complex works, through case law that has taken a distributive approach, in other words, protecting parts of rather than the whole work. For example, websites are protected through their components, such as graphics, text, sound, etc., whereas video games are partly protected as software and partly as audiovisual works, multimedia works, etc. Yet, in all these cases, the initial beneficiary (the first owner of copyright) is the natural person who created the work, and any protection afforded has always been based on that person's creativity.

In the UK (as well as in other common law jurisdictions, such as Ireland, New Zealand, India, Hong Kong, and South Africa)[9] there is an exception to the rule of human authorship.[10] According to s 178 of the UK Copyright, Patents and Designs Act 1988 (CDPA), 'computer-generated' means that the work is generated by computer in circumstances such that there is no human author of the work. Section 9.3 provides that when we deal with a computer-generated work with no human author, the author 'shall be taken to be the person by whom the arrangements necessary for

[9] Ireland (Copyright and Related Rights Act of 2000, section 2): 'a computer-generated work means that the work is generated by a computer in circumstances where the author of the work is not an individual. The author of such a work is the computer programmer'; New Zealand (Copyright Act of 1994, section 2(1): 'a computer-generated work means a work that is generated by a computer and where there is no human author. The author of such computer-generated works is the programmer'; India (Copyright Act of 1957, section 2(d)(vi)); Hong Kong (Copyright Ordinance, sections 198, 11(3), 17960, 91(2)(c), and 93(2)); South Africa (Copyright Act of 1978, amended in 1992, adding paragraph (h) to the definition of author in section 1(iv)).

[10] Robert Denicola, 'Ex Machina: Copyright Protection for Computer-Generated Works' (2016) 69 *Rutgers University Law Review* 251, available at https://ssrn.com/abstract=3007842; Jyh-An Lee, 'Computer-generated Works under the CDPA 1988' in Jyh-An Lee, Reto Hilty & Kung-Chung Liu (eds), *Artificial Intelligence and Intellectual Property* (Oxford, UK: Oxford University Press, 2021) 177; Massimo Maggiore, 'Artificial intelligence, computer generated works and dispersed authorship: spectres that are haunting copyright' in Enrico Bonadio & Nicola Lucchi (eds), *Unconventional Copyrigh*t (Cheltenham, UK: Edward Elgar, 2018); Jani McCutcheon, 'Curing the Authorless Void: Protecting Computer-Generated Works following IceTV and Phone Directories' (2013) 37(1) *Melbourne University Law Review*, available at http://www5.austlii.edu.au/au/journals/MelbULawRw/2013/9.html; Pamela Samuelson, 'Allocating ownership rights in computer-generated works' (1985) 47 *University of Pittsburgh Law Review* 1,185; Ingrida Veiksa, 'Protection of computer-generated works' (2021) 10(1) *International Journal of Artificial Intelligence (IJ-AI)* 234.

the creation of the work are undertaken'. 'Computer-generated' works need to be distinguished from 'computer-assisted' works in the same way we distinguish between 'transformative AI', where a sophisticated assistive AI is used as a tool by the author and a degree of creative input subsists in the output, and 'generative AI', where there is no human intervention, and no human input is discernible in the output.[11]

Computer-generated works are not entitled to moral rights protection, and their term of protection amounts to 50 years from the date of publication (CDPA sections 81 and 12.7, respectively). Not much case law exists on the issue. In *Nova Productions v Mazooma Games and Others*,[12] the 'work' at hand was the display of a series of composite frames generated by a computer programme using bitmap files.

The provision for the protection of computer-generated works (as will also be demonstrated in the next section of this article) contradicts the basic premise of copyright law that only human beings can create and that it is only them that should be entitled to incentives to create further. Thus, the concept of computer-generated works, which is also difficult to discern from computer-assisted works, finds no justification in copyright (at least when it comes to conferring the full set of rights to someone that did not have actual creative involvement in the 'production' process of a 'work').[13] Apart from that, the concept of computer-generated works does not appear to be compatible with EU copyright law,[14] which requires human authorship and that the work is its

[11] Stefania Ercolani, 'Computer-Generated Works' in Paul Torremans & al. (eds), *Elgar Encyclopedia of Intellectual Property Law* (Cheltenham, UK: Edward Elgar, 2026): 'the term "generative" applies to a process autonomously determining the output, while "transformative" or "assistive" applies to a human-directed process whose result reflects the author's personality'. This paper focuses exclusively on generative AI.

[12] [2007] EWCA Civ 219, [2007] Bus LR 1032.

[13] See also James Grimmelmann, 'There's No Such Thing as a Computer-Authored Work - And It's a Good Thing, Too' (2016) 39 *Columbia Journal of Law & the Arts*, available at https://scholarship.law.cornell.edu/facpub/1483/; Gwilym Harbottle, Nicholas Caddick & Uma Suthersanen (eds), *Copinger and Skone James on Copyright* (18th edn., London, UK: Sweet & Maxwell, 2000) 3–273; Jane C. Ginsburg & Luke A. Budiardjo, 'Authors and Machines' (2019) 34 *Berkeley Technology Law Journal* 343, available at https://scholarship.law.columbia.edu/faculty_scholarship/2323; Ana Ramalho, 'Will Robots Rule the (Artistic) World? A Proposed Model for the Legal Status of Creations by Artificial Intelligence Systems' (2017) 21 *Journal of Internet Law* 11, 17. The UK is considering to abolish the provision on computer-generated works. See the Report on Copyright and Artificial Intelligence, 18/3/2026. https://www.gov.uk/government/publications/report-and-impact-assessment-on-copyright-and-artificial-intelligence/report-on-copyright-and-artificial-intelligence

[14] E.g. Case C-5/08 *Infopaq International A/S v. Danske Dagblades Forening* (2009), Judgment of 16 July 2009 ECR I-656; Cases C-403/08 and C-429/08 *Football Association Premier League Ltd and Karen Murphy*, Judgment of 4 October 2011, (2011) ECR I-9083; Case C-145/10 *Eva-Maria Painer v. Standard VerlagsGmbH and others* (2013), Judgment of 7 March 2013 (2011) ECR I-12533; Case C-604/10, *Foot-*

author's own intellectual creation.[15] According to EU law (which has been clarified by the Court of Justice of the European Union's (CJEU) recent case law), there is no criterion for copyright protection apart from originality, to the extent, of course, that we are dealing with copyright subject matter expressed in a form that makes it identifiable with sufficient precision and objectivity, even though that expression is not necessarily in permanent form.[16] The case of computer-generated works should not be compared to instances where the law permits legal entities to be designated as rightsholders, as in these instances, the creator remains a natural person.[17]

The UK provision on computer-generated works is also not very clear as to: how it applies to joint authorship/multiple authors;[18] how one distinguishes between computer-generated and computer-assisted works;[19] who made the arrangements; and

ball Dataco Ltd and Others v. Yahoo! UK Ltd and Others (2012), Judgment of 1 March 2012 (published in the electronic Reports of Cases: https://curia.europa.eu/jcms/jcms/P_106320/en/?rec=RG&jur=C&anchor=201203C0031); Case C-393/09, *Bezpečnostní softwarová asociace – Svaz softwarové ochrany v. Ministerstvo kultury* (2010) ECR I-13971, Judgment of 22 December 2010, (2010) ECR I-13971; Case C-406/10, *SAS Institute Inc. v. World Programming Ltd* (2012), Judgment of 2 May 2012 (published in the electronic Reports of Cases: https://curia.europa.eu/jcms/jcms/P_106320/en/?rec=RG&jur=C&anchor=201205C0081#201205C0081); Irini Stamatoudi, 'Originality under EU Copyright Law' in Paul Torremans (ed), *Copyright Law: A Handbook of Contemporary Research* (Cheltenham, UK: Edward Elgar, 2017) 57; E. Rosati, *Originality in EU Copyright* (Cheltenham, UK: Edward Elgar, 2013). See also Kevin Garnett, Gillian Davies & Gwilym Harbottle (eds), *Copinger and Skone James on copyright* (16th edn., UK: Sweet & Maxwell, 2010); Irini Stamatoudi, 'The concept of originality in Greece' in Indranath Gupta, Vishwas H. Devaiah & Manveen Singh (eds), *Handbook on Originality in Copyright: Cases and Materials* (Springer Nature, 2024) available at https://link.springer.com/referencework/10.1007/978-981-19-1144-6; Irini Stamatoudi, '"More" Originality for Cypriot Copyright Law according to the CJEU's Case Law' (2018) 30(1) *Cyprus Review* 207–231; Lionel Bently & al., *Intellectual Property Law* (6th edn., Oxford, UK: OUP, 2022).

[15] The UK originality criterion is not compatible with EU copyright law. D.J. Gervais 'The compatibility of the skill and labour standard with the Berne Convention and the TRIPs Agreement' (2004) 26 *EIPR* 75, 79; J. Griffiths 'Infopaq, BSA and the "Europeanisation" of United Kingdom copyright law' (2011) 16 *Media & Arts Law Review*, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1777027.

[16] C-310/17, *Levola Hengelo BV v Smilde Foods BV* (2018), Judgment of 13 November 2018, para 40 and 42–44.

[17] See, for example, Council Directive 91/250/EEC, 2009/24/EC, Art 2(1) according to which the author is the natural person who created the programme. Member States are allowed to designate a legal person as the rightsholder. See also the Database Directive 96/9/EC, Art 4(1) according to which the author of a database is the natural person or group of natural persons who created the database or, where the legislation of the Member States so permits, the legal person designated as the rightsholder by that legislation.

[18] Michael Tappin & al., *Laddie, Prescott and Vitoria: The Modern Law of Copyright* (5th edn., Lexis Nexis, 2018) 36.41.

[19] Robert Denicola, 'Ex Machina: Copyright Protection for Computer-Generated Works' (2016) 69 *Rutgers University Law Review* 251, available at https://ssrn.com/abstract=3007842.

whether the notion of originality plays any role, what that role is, and how it is assessed concerning the fictitious author. Additionally, the fact that relevant case law is scarce suggests that there was a social need for such a legal fiction. These provisions may well be viewed as a form of related right rather than copyright.

Legislation that offers an exception to the human authorship rule can also be found in Kenya. The Kenyan Copyright Act[20] provides that an artwork is 'an original work of visual art created by an artist or artists or *produced under their authority*' (emphasis added).[21] The provision refers to a commission or basic instruction to produce an artistic work; this, again, contradicts the basic premise of copyright law as found in the international conventions.

In this section, we demonstrated that human authorship and originality are the cornerstones of copyright law and its justification. It is from this angle that AI-generated works should be approached. The next section considers whether, according to EU law, an AI work can be regarded as original and thus protectable. We will then explore how different jurisdictions have dealt with the matter and draw conclusions as to whether the law, as it stands today, can meet the needs of the era or if we should seek other legislative solutions.

Originality Under EU Copyright Law

According to settled case law, the need for a uniform application of EU law and the principle of equality require that the terms of a provision of EU law, which makes no express reference to the law of the Member States for the purpose of determining its meaning and scope, must be given an independent and uniform interpretation throughout the EU.[22] In this light, Member States are not free to interpret originality and the 'author's own intellectual creation' criterion using their national legal tradition.

In a series of groundbreaking cases, the CJEU has clarified that the originality criterion—which was initially provided expressly only for photographs, computer programmes, and databases—applies to all types of works, along with the content of this criterion and how it is assessed.

---

[20] Kenyan Copyright Act No. 12 of 2001.

[21] Ibid, Section 2.1.

[22] See, inter alia, Case C-5/08 *Infopaq International* [2009] ECR I-6569, paragraph 27; Case C-34/10 *Brüstle* (nyp), paragraph 25; and Case C-510/10 *DR and TV2 Danmark* (nyp), paragraph 33, as referred to in Case C-128/11 *UsedSoft GmbH v. Oracle International Corp*. (nyp) (para 39). See also see, in particular, Case C-357/98 *Yiadom* [2000] ECR I-9265 (para 26), Case C-245/00 *SENA* [2003] ECR I-1251 (para 23) and Case C-306/05 *SGAE* [2006] ECR I -11519 (para 31).

Infopaq: The EU Originality Criterion

Infopaq[23] is a landmark case for the concept of originality in the EU. It examined whether 11-word summaries of articles from Danish newspapers, created through a 'data capture process', were protected by copyright.

The CJEU stated that the author's right to authorise or prohibit reproduction applies to a 'work'. According to the Berne Convention (and in particular Articles 2(5) and (8)), the protection of certain subject-matter as artistic or literary works presupposes that they are 'intellectual creations'. EU law provides that works such as computer programmes, databases, or photographs, are protected by copyright only if they are original in the sense that they are their 'author's own intellectual creation'. In addition, the Court stated that protection also extends to parts of a work since they share the originality of the whole work and contain elements that express the author's intellectual creation.

Words themselves are not protected, as they do not constitute elements eligible for protection, nor are they the intellectual creation of the author who employs them.[24] It is only through the choice, sequence, and combination of words that the author may express their creativity in an original manner and achieve a result, which is an intellectual creation. Certain isolated sentences, or even certain parts of sentences in a text, may be suitable for conveying to the reader the originality of a publication, such as a newspaper article, by communicating to that reader an element which is, in itself, the expression of the intellectual creation of the author of that article. Such sentences or parts of sentences may therefore be protected.[25]

Murphy, Painer, and Football Dataco: Creative Choices and a Personal Touch

Murphy

The issue of originality was further clarified in a number of cases that followed *Infopaq,* including *Murphy*.[26] The issue was whether sporting events, which were the subject of a retransmission, were protected by copyright. The CJEU ruled that sport-

[23] C-5/08 *Infopaq International A/S v Danske Dagblades Forening*, Judgment of 16 July 2009 [2009] ECR I-656.

[24] Ibid, paras 45–46.

[25] Ibid, para 47.

[26] C-403/08 and C-429/08 *Football Association Premier League Ltd and Karen Murphy*, Judgment of 4 October 2011, [2011] ECR I-9083.

ing events, including football games, constitute subject-matter that falls outside the scope of copyright protection because they are not intellectual creations. Intellectual creations require creative freedom, while sporting events are dictated by the rules of the game, leaving no room for creative freedom. Thus, a new parameter (which was not mentioned in EU copyright law) was introduced in assessing originality: the author's creative freedom.

Painer

Almost two years later, *Painer*[27] *shed more light on originality by dealing with that of photographs. The CJEU used an a contrario* argument from *Murphy*.[28] A work is original if the author expresses their creative abilities in producing the work by making free and creative choices. As regards a portrait photograph, the photographer can make free and creative choices in several ways and at various points in its production. For example, in the preparation phase, they can choose the background, the subject's pose and the lighting; when taking the photograph, they can choose the framing, the angle of view, and the atmosphere; finally, when selecting the snapshot, the photographer may choose from a variety of developing techniques or use computer software. By making these choices, the photographer can mark the work created with their 'personal touch'. From this perspective, the freedom for the photographer to exercise their creative abilities can be substantial.

This case also dealt with the issue of whether the protection afforded to portrait photography is inferior to that enjoyed by other types of photographic works. The Court held that nothing in the Information Society Directive or any other directive applicable in this field, supports the view that the extent of protection should depend on the degree of creative freedom in the production of the work. So, the protection of a portrait photograph cannot be inferior to that enjoyed by other works, including other photographic works.[29] The judgment thus confirms that the criterion for protection is one, that of originality; and it is the same for all types of works.

[27] Case C-145/10, *Eva-Maria Painer v. Standard VerlagsGmbH and others*, Judgment of 7 March 2013.

[28] Joined Cases C-403/08 and C-429/08 *Football Association Premier League and Others* [2011] para 98.

[29] Ibid, paras 85-99. See also the dictum (2) of the judgment.

Football Dataco

Football Dataco[30] tested whether the UK criterion of skill, labour, and judgment was compatible with EU law. The case concerned annual fixture lists of the football leagues (which qualified as databases) in England and Scotland that are based on rules and procedures. The process for their production was only partially computerised, leaving room for judgment and discretion. One of the issues was whether the notion of the 'author's own intellectual creation' requires more than significant labour and skill from the author and, if so, what that additional requirement is.

According to the Court, the author's significant skill and labour cannot, as such, justify copyright protection in the sense that they do not express any originality in the selection or arrangement of the data contained in the database. If the procedures for creating the fixture lists are not supplemented by elements reflecting originality in the selection or arrangement of the data contained in those lists, these lists are not eligible for copyright protection. In effect, the Court (once again) linked originality to the author's free and creative choices and their personal touch on the work.[31]

Interim Conclusions

EU copyright law—clarified by relevant case law—states that only humans qualify as authors to the extent they have exercised creativity, meaning that the output (work) is its author's intellectual creation, marked by their personal touch. Derivative works are also protected by copyright to the extent they demonstrate originality and only in relation to those parts that are original.

Transformative AI works may qualify as derivative works to the extent a natural person uses AI as a tool to alter, retouch, adapt, etc. a preexisting work, irrespective of whether the preexisting work is protected or not and who holds the rights. Both the work (if a human creates it and is original) as well as the alterations/changes/adap-

[30] C-604/10, *Football Dataco Ltd and Others v. Yahoo! UK Ltd and Others*, Judgment of 1 March 2012.

[31] There have also been cases of the CJEU concerning the issue of originality for computer programmes and designs, which fall out of the scope of this paper. E.g., Case C-393/09, *Bezpečnostní softwarová asociace – Svaz softwarové ochrany v Ministerstvo kultury*, Judgment of 22 December 2010; Case C-406/10, *SAS Institute Inc. v. World Programming Ltd*, Judgment of 2 May 2012; C-355/12, Nintendo Co. Ltd, Nintendo of America Inc., Nintendo of Europe GmbH v PC Box Srl, Net Srl, Judgment of 23 January 2014; Case C-168/09 *Flos SpA v Semeraro Case Famiglia SpA*, Judgment of 27 January 2011; C-683/17 *Cofemel – Sociedade de Vestuário SA v G-Star Raw CV,* Judgment of 11 April 2019; C-833/18 *Brompton Bicycle Ltd v Chedech/Get2Get*, Judgment of 11 June 2020; Joint Cases C-580/23 & C-795/23 *Mio/Konektra*, Judgments of 4 December 2025.

tations are protected (to the extent they are original), and whoever has authored the work and/or the alterations is entitled to copyright protection and only for the parts they have authored and which are original.[32]

Transformative AI does not itself 'create' but assists in the creative process. The author's role goes beyond merely initiating the creation of the 'work' to include the exercise of actual control over the specifics and content of the 'creative' output. In AI-generated works, the user has no actual control over the application's outcome. This is because the resulting content cannot be predicted by the user or the programmer, even though it very much depends on: a) what the system has been fed (i.e., the data input); b) the training of the algorithm used to produce the output; and c) the prompt used. So, the user prompts the production process but has no control over the result. The prompt itself equates an idea, and it is a well-established principle in copyright law that ideas are not protected[33] since their free propagation, proliferation, and circulation are

[32] Since November 2022, several class action suits have been filed in the US concerning the production of derivative works by AI applications (in comparison to the works fed into the system) against companies operating generative AI systems. One side included the copyright owners of works that were allegedly fed into the AI systems, and on the other, companies like OpenAI, Microsoft, Meta, Alphabet, Stability AI, Midjourney, and DeviantArt. Plaintiffs' claims were based on grounds including copyright, privacy, and contract law. The District Court of the Northern District of California (30.10.2023) re class action that was filed by Sarah Andersen; Kelly McKernan; and Karla Ortiz against Stability AI; Midjourney; and DeviantArt on 13 January 2023, was not convinced that these works were all used by the AI systems and asked the plaintiffs to provide more clarity on their theory regarding the use of 'compressed files' of the training images in each product that uses the Stable Diffusion library. It also asked the Plaintiffs to demonstrate substantial similarity between works or styles in the specific AI outputs and the specific training images to show that certain outputs are derivative works of the training images. See also the *Porsche 911* case (German Federal Court of Justice (BGH) GRUR 2022, 899 – Porsche 911). The dispute concerned the eighth series of the Porsche 911, which has been marketed since 2011. The daughter of the designer of Porsche 356 and Porsche 911 models produced from 1950 to 1956 argued that her father's models were partly reproduced in the new Porsche 911. According to the BGH, the previous courts that heard the case should not have referred solely to the fact that it was not enough that the personal features borrowed from the protected older work had 'faded' (this is what Section 24 UrhG aF provided and which was considered by the CJEU to be incompatible with Directive 2001/29/EC). The criterion of fading should be understood in conformity with EU law in the sense of the criterion of a lack of recognisability of the creative elements giving rise to the protection of the used work. (Section 23(1) sentence 2 UrhG nF now provides that there is no adaptation or transformation if the newly created work maintains a sufficient distance from the used work). So, there should be recognisable parts in the new work carrying elements of the originality of the used work. C-476/17 *Pelham & Others*, Judgment of 29 July 2019; Joint Cases C-580/23 & C-795/23 *Mio/Konektra*, Judgments of 4 December 2025

[33] The 1996 WIPO Copyright Treaty (WCT) and the 1994 Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) have explicitly declared that copyright protection extends to expressions and not to ideas, procedures, methods of operation, or mathematical concepts as such (Articles 2 and 9§2 respectively).

prerequisites for the advancement of knowledge, science, arts, and culture. Any other approach would create unjustified monopolies and stifle competition.

The next section examines how the different jurisdictions have dealt with the issue of copyright in AI-generated works.

National Approaches on AI-Generated Works

The US Copyright Office (USCO) specifies that one can 'register an original work of authorship, provided that the work was created by a human being'.[34] It rejects material produced through prompted technology where the use of AI is more than *de minimis*. This is in accordance with the judgment in *Feist*, which specified that copyright law only protects 'the fruits of intellectual labor [...] founded in the creative powers of the mind'.[35]

Later cases confirmed this stance. In 2019, the USCO refused to register a claim of copyright in an artwork called *A Recent Entrance to Paradise*, which was produced by Stephen Thaler's AI system Creativity Machine, because of lack of human authorship. The decision was confirmed by the Review Board of the Copyright Office in February 2022.[36] In August 2023, it was also affirmed by a US District Court in *Thaler v Perlmutter*.[37] The USCO also refused to register claims of copyright in works (images) created through substantial iterative prompting of a generative AI tool, Midjourney, (*Théâtre D'opéra Spatial*[38] *and Zarya of the Dawn*[39]) because the changes by the alleged author were too minor and imperceptible to supply the necessary creativity for copyright protection. These rulings take a narrower view of authorship compared

[34] See the US Copyright Office's *Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence* (issued March 16, 2023), available at https://www.copyright.gov/comp3/chapter300.html and Notice of Inquiry on Copyright and Artificial Intelligence, published in the Federal Register on August 30, 2023. See also National Policy Framework. Artificial Intelligence, The White House. Legislative Recommendations, US, March 2026.

[35] *Feist Publications* v *Rural Telephone Service Company, Inc. 499 U.S. 340 (1991)*, available at https://supreme.justia.com/cases/federal/us/499/340/.

[36] The decision was confirmed by the Review Board of the Copyright Office in February 2022 as well as in the recent decision by the US District Court of Columbia of 18 August 2023, No 22-1564, specifying that 'human authorship is a bedrock requirement of copyright', available at https://ecf.dcd.uscourts.gov/cgi-bin/show_public_doc?2022cv1564-24.

[37] US District Court for the District of Columbia [2023]: *Thaler v. Perlmutter*, No. 22-CV-384-1564-BAH, available at https://www.wipo.int/wipolex/en/judgments/details/1840.

[38] Available at https://www.copyright.gov/rulings-filings/review-board/docs/Theatre-Dopera-Spatial.pdf.

[39] 21 February 2023, Registration No. VAu001480196, available at https://www.copyright.gov/docs/zarya-of-the-dawn.pdf.

to the previous one (*A Recent Entrance to Paradise*, concerning substantial iterative prompting), under which a human prompter of an AI tool will not be considered an author of the output if the output details cannot be predicted in advance.

Similarly, the USCO rejected the registration of a computer image titled *Suryast*, created by Ankit Sahni. Sahni created the work with three elements: a) his own original photograph; b) an image of Vincent van Gogh's *The Starry Night*; and c) a software programme called RAGHAV, which enables a user to apply the style of one image to the style of another image to generate a new third image.

Sahni applied to the USCO to register this third image—*Suryast*—on 1 December 2021 and listed himself and RAGHAV (the software) as the two authors. The application was rejected by reason of lack of human authorship; and subsequently, on reconsideration, on the basis that it was a derivative work, being a digital adaptation of a photograph, and that RAGHAV generated the new aspects of the work and therefore they were not the result of human creativity or authorship. On 10 July 2023, Sahni requested that the USCO again reconsider its refusal to register the work and argued that he was responsible for key creative decisions, including selecting the original photograph, *The Starry Night* as the style input, and the variable value for the amount of style transfer; all these choices, according to him, amounting to sufficient human authorship. Sahni also argued that the work was not derivative because the original photograph was 'an early stage of what would ultimately become the work'. The Copyright Review Board affirmed the USCO's decision on the basis that the work was 'not the product of human authorship because the expressive elements of pictorial authorship were not provided by Mr Sahni' but by RAGHAV. Sahni's contribution consisted of selecting the base image, the style image, and the level of style transfer. These constituted nothing more than the unprotectable idea behind the work, which was an altered version of his photograph in the style of *The Starry Night*. In contrast, the Canadian Intellectual Property Office, before which Sahni also brought the case, was more liberal. In 2021, it registered *Suryast*,[40] and listed RAGHAV as a co-author.[41]

In the 2012 case *Acohs Pty Ltd v Ucorp Pty Ltd*, the Federal Court of Australia found that a work generated with the intervention of a computer could not be protected by copyright because it was not produced by a human.[42]

[40] Available at https://www.ic.gc.ca/app/opic-cipo/cpyrghts/dtls.do?fileNum=1188619.

[41] The Canadian Intellectual Property Office registers as copyrighted works even works 100% created by AI.

[42] Available at http://www8.austlii.edu.au/cgi-bin/viewdoc/au/cases/cth/FCAFC/2012/16.html.

The Municipal Court of Prague (in the Czech Republic) held that an image generated by the AI programme DALL-E, based on a prompt saying 'Create a visual representation of two parties signing a business contract in a formal setting, such as a conference room or a law firm office in Prague. Show only hands', was not capable of being protected by copyright, as it was not authored by a natural person.[43]

China seems to be divided on the issue of copyrighting AI-generated works. In 2018, the Beijing Internet Court (BIC) in *Feilin v Baidu* (the first reported Chinese case on the issue) dealt with the copyrightability of an 'Analysis Report on Judicial Big Data of the Entertainment Industry - Film Volume Beijing', which was written using legal statistical data software.[44] The report consisted of both visuals and text. The visuals were generated by the software and then modified and coloured manually by the plaintiff's staff. As for the text, the plaintiff's team examined the search results and removed irrelevant court decisions. After this, the team created the report by analysing the statistics.[45] The Court found that the report was original, made by natural persons, and reflected the choice, judgement, and analysis of the relevant data. According to the BIC, only a production by a natural person could be regarded as a 'work'. AI software cannot be considered an author. Also, a developer or a user cannot qualify as an author because they do not create or generate production. The Beijing Intellectual Property Court (on appeal) affirmed the ruling of the lower court.[46]

In the 2019 case *Tencent v Yingxun*, the Shenzhen District Court granted copyright protection to an AI-generated stock market financial review consisting of nine paragraphs, which was generated by the AI application Dreamwriter.[47] The Court found that the work was original because the form of expression in the review was individually determined and arranged by Tencent employees, and the form of expression had a certain originality even though it was not unique. It also noted that, to assess whether

[43] https://justice.cz/web/msp/rozcestnik.

[44] Available at https://www.chinadaily.com.cn/specials/BeijingInternetCourtCivilJudgment(2018)Jing0491MinChuNo.239.pdf. See also L. Ju Yoen, 'Artificial Intelligence Cases in China: Feilin v. Baidu and Tencent Shenzhen v. Shanghai Yingxin' (2021) 1 *China & WTO Review* 211–222, available at https://pdfs.semanticscholar.org/b61c/cb98fd748b7ee654895ee133d5363e4d8556.pdf.

[45] Kan He, 'Feilin v. Baidu: Beijing Internet Court tackles protection of AI/software-generated work and holds that copyright only vests in works by human authors' (*The IPKat*, 9 November 2019), available at https://ipkitten.blogspot.com/2019/11/feilin-v-baidu-beijing-internet-court.html.

[46] Beijing Intellectual Property Court Civil Judgment (2019) Jing 73 MinZho No. 2030.

[47] Guangdong Shenzhen Nanshan District People's Court Civil Judgment (2019) Yue 0305 MinChu No. 14010 (First Instance). For an English translation, see Xiaoshuai Ren, Tencent Dreamwriter: Decision of the People's Court of Nanshan (District of Shenzhen) 24 December 2019-Case No. (2019) Yue 0305 Min Chu No. 14010, 51(1) INt'l reV. INtell. ProP. & comPetItIoN L. 652-9 (2020).

an act is a creative one, one should consider whether there's intellectual activity and a direct connection between the act and the specific form of expression in the work.[48]

On 27 November 2023, in *Li v Liu*, the BIC ruled in an infringement lawsuit that an image of a woman generated using Stable Diffusion (an open-source generative AI model that creates images from textual prompts) is copyrightable as it reflects a human's original intellectual investment and should be recognised as a work and protected by copyright law.[49] It also ruled that the person who prompted the AI-generated image is entitled to the right of authorship under Chinese copyright law.[50]

On 13 February 2026, the District Court of Munich[51] denied copyright protection to images of three logos (a handshake/bell, an envelope, and a laptop) generated by a text-to-image generative AI system. The Plaintiff argued that his extensive iterative prompting (including a 1,700-character instruction for a single logo) constituted a "personal intellectual creation".[52] The Court stated that the work should "reflect the personality of its author by expressing their free creative decisions" (para. 18). For copyright to subsist, the AI model must be "closer to a mere tool than to an independent instrument of creation" (para. 21), meaning it is not enough to merely "trigger" a process or select from multiple outputs. The Court noted that "merely generally formulated, open-ended instructions" leave the actual "design decision" to the AI, thus failing the test for authorship.[53]

In September 2025, Italy ammended its Copyright Law and clarified that AI-assisted works are protected only where there is a substantial human contribution.[54]

---

[48] Ju Yoen (no 44) 211–222.

[49] Beijing Internet Court Civil Judgment (2023) Jing 0491 Min Chu No. 11279, available at https://english.bjinternetcourt.gov.cn/pdf/BeijingInternetCourtCivilJudgment112792023.pdf. This is China's first AI-image copyright case.

[50] An appeal is pending.

[51] AG München, Judgment of Feb 13, 2026 – 142 C 9786/25.

[52] § 2 Abs. 2 UrhG.

[53] https://legalblogs.wolterskluwer.com/copyright-blog/the-prompter-as-copyright-author-lessons-from-munich .

[54] Italy, Law No. 132 of 23 September 2025, *Disposizioni e deleghe al Governo in materia di intelligenza artificiale* ("Provisions and Delegations to the Government on Artificial Intelligence"), G.U. No. 223 (25 Sept. 2025). Article 25 of Law No. 132/2025 amended the Italian Copyright Act. Specifically it amended Article 1 of the Italian Copyright Law to clarify that AI-assisted works are protected only where there is a substantial/intellectual human contribution, and inserted Article 70-septies concerning AI-related text and data mining.

The Way Forward

The basic premise of copyright law is originality, which is inextricably linked to human authorship. To the extent AI software is used as a tool for an author to exercise their creative choices and mark the work with their personal touch (according to EU law), and this creativity is evident in the output (which comes within the domain of literature, arts, or science), this output may qualify as a copyrighted work. Specifically, it is the user who is entitled to authorship, not the programmer of the software, unless cooperation between the two has taken place.

The above applies to AI-assisted rather than AI-generated works. The latter category refers to productions that emerge from the 'black box' or 'off the shelf' applications where the user has no control or cannot predict the final output. If, of course, the final output is processed creatively or 'worked further', and this additional input is creative and thus original, it is this additional input that is protected.

The 2020 study 'Challenges to the Intellectual Property Rights Framework' (The Institute for Information Law (IVIR)) commissioned by the European Commission, concurred with *Painer*[55] and concluded that one should distinguish between 'three distinct phases of the creative process in AI-assisted production: "conception" (design and specifications), "execution" (producing draft versions) and "redaction" (editing, finalisation)'.[56] In these instances, any creative input in any of those three stages would allow a work to be protected, but then again, this would make it an AI-assisted work and not an AI-generated work.[57]

The use of prompts (in their current form) does not qualify as human creative input in the phase of 'conception'; prompts are ideas, and ideas are not protected under copyright. Iterative prompts also fall in this category. In the future, prompts may become more detailed, allowing perhaps users to dictate the output and predict the outcome, but this is not currently the case.

Difficulties apply when one is to distinguish in practice between AI-assisted works and AI-generated works. Difficulties may also apply when one makes use of Article

[55] Case C-145/10, *Eva-Maria Painer v. Standard VerlagsGmbH and others*, judgment of 7 March 2013.

[56] The 2020 study 'Challenges to the Intellectual Property Rights Framework' (The Institute for Information Law (IVIR)) <https://op.europa.eu/en/publication-detail/-/publication/394345a1-2ecf-11eb-b27b-01aa75ed71a1/language-en>.

[57] The authors of the study reach the conclusion that the requirement of human intellectual effort excludes from copyright protection outputs produced without any human intervention and, among these, outputs wholly generated by an AI system without any human intellectual effort. https://copyrightblog.kluweriplaw.com/2020/12/16/trends-and-developments-in-artificial-intelligence-challenges-to-copyright/.

5(1) of the Enforcement Directive,[58] which provides for a presumption of authorship; in particular the author of a literary or artistic work, in the absence of proof to the contrary, will be regarded as such, and consequently be entitled to institute infringement proceedings, if their name appears on the work. But this is a matter of proof and not of legal doctrine.

There are legal means, though, that can be used for the protection of AI-generated works, one of them being related rights. It could be argued that it is in this category that the legal form of computer-generated works falls since, as we discussed above, EU copyright law is not compatible with systems that allow non-humans to qualify as authors. Apart from that, there are also the provisions for phonogram producers, audiovisual producers, and broadcasters, depending on the subject matter and its focus. The *sui generis* right for databases may also prove useful if the subject matter at hand involves a substantial investment in either the obtaining, verification, or presentation of the data included in the database. The protection of data as such is a separate issue.

The prevailing view in the EU is that AI-generated works are authorless and there is no economic evidence that justifies (at least up to now) legal intervention. It makes no sense to afford absolute and exclusive rights to producers of non-original output, i.e., to someone who had no control over what came out of the application. The thinking should remain that something is unique if a natural person creates it, otherwise it is an industrialised product that can be repeated, and which does not justify exclusivity and (at least) legal monopoly.[59]

References

Bently L., B. Sherman, D. Gangjee, and Ph. Johnson (2022) *Intellectual Property Law* (6th ed, OUP)

Bently L., B.Sherman, D. Gangjee, and Ph. Johnson, Intellectual Property Law, 6th Ed. OUP, 2022.

Caddick N., G. Harbottle and U. Suthersanen, Copinger and Skone James on Copyright, 18th ed., Sweet & Maxwell, 2000

Denicola R. 'Ex Machina: Copyright Protection for Computer-Generated Works', (2016) 69 Rutgers University Law Review 251 https://ssrn.com/abstract=3007842.

[58] Directive 2004/48/EC on the Enforcement of Intellectual Property Rights.

[59] A recent EU document on AI (apart from the EU AI Act) is the European Parliament, Committee on Legal Affairs, Report on Copyright and Generated Artificial Intelligence - Opportunities and Challenges (2025/2058 (INI)), 25/2/2026, Rapporteur Axel Voss.

Ercolani St., "Computer-Generated Works" (Elgar Encyclopedia of Intellectual Property Law, P. Torremans, I. Stamatoudi, P. Yu and J. Jutte, 2025)

Garnett K., G. Davies and G. Harbottle (2010) *Copinger and Skone James on copyright*. UK: 16th ed., Sweet & Maxwell; "The concept of originality in Greece" in Indranath Gupta, Vishwas H. Devaiah and Manveen Singh (eds), "Handbook on Originality in Copyright: Cases and Materials", Springer Nature, 2024 https://link.springer.com/referencework/10.1007/978-981-19-1144-6

Gervais D.J., 'The compatibility of the skill and labour standard with the Berne Convention and the TRIPs Agreement' (2004) 26 *EIPR* 75, 79

Ginsburg J. & L. Budiardjo, *Authors and Machines*, 34 Berkeley Tech. L. J. 343 (2019). Available at https://scholarship.law.columbia.edu/faculty_scholarship/2323

Griffiths, J. (2011) 'Infopaq, BSA and the "Europeanisation" of United Kingdom copyright law', *Media & Arts Law Review*, Vol. 16 <http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1777027>.

Grimmelmann J., "There's No Such Thing as a Computer-Authored Work - And It's a Good Thing, Too," (2016) 39 Columbia Journal of Law & the Arts https://scholarship.law.cornell.edu/facpub/1483/

Ju Yoen L, "Artificial Intelligence Cases in China: Feilin v. Baidu and Tencent Shenzhen v. Shanghai Yingxin", China & WTO Rev. 2021:1; 211-222 https://pdfs.semanticscholar.org/b61c/cb98fd748b7ee654895ee133d5363e4d8556.pdf

Kan He, "Feilin v. Baidu: Beijing Internet Court tackles protection of AI/software-generated work and holds that copyright only vests in works by human authors", IPKat Blog, 9.11.2019, https://ipkitten.blogspot.com/2019/11/feilin-v-baidu-beijing-internet-court.html.

Laddie, Prescott and Vitoria: The Modern Law of Copyright, 5th ed., Lexis Nexis, 2018

Lee Jyh-An (2021) 'Computer-generated Works under the CDPA 1988' in Lee Jyh-An, R. Hilty, and Liu Kung-Chung Liu (eds), *Artificial Intelligence and Intellectual Property* (Oxford University Press) 177

Maggiore M. (2018) 'Artificial intelligence, computer generated works and dispersed authorship: spectres that are haunting copyright' in Bonadio E. and N. Lucchi (eds), *Unconventional Copyrigh*t (Edward Elgar Publisher)

McCutcheon J., (2013) 'Curing the Authorless Void: Protecting Computer-Generated Works following IceTV and Phone Directories' 37(1) Melbourne University Law Review http://www5.austlii.edu.au/au/journals/MelbULawRw/2013/9.html

Ramalho A., 'Will Robots Rule the (Artistic) World? A Proposed Model for the Legal Status of Creations by Artificial Intelligence Systems' (2017) 21 Journal of Internet Law 11, 17.

Ricketson S. and J. Ginsburg, (2006) *International Copyright and Neighbouring Rights: The Berne Convention and Beyond* (2nd ed, OUP)

Ricketson Sam, 'People or Machines? The Berne Convention and the Changing Concept of Authorship' (1991) 16(1) Columbia Journal of Law and the Arts 1..

Rosati, E. (2013) *Originality in EU Copyright*. UK: Edward Elgar.

Samuelson P. (1985) 'Allocating ownership rights in computer-generated works' 47, University of Pittsburgh Law Review 1185

Simone D. "Authorship" (Elgar Encyclopedia of Intellectual Property Law, P. Torremans, I. Stamatoudi, P. Yu and J. Jutte, 2025)

Stamatoudi I., "Originality under EU Copyright Law" in P. Torremans (ed), Copyright Law: A Handbook of Contemporary Research, Edward Elgar Publishing, Cheltenham (UK) - Northampton (US), 2017, 57

Stamatoudi, I., "'More' Originality for Cypriot Copyright Law according to the CJEU's Case Law", The Cyprus Review, Spring 2018, Vol. 30:1, pp. 207-231

Veiksa I., 'Protection of computer-generated works' (2021) 10(1) International Journal of Artificial Intelligence (IJ-AI) 234

Xiaoshuai Ren, Tencent Dreamwriter: Decision of the People's Court of Nanshan (District of Shenzhen) 24 December 2019-Case No. (2019) Yue 0305 Min Chu No. 14010, (2020) 51(1) INt'l reV. INtell. ProP. & comPetItIoN L. 652-9.

The Employment Status of Providers in the Gig Economy

Konstantinos Dimarellis[1]

Abstract

The gig economy is continually expanding and now accounts for a growing share of service provision. Consumers increasingly rely on digital platforms to connect with providers who offer the services they seek. These platforms typically classify service providers as freelancers rather than employees. However, work performed through platforms often exhibits strong characteristics of subordinate labour. This article seeks to outline the key parameters for classifying gig workers and for regulating the conditions under which their work is performed.

Keywords: gig economy; provider; subordination; freelancer; platform

Introduction

The rising dominance of technology has instigated new concerns and expectations in the labour market. It has also caused new challenges for employees who increasingly use technological means to provide their work.[2] The gig economy has emerged as a new form of entrepreneurship.[3] Ensuring fair working conditions is a necessity[4] to balance out the rapid development of the gig economy.[5]

The article attempts a doctrinal legal analysis of the employment status of providers in the gig economy. It starts by analysing the concepts of 'crowdsourcing' and the 'gig economy'. It then contrasts the status of freelancers with that of employees

[1] Assistant Professor of Labour Law, University of Nicosia.

[2] Geoffrey Parker, Marshall Van Alstyne & Sangeet Paul Choudary, *Platform revolution* (W.W. Norton & Company, 2017) 3; Izabela Florczak & Marta Otto, 'Precarious work and labour regulation in the EU: current reality and perspectives', in Jeff Kenner, Izabela Florczak & Marta Otto (eds), *Precarious work. The challenge for Labour Law in Europe* (Edward Elgar Publishing, 2019) 2; Ursula Huws, *The making of a cybertariat: Virtual work in a real world* (Monthly Review Press, 2003) 45.

[3] Manuel Castells, *The rise of the network society* (2nd edn., Wiley-Blackwell, 2009) 77; Orly Lobel, 'The law of the platform' (2016) 101(1) *Minesota Law Review* 89.

[4] Brhmie Balaram, Josie Warden & Fabian Wallace-Stephens, *Good gigs: A fairer future for the UK's gig economy* (Royal Society of Arts, 2017) 34.

[5] Tim O'Reilly, *What's the future of work?* (O'Reilly Media Inc., 2016) 89.

and examines the choice that gig economy platforms that facilitate the provision of services have made to designate providers as freelancers. The article focuses on the criterion of subordination, which is often found in platform work. The article concludes with an opinion about the appropriate status of platform providers and recommendations on the regulation of platform work.

1. From Outsourcing to Crowdsourcing

Outsourcing constitutes the permanent assignment of an activity from the company to a third party that has previously been carried out by the company.[6] The third party uses its employees for the execution of the activity and contributes to the achievement of the business goals. Companies outsource their activities to reduce operating costs and exempt themselves from labour and social insurance legislation, since the company is not considered the employer of the personnel used to carry out the outsourced activities.[7]

Crowdsourcing[8] is the digital form of outsourcing, where the company (platform) assigns online to a provider, who does not necessarily belong to its staff, a task or a service that is usually repetitive and low-skilled.[9] Technological advancements have

[6] Gerhard Bosch, 'Towards a new standard employment relationship in western Europe' (2004) 42(4) *British Journal of Industrial Relations British Journal of Industrial Relations* 627.

[7] Arne Kalleberg, 'Precarious work, insecure workers: employment relations in transition' (2008) 74(1) *American Sociological Review* 2.

[8] The term was first introduced by Jeff Howe, 'The rise of crowdsourcing', *Wired Magazine* (2006) 14(6) 176–183.

[9] Daren Brabham, *Crowdsourcing* (The MIT Press, 2013) 3 and 19; Isabelle Daugareilh, Cristophe Degryse & Philippe Pochet, *The platform economy and social law: Key issues in comparative perspective* (ETUI, 2019) 21; Mark Graham, Isis Hjorth & Vili Lehdonvirta, *Digital labour and development: impacts of global digital labour platforms and the gig economy on worker livelihoods* (ETUI, 2017) 137; Gionata Cavallini & Matteo Avogaro, '"Digital work" in the "platform economy": the last (but not least) stage of precariousness in labour relationships', in Jeff Kenner, Izabela Florczak & Marta Otto (eds), *Precarious work. The challenge for Labour Law in Europe* (Edward Elgar Publishing, 2019) 179; Jeremias Prassl & Martin Gruber-Risak, 'Uber, TaskRabbit & Co: Platforms as employers? Rethinking the Labour Law of Crowdwork', in Kerstin Ahlberg & Niklas Bruun (eds), *The new foundations of Labour Law* (PL Academic Research, 2017) 99; Mijke Houwerzijl, 'European Union cross-border worker mobility in light of digitalization of labour – More fragmentation underway?', in Kerstin Ahlberg & Niklas Bruun (eds), *The new foundations of Labour Law* (PL Academic Research, 2017) 238; Dave Murray-Rust & Dave Robertson, 'Bootstrapping the next generation of social machines', in Wei Li, Michael Huhns, Wei-Tek Tsai & Wenjun Wu (eds), *Crowdsourcing; Cloud-based software development* (Springer, 2015) 40; Birgitta Bergvall-Kareborn & Debra Howcroft, 'Amazon Mechanical Turk and the commodification of labour' (2014) 29(3) *New Technology Work and Employment* 214; Enrique Estellés Arolas & Fernando González Guevara, 'Towards an integrated crowdsourcing definition' (2012) 38(2) *Journal of Information Science* 1.

made crowdsourcing even more popular, with online task assignment and delivery becoming more efficient on a global level.[10] A consumer expresses interest in receiving a service and the platform mediates to assign the task to a provider who has expressed interest in providing that service.[11] The provider is not employed by nor part of the platform's staff;[12] as a result, the provider is deprived of the protection of labour and social security legislation.[13] Platforms are thus able to employ minimal permanent staff[14] and turn to crowdsourcing in order to find willing casual staff to meet their needs at any given time.[15]

The acquisition of qualifications and skills related to technology has coincided with the desire of employers to turn to outsourcing and nowadays the online assignment and provision of a service is obvious.[16] Undoubtedly, crowdsourcing has allowed various categories of people, such as the unemployed, young people, and immigrants, who were largely excluded from the labour market and fall into the cate-

[10] Vassilis Hatzopoulos, *The collaborative economy and EU Law* (Hart Publishing, 2018) 149; Klaus Schwab, *The fourth industrial revolution* (Penguin, 2017) 41; Brabham (no 9) 10; Ursula Huws, *New forms of platform employment or crowd work* (Foundation for European Progressive Studies, 2016) 67; Alain Strowel & Wouter Vergote, 'Digital platforms: to regulate or not to regulate?', in Bram Devolder (ed.), *The platform economy* (Intersentia, 2019) 8; Antonio Aloisi, 'Commoditized workers: Case study research on Labor Law issues arising from a set of "on-demand/gig economy" platforms' (2016) 37(3) *Comparative Labor Law & Policy Journal* 659; Alek Felstiner, 'Working the crowd: employment and labor law in the crowdsourcing industry' (2011) 32(1) *Berkeley Journal of Employment and Labor Law* 145.

[11] Jeremias Prassl, *Humans as a service*, (Oxford University Press, 2018) 11; Werner Wobbe, *Digitalisation of work and the social impact of the platform economy* (Foundation for European Progressive Studies, 2016) 17–18; Sahil Deva & Veronica Wasza, 'Testbirds. Software testing solutions in the crowdsourcing industry', in Werner Wobbe, Elva Bova & Catalin Dragomirescu-Gaina (eds), *The digital economy and the single market – Employment prospects and working conditions in Europe* (Foundation for European Progressive Studies, 2016) 57; Irene Mandl, *Working conditions in crowd employment and ICT-based mobile work* (Foundation for European Progressive Studies, 2016) 115; David Durward, Ivo Blohm & Jan Marco Leimeister, 'Principal forms of crowdsourcing and crowd work', in Werner Wobbe, Elva Bova & Catalin Dragomirescu-Gaina (eds), *The digital economy and the single market – Employment prospects and working conditions in Europe* (Foundation for European Progressive Studies, 2016) 39; Ursula Holtgrewe, *Working in the low-paid service sector: what is to be learned from the analogue world?* (Foundation for European Progressive Studies, 2016) 94.

[12] Durward, Blohm & Leimeister (no 11) 46–47.

[13] Prassl (no 11) 4; Wobbe (no 11) 24.

[14] Schwab (no 10) 10; Huws (no 2) 57.

[15] Hatzopoulos (no 10) 154; Valerio De Stefano & Antonio Aloisi, *European legal framework for "digital labour platforms"* (European Commission, 2018) 38; Valerio De Stefano, *The rise of the "just-in-time workforce": on-demand work, crowd work and labour protection in the "gig-economy"* (International Labour Office, 2016) 3.

[16] Huws (no 10) 67.

gory of 'outsiders', to work and gain an income.[17] The rapid growth of crowdsourcing has contributed decisively to the emergence of the gig economy and sparked discussions about its influence on the labour market.[18]

2. The Emergence of the Gig Economy

The gig economy represents a business model where the overall service or work is divided into several parts (gigs)[19] and each gig is assigned to a provider using modern technological means, mainly through a platform, which creates an open market for temporary use of goods and services.[20] The division of the service into smaller parts, especially in the case of services whose execution does not necessarily require specialised knowledge or skills,[21] and the concentration of the parts that shape the overall service are the key aspects of the function of the gig economy.[22]

Gig economy businesses have introduced a new form of organisation and provision of services on an international level.[23] A platform is open to the consumers who desire the provision of a service and an impersonal,[24] alternating,[25] heterogeneous,[26]

[17] Steven Hill, *Raw Deal: How the "Uber Economy" and Runaway Capitalism Are Screwing American Workers* (St. Martin's Press, 2015) 129; Holtgrewe (no 11) 97; James Manyika et al., *Independent work: choice, necessity, and the gig economy* (McKinsey & Company, 2016) 73–74.

[18] Manyika et al. (no 17) 66; Mandl (no 11) 111.

[19] The term 'gig' originated from the field of music performance as another word for 'concert'. The term was then used to denote work over a specific period. In the labour market, it is used to denote an economy of micro-tasks, where a multitude of providers carry out piecemeal activities and are paid 'by the piece'.

[20] Parker, Van Alstyne & Choudary (no 2) 19; Hill (no 17) 2 and 39; Balaram, Warden & Wallace-Stephens (no 4) 10; Durward, Blohm & Leimeister (no 11) 40-41; Mandl (no 11) 115; Friso Bostoen, 'Competition Law in the peer-to-peer economy', in Bram Devolder (ed.), *The platform economy* (Intersentia, 2019) 144; Cavallini & Avogaro (no 9) 177; Lobel (no 3) 108.

[21] Prassl (no 11) 75-76; Mandl (no 11) 116; Christophe Degryse, *Digitalisation of the economy and its impact on labour markets* (ETUI, 2016) 36.

[22] Parker, Van Alstyne & Choudary (no 2) 280; Durward, Blohm & Leimeister (no 11) 49–50. See also case BAG of 25.8.2020 – 9 AZR 373/19, NZA 2020.1537.

[23] Huws (no 10) 65.

[24] Trebor Scholz, *Uberworked and underpaid: How workers are disrupting the digital economy* (Polity Press, 2017) 26; Felstiner (no 10) 145.

[25] Ognjen Scekic, Hong-Linh Truong & Schahram Dustdar, 'Supporting multilevel incentive mechanisms in crowdsourcing systems: an artifact-centric view', in Wei Li, Michael Huhns, Wei-Tek Tsai & Wenjun Wu (eds), *Crowdsourcing; Cloud-based software development* (Springer, 2015) 93.

[26] S.A.M. Geerts, *Discovering crowdsourcing: theory, classification and directions for use*, (TUE, 2009) 38.

and globally distributed[27] crowd of providers who wish to provide a service and contribute to the economic activity carried by the platform.[28] The platform functions as the 'intermediary', connecting the consumer looking for the specific service and the provider who is able to provide it.

The gig economy restructures the way that the three parties (company, consumer, and provider) operate. First, the companies eschew conventional models of entrepreneurship, instead adopting a decentralised system of service provision, where the company lacks concrete facilities and a structured network of employees.[29] Consumers turn to the gig economy to enjoy services at cheaper cost.[30] Finally, workers turn to the gig economy as a means of combating high unemployment rates and exclusion from the labour market.[31] The gig economy has thus evolved into a global business phenomenon, with an increasing variety of services provided through platforms. As a result, long-standing companies with conventional models of entrepreneurship are adapting their activity to the gig economy model.[32]

3. The Legal Status of the 'Employee' vs. the 'Freelancer'

An employee is a person who provides their work to an employer under an employment contract in return for a salary. A key element of the employment contract is 'subordination'—the employer's managerial right to instruct and order the employee regarding the type, manner, time, and place of their work, as well as the employer's power to control and ascertain the correct execution of work. The concept of the employee in EU law has been shaped through the jurisprudence of the European Court of Justice (ECJ). The ECJ has repeatedly pointed out that there is no single defini-

[27] Brabham (no 9) 14, Siou Chew Kuek, Cecilia Paradi-Guilford, Toks Fayomi, Saori Imaizumi & Panos Ipeirotis, *The global opportunity in online outsourcing* (World Bank Group, 2015) 1.

[28] O'Reilly (no 5) 20; Jörg Flecker & Annika Schönauer, 'The production of "placeness": Digital service work in global value chains', in Jörg Flecker (ed.), *Space, place and global digital work* (Palgrave Macmillan, 2016) 23; Wei Li, Wei-Tek Tsai & Wenjun Wu, 'Crowdsourcing for large-scale software development', in Wei Li, Michael Huhns, Wei-Tek Tsai & Wenjun Wu (eds), *Crowdsourcing; Cloud-based software development* (Springer, 2015) 3.

[29] Arun Sundararajan, *The sharing economy: The end of employment and the rise of crowd-based capitalism* (The MIT Press, 2016) 172.

[30] Deva & Wasza (no 11) 60.

[31] Hill (no 17) 12 and 109.

[32] Lobel (no 3) 95.

tion for the concept of employee,[33] but also expressed the need to formulate such a definition to prevent national approaches that may exclude certain persons, and, by extension, those persons' protection under labour law.[34] The ECJ interpreted for the first time the concept of the employee according to Art. 45 of the Treaty on the Functioning of the European Union (TFEU) in the *Lawrie-Blum* case.[35] The employee was defined as a person who provides services for a certain time to another person under their direction and receives remuneration.[36] Three conceptual elements could thus be derived: a) payment of remuneration; b) subordination of the employee to the employer's authority; and c) the genuine and authentic activity of the employee.

In contrast, a freelancer provides their services under a contract of independent services, without being subject to the instructions, orders, and control of a counterparty, in particular with regard to the time and manner of providing the service.[37] The freelancer maintains a high degree of autonomy, determines their own working hours and remuneration, chooses their clientele freely,[38] has their own organisational structure, maintains their own technical means, can employ staff, and carries the risks and enjoys the benefits of their activity.[39]

The classification of a provider as an employee would carry the basic legal consequence of the application of labour and social insurance legislation. As a result, the platform would be bound by certain aspects of labour legislation, such as minimum wage, working time limits, health and safety rules, and dismissal restrictions. Thus, most platforms prefer to employ providers as freelancers to avoid liability as employers.

[33] See cases C-85/96 of 12.5.1998 thought 31, C-543/03 of 7.6.2005 thought 27, C-256/01 of 13.1.2004 thought 63.

[34] Martin Gruber-Risak & Thomas Dullinger, *The concept of "worker" in EU law* (ETUI, 2018) 41.

[35] See case C-66/85 of 3.7.1986 thought 17.

[36] Gruber-Risak & Dullinger (no 34) 11.

[37] Mark Freedland & Nicola Kountouris, *The legal construction of personal work relations* (Oxford University Press, 2011) 269 ff.; Brishen Rogers, 'Employment rights in the platform economy: Getting back to basics' (2016) 10(2) *Harvard Law & Policy Review* 485.

[38] Hatzopoulos (no 10) 107.

[39] Manyika et al. (no 17) 2 and 20 ff.; Seth Harris & Alan Krueger, *A proposal for modernizing labor laws for twenty-first-century work: The "independent worker"* (The Hamilton project, 2015) 7; Rogers (no 37) 485.

4. Platforms and 'Uberisation'

The lack of legislative framework regarding the status of gig economy providers has allowed the platforms to choose the legal status of providers according to their interests.[40] Platforms opt for independent services contracts that exempt them from labour laws.[41] This choice can be captured by the term 'Uberisation', which derives from the pioneering Uber platform.[42] When they join the platform, providers must agree to a condition stating that they work as freelancers,[43] even though their legal status is in fact defined by jurisdictional authority, who consider, analyse, and assess the circumstances under which their work is provided.[44]

The industry terminology exacerbates the problem, as platforms favour terms used in independent services contracts.[45] The platform assumes the role of the mediator and not the employer.[46] The provider is presented not as an employee but as self-employed: a freelancer, independent contractor, partner, or microentrepre-

[40] Debra Howcroft & Birgitta Bergvall-Kareborn, 'A typology of crowdwork platforms' (2019) 33(3) *Work Employment and Society* 30; Vanessa Katz, 'Regulating the sharing economy' (2015) 30 *Berkeley Technology Law Journal* 1072.

[41] Hill (no 17) 25, 88 and 143; Parker, Van Alstyne & Choudary (no 2) 250; Rebecca Smith & Sarah Leberstein, *Rights on demand: Ensuring workplace standards and worker security in the on-demand economy* (National Employment Law Project, 2015) 4; Gruber-Risak & Dullinger (no 34) 12; De Stefano & Aloisi (no 15) 42; Joanna Unterschütz, 'Digital work – real bargaining: how can the sustainability of social dialogue be ensured in the digital era?', in Jeff Kenner, Izabela Florczak & Marta Otto (eds), *Precarious work. The challenge for Labour Law in Europe* (Edward Elgar Publishing, 2019) 226; Jeff Kenner, 'Uber drivers are "workers": the expanding scope of the "worker" concept in the UK's gig economy', in Jeff Kenner, Izabela Florczak & Marta Otto (eds), *Precarious work. The challenge for Labour Law in Europe* (Edward Elgar Publishing, 2019) 199; Niklas Bruun, 'The old and new foundations of Labour Law', in Kerstin Ahlberg & Niklas Bruun (eds), *The new foundations of Labour Law* (PL Academic Research, 2017) 19; Lobel (no 3) 105; Harris & Krueger (no 39) 5–7; Howcroft & Bergvall-Kareborn (no 40) 24.

[42] Hatzopoulos (no 10) 158; Hill (no 17) 6; De Stefano & Aloisi (no 15) 13; De Stefano (no 15) 12; Bostoen (no 20) 145; Victoria Daskalova, 'Regulating the new self-employed in the Uber economy: What role for EU Competition Law?' (2018) 19(3) *German Law Journal* 479; Howcroft & Bergvall-Kareborn (no 40) 24; Graham, Hjorth & Lehdonvirta (no 9) 140; Aloisi (no 10) 663.

[43] De Stefano & Aloisi (no 15) 17.

[44] Colin Williams & Frédéric Lapeyre, *Dependent self-employment: Trends, challenges and policy responses in the EU* (International Labour Office, 2017) 36; Kenner (no 41) 219; Aloisi (no 10) 663. See cases BAG of 25.8.2020 – 9 AZR 373/19, NZA 2020.1537, *Aslam and Farrar v Uber*, 2202550/2015, 87-89, *London Employment Tribunal, Uber v Aslam and Farrar*, UKEAT/0056/17/DA, 116, *Employment Appeal Tribunal, Dewhurst v Citysprint UK Ltd*, 2202512/2016, 49, London Employment Tribunal.

[45] De Stefano & Aloisi (no 15) 11.

[46] Prassl (no 11) 4; Holtgrewe (no 11) 99.

neur.[47] The provider does not sign an employment contract, but a supplier agreement; the provider does not receive salary, but issues invoices; the provider is not fired, but their cooperation with the platform may be terminated. Finally, platforms use terms such as 'ride', 'task', or 'favour' instead of 'work'.[48]

Platforms have preferred to classify workers as freelancers from the start. This classification has been facilitated by modern technological means and methods:[49] services are provided immediately, without instructions from the platform, and at the initiative of the provider. These conditions favour the platforms' argument that they are not employers, but simply mediate between providers and consumers[50] or provide the necessary software to facilitate the provision of the service.[51] Therefore, it is particularly important to identify the appropriate legal status of the providers based on the conditions under which they provide their work and especially based on the existence of criteria indicating subordination, regardless of the status under which the platforms prefer to classify providers.

5. Defining Subordination in the Digital Labour Market

As mentioned, subordination is the key element of an employment contract. However, economic, labour, social, and technological developments have made it difficult to identify the element of subordination.[52] The distinction between subordinate work and independent services provision is at times unclear—employees may take initiative when providing their work, while freelancers may be controlled through instructions and supervision from their counterparty.[53]

The platforms claim that they are merely providing the software that facilitates the connection between providers and consumers. However, this simple mediating

[47] Sundararajan (no 29) 177 ff.; Balaram, Warden & Wallace-Stephens (no 4) 10 and 35; Alex Rosenblat & Luke Stark, 'Algorithmic labor and information asymmetries: a case study of Uber's drivers' (2017) 10 *International Journal of Communication* 3761.

[48] Prassl (no 11) 42 ff.; De Stefano (no 15) 5.

[49] Brhmie Balaram, *Fair share: Reclaiming power in the sharing economy* (Royal Society of Arts, 2016) 12.

[50] Parker, Van Alstyne & Choudary (no 2) 249, Balaram (no 49) 42, Balaram, Warden & Wallace-Stephens (no 4) 35; Smith & Leberstein (no 41) 5.

[51] Sundararajan (no 29) 159.

[52] Prassl (no 11) 100; Simon Deakin, The many futures of the contract of employment, in Joanne Conaghan, Richard Michael Fischl & Karl Klare (eds), *Labour Law in an era of globalization: Transformative practices and possibilities* (Oxford University Press, 2002) 189; Rogers (no 37) 485.

[53] Daugareilh, Degryse & Pochet (no 9) 116; Daskalova (no 42) 476.

role often fails to correspond to reality.[54] The fact that the platforms operate in a digital environment does not negate the fact that they exercise a managerial right, since they define the conditions under which a service is provided.[55] The platform determines the conditions under which one can join the platform and the provider is obliged to agree to a set of conditions in order to use the platform and provide a service.[56] In fact, the platform often reserves the right to unilaterally change the working conditions by setting new terms, which are presumed to have been acknowledged by the provider upon entry to the platform.[57]

The platform has the power to control the execution of the service and whether the service corresponds to the desired parameters.[58] The fact that the control is exercised digitally does not undermine its value as a criterion of subordination.[59] Supervision and monitoring of providers may also be rigorous, and may even include the imposition of sanctions should the provider not comply with instructions;[60] often, payment depends on the provider's compliance with the instructions given by the platform.[61] This high level of control on the part of the platform through technological means[62] can violate providers' personal data,[63] especially in cases where the service is provid-

[54] Daugareilh, Degryse & Pochet (no 9) 22; De Stefano (no 15) 15; Cristiano Codagnone, Federico Biagi & Fabienne Abadie, *The passions and the interests: Unpacking the "sharing economy"* (JRC Research reports, 2016) 34; Nicolas Van Damme, 'Law evasion in the platform economy: The Uber case', in Bram Devolder (ed.), *The platform economy* (Intersentia, 2019) 110; Cavallini & Avogaro (no 9) 179-180; Prassl & Gruber-Risak (no 9) 97; Aloisi (no 10) 655; Felstiner (no 10) 149; Tarleton Gillespie, 'The politics of "platforms"' (2010) 12(3) *New Media & Society* 356.

[55] Bergvall-Kareborn & Howcroft (no 9) 220. See also cases BAG of 25.8.2020 – 9 AZR 373/19 NZA 2020.1537, FNV v Uber BV ECLI:NL:RBAMS:2021:5029, *O' Connor v Uber Techs.* – 82 F. Supp. 3d 1133 (N.D. Cal. 2015). Contrary Hatzopoulos (no 10) 106.

[56] Anna Wood, *Algorithmic management: Consequences for work organization and working conditions* (Joint Research Center, 2021) 1; Aloisi (no 10) 671.

[57] Aloisi (no 10) 676. See case *Aslam and others v Uber BV and others* [2017] IRLR 4, thought 35.

[58] Wobbe (no 11) 20; Degryse (no 21) 40; Prassl & Gruber-Risak (no 9) 116; Katz (no 40) 1071; Felstiner (no 10) 175.

[59] Massimiliano Delfino, 'Work in the age of collaborative platforms between innovation and tradition' (2018) 9(4) *European Labour Law Journal* 348. See case *FNV v Uber BV* ECLI:NL:RBAMS:2021:5029.

[60] Prassl (no 11) 61-63; Smith & Leberstein (no 41) 5; Jon Messenger, *Working time and the future of work* (ILO Research paper series, 2018) 23; Kenner (no 41) 219; Prassl & Gruber-Risak (no 9) 114. See cases *Aslam and others v Uber BV and others* [2017] IRLR 4, thought 55, ECLI:FR:CCASS:2020:SO00374, n. 19-13.316, *O' Connor v Uber Techs.* – 82 F. Supp. 3d 1133 (N.D. Cal. 2015).

[61] See case BAG of 1.12.2020 – 9 AZR 102/20.

[62] Felstiner (no 10) 175.

[63] Mandl (no 11) 123. For example, the platform Honor controls the provision of home care services by monitoring the home where the employee is sent to work, in order to verify that they arrived on time,

ed digitally.[64] For example, the platform may regularly take screenshots to check that the provider is engaged in the service and then count the working hours in order to calculate the fee.[65] Platforms can also count how many times the provider presses the keyboard per minute or how many times the mouse moves to check that the provider is working hard enough and not taking excessive breaks.[66] These technological practices have given new meaning to the concepts of online surveillance and control.[67]

Control is also related to the evaluation of providers by consumers after the service has been provided.[68] Platforms control, organise, and manage an extensive evaluation system, where consumers are asked to rate and evaluate the provider to establish the latter's level of performance.[69] A positive evaluation does not constitute a simple moral or psychological reward for the provider, but is a basic condition upon which their participation in the platform and future work is decided.[70] The evaluations lead to a categorisation that determines which shifts or jobs the providers get;[71] specifically, a high rating boosts the provider's reputation and leads to more and better assignments.[72] Conversely, the platform retains the power to exercise a quasi-disciplinary role and to impose sanctions on providers according to the evaluations, which may go as far as exclusion from the platform.[73] The platform may deliberately set particularly high standards of performance, which are difficult or impossible to achieve, and, in case of non-response from the providers, may result in the imposition of sanctions.[74] At the same time, providers may be unable to rebut bad reviews or challenge the validity of a review that they consider unfair.[75]

avoided using social media, and were not resting when they were meant to be completing a task (e.g. cooking), see Smith & Leberstein (no 41) 4.

64 Aneesh Aneesh, 'Global labor: Algocratic modes of organization' (2009) 27(4) *Sociological Theory* 349.

65 Prassl (no 11) 56-57; Wood (no 56) 4; De Stefano & Aloisi (no 15) 21; De Stefano (no 15) 17.

66 Hill (no 17) 104-105; Wood (no 56) 5; De Stefano & Aloisi (no 15) 21; Degryse (no 21) 40.

67 Huws (no 2) 53; Felstiner (no 10) 191.

68 Rachel Botsman & Roo Rogers, *What's mine is yours* (Collins, 2011) 217 ff.; Rosenblat & Stark (no 47) 3772; Katz (no 40) 1075. See case *O' Connor v Uber Techs.* – 82 F. Supp. 3d 1133 (N.D. Cal. 2015).

69 Andrew McAfee & Erik Brynjolfsson, *Machine, platform, crowd* (W. W. Norton & Company, 2017) 208; Lobel (no 3) 152; Aloisi (no 10) 662.

70 Botsman & Rogers (no 68) 218; Lobel (no 3) 152.

71 Wood (no 56) 8.

72 De Stefano & Aloisi (no 15) 21; Prassl & Gruber-Risak (no 9) 103; Aloisi (no 10) 671.

73 Sundararajan (no 29) 200; Balaram (no 49) 45; De Stefano & Aloisi (no 15) 21; Howcroft & Bergvall-Kareborn (no 40) 30. See case *Aslam and others v Uber BV and others* [2017] IRLR 4, thought 54.

74 Prassl & Gruber-Risak (no 9) 114; Wood (no 56) 8; Rosenblat & Stark (no 47) 3774–3775.

75 De Stefano & Aloisi (no 15) 22; Katz (no 40) 1119.

The platform usually determines unilaterally the exact fee paid to the provider for the service, as well as its own commission on that fee.[76] The consumer pays the fee to the platform and the platform withholds a commission amount and pays the rest to the provider. The unequal distribution of the bargaining power between the platform and the provider is made evident by the latter's inability to negotiate the amounts of the fee and the commission.[77]

Providers are usually free to choose their working hours, a much-lauded indication of autonomy.[78] Providers themselves often cite this flexibility as an incentive to work in the gig economy.[79] In reality, they are often forced to work for many hours to secure a satisfactory income.[80] While the platform may not set mandatory working hours, it will usually set deadlines within which the provider must provide their service.[81] Platforms may also indirectly force providers to work at certain times—for example, periods of increased demand—by paying higher fees or rating them more positively.[82]

Platforms also often use exclusivity clauses, according to which the provider cannot work on a competing platform or develop their own business activity.[83] An exclusivity clause is an indication of employment and runs contrary to what it means to be a freelancer—a professional who is meant to retain the right to provide their services freely and grow their clientele.[84] Providers are hindered from developing their own

[76] Parker, Van Alstyne & Choudary (no 2) 115 ff.; Huws (no 10) 76; Prassl & Gruber-Risak (no 9) 115; Rosenblat & Stark (no 47) 3763; Katz (no 40) 1071. See cases *Uber BV and others v Aslam and others* [2021] UKSC 5, thought 94 and *O' Connor v Uber Techs*. – 82 F. Supp. 3d 1133 (N.D. Cal. 2015).

[77] Huws (no 10) 80; Hill (no 17) 110; Balaram, Warden & Wallace-Stephens (no 4) 35; Harris & Krueger (no 39) 8; Holtgrewe (no 11) 97; Janine Berg, *Income security in the on-demand economy: Findings and policy lessons from a survey of crowdworkers* (International Labour Office, 2016) 18; R. Kuttner, 'The TaskRabbit economy', *The American Prospect* (2013) 46.

[78] Sundararajan (no 29) 194-195; Balaram, Warden & Wallace-Stephens (no 4) 36; Gruber-Risak & Dullinger (no 34) 12; Rogers (no 37) 494.

[79] Mandl (no 11) 117 and 123; Messenger (no 60) 23; Durward, Blohm & Leimeister (no 11) 51; Rosenblat & Stark (no 47) 3758.

[80] Scholz (no 24) 58.

[81] De Stefano (no 15) 7-8; Felstiner (no 10) 175. See case BAG of 1.12.2020 – 9 AZR 102/20, where it was judged that the provider was obliged to complete the assigned task within two hours.

[82] Wouter Verheyen & Fiona Unz, 'And what about the gig worker? Transport law as a lifeline to gig workers falling outside labour law protection', in Bram Devolder (ed.), *The platform economy* (Intersentia, 2019) 318. See case STS 805/2020 ES:TS:2020:2924.

[83] Hill (no 17) 19-20; Felstiner (no 10) 177 and 192. See case *O' Connor v Uber Techs*. – 82 F. Supp. 3d 1133 (N.D. Cal. 2015).

[84] De Stefano (no 15) 14.

business as platforms do not allow direct interaction with the consumer.[85] Specifically, the provider tends not to have contact with the consumer before or after provision of the service, preventing them from pursuing independent business activity to their financial benefit.[86]

Finally, platforms typically reserve the right to unilaterally expel the provider from the platform, which resembles the employer's right to terminate the employment contract.[87] This power is usually included in the terms to which the provider agrees when entering the platform, which means that the right of termination is the sole initiative of the platform and providers have no right to object.

Platforms often attempt to downplay that the provider is not simply providing a service, but is in fact working under conditions of subordination, as the platform largely determines the terms of the work. The classification of providers depends on the level of subordination, which determines if labour and social insurance legislation is applicable.

Concluding Remarks

The gig economy is promoted as a new form of entrepreneurship that enables more transactions between providers and consumers, giving a wider range of people the opportunity to work and earn income. However, work in the gig economy remains largely unregulated. Providers feel the pressure of competition and end up providing their services at the lowest possible cost, all while classified as freelancers and deprived of labour rights protection. Meanwhile, platforms go beyond acting as mediators that merely connect the consumer with the appropriate provider—they often largely determine the terms and conditions of the work in a way that resembles subordinate employment. That being the case, two major and interdependent questions arise: how should gig workers be classified and how should their work be regulated?

Work in the gig economy is very heterogeneous and displays elements of both subordinate and freelance work, depending on the conditions under which the service is provided. For example, the provision of services differs remarkably between: a) a provider who undertakes the cleaning of a house after the activity has been as-

[85] Felstiner (no 10) 145 and 182. See cases *Uber BV and others v Aslam and others* [2021] UKSC 5 thought 100, ECLI:FR:CCASS:2020:SO00374, n. 19-13.316.

[86] De Stefano (no 15) 5; Prassl & Gruber-Risak (no 9) 113 and 116; Felstiner (no 10) 162. See case Uber BV et als v Aslam et als [2018] EWCA Civ 2748, thought 101.

[87] Degryse (no 21) 35; Prassl & Gruber-Risak (no 9) 113-115; Katz (no 40) 1117; Felstiner (no 10) 193.

signed by a platform that does not specify any other terms for the provision of the work but simply mediates the connection between the provider and the consumer and receives a commission for this connection; and b) a provider who transfers or delivers products to a consumer following particular rules set by a platform, such as the remuneration and the conduct of the provider, the timeframe for performing the work, and the imposition of sanctions on the provider in case of unsatisfactory performance or failure to deliver.

What is an appropriate employment status for platform providers is not always clear and depends on whether and to what extent the parameters that have been established over time as subordination criteria are met, adapted of course to the provision of work through a platform. The coexistence of criteria for subordinate and freelance work is a well-known situation for modern and flexible labour relations. It is commonly said that workers who provide their labour under these conditions fall into a 'grey area' between subordinate and freelance work, where it is not always clear whether they work as employees or freelancers. This causes particular tension in the gig economy due to the looser bond between the platform and the providers and the self-evident development of initiatives during the provision of services. The coexistence of criteria of subordinate and independent work is obvious—platforms set specific terms of service provision and providers are bound by the instructions they receive from the platforms, while at the same time providers have a level of autonomy, ranging from setting working hours to using their own technical means (e.g. a privately owned car). This coexistence of criteria leads to the conclusion that we must establish an intermediate category between employees and freelancers, characterised as 'independent workers'.[88] The intermediate category would not supersede cases where it is clear that the work is either wholly independent (freelancers) or wholly subordinate (employees).

The classification of the provider as employee or freelancer or their inclusion in the intermediate category of independent worker would directly affect the way in which work in the gig economy is regulated. Some States have already chosen to apply labour legislation to platform providers in terms of limiting the duration of fixed-term employment contracts, establishing equal pay for the same work or work of equal value, and obliging platforms to contribute to the social insurance of the providers.[89]

[88] Sundararajan (no 29) 184; Wobbe (no 11) 24; Balaram (no 49) 27; Balaram, Warden & Wallace-Stephens (no 4) 37.

[89] Hill (no 17) 27-28, see also the Regulatory Decree 2017-774 of 4.5.2017 of France.

Establishing the intermediate category of independent workers also leads the need for standalone regulation due to the characteristics under which the platform providers work. On the one hand, they do not provide their work as employees and as a result labour legislation cannot be applied as is. On the other hand, they cannot be completely exempt from the provisions of labour legislation, because a level of subordination exists and therefore there is a need for protection. So, platform providers do need protection, but not to the extent that a regular employee does. In practice, targeted regulation would entail specific derogations from broader labour legislation. This is not about creating legal rules that contradict the core protective regulations of labour law, but about adapting the rules to the special characteristics of the gig economy. The regulation of work in the gig economy could also address issues like the termination of the relationship between the platform and the provider, payment of a minimum wage depending on the hours worked, protection against work accidents, protection against unemployment, and an obligation for platforms to contribute to the social insurance of providers.

The gig economy is growing rapidly and in the coming years is likely to give rise to more platforms, engage more consumers, and employ more providers, firmly establishing itself as a new business model. The lack of legislative framework currently allows platforms to turn to independent services contracts, which, in reality, often force providers to work under conditions of subordination. Identifying the negative aspects of the work in the gig economy and trying to solve the problems that arise is not anathema to the gig economy. A new legislative framework for the regulation of work in the gig economy is instead an urgent necessity to manage the social and economic ramifications of the current situation.

References

Aloisi A., 'Commoditized workers: Case study research on Labor Law issues arising from a set of "on-demand/gig economy" platforms' (2016) 37(3) *Comparative Labor Law & Policy Journal* 653–690.

Aneesh A., 'Global labor: Algocratic modes of organization' (2009) 27(4) *Sociological Theory* 347–370.

Balaram B., *Fair share: Reclaiming power in the sharing economy* (Royal Society of Arts, 2016).

Balaram B., J. Warden & F. Wallace-Stephens, *Good gigs: A fairer future for the UK's gig economy* (Royal Society of Arts, 2017).

Berg J., *Income security in the on-demand economy: Findings and policy lessons from a survey of crowdworkers* (International Labour Office, 2016).

Bergvall-Kareborn B. & D. Howcroft, 'Amazon Mechanical Turk and the commodification of labour' (2014) 29(3) *New Technology Work and Employment* 213–223.

Bosch G., 'Towards a new standard employment relationship in western Europe' (2004) 42(4) *British Journal of Industrial Relations* 617–636.

Bostoen F., 'Competition Law in the peer-to-peer economy', in B. Devolder (ed.), *The platform economy* (Intersentia, 2019) 143–172.

Botsman R. & R. Rogers, *What's mine is yours* (Collins, 2011).

Brabham D., *Crowdsourcing* (The MIT Press, 2013).

Bruun N., 'The old and new foundations of Labour Law', in K. Ahlberg & N. Bruun (eds), *The new foundations of Labour Law* (PL Academic Research, 2017) 17–35.

Castells M., *The rise of the network society* (2nd edn., Wiley-Blackwell, 2009).

Cavallini G. & M. Avogaro, '"Digital work" in the "platform economy": the last (but not least) stage of precariousness in labour relationships', in J. Kenner, I. Florczak & M. Otto (eds), *Precarious work. The challenge for Labour Law in Europe* (Edward Elgar Publishing, 2019) 176–196.

Chew Kuek S., C. Paradi-Guilford, T. Fayomi, S. Imaizumi & P. Ipeirotis, *The global opportunity in online outsourcing* (World Bank Group, 2015).

Codagnone C., F. Biagi & F. Abadie, *The passions and the interests: Unpacking the "sharing economy"* (JRC Research reports, 2016).

Daskalova V., 'Regulating the new self-employed in the Uber economy: What role for EU Competition Law?' (2018) 19(3) *German Law Journal* 461–508.

Daugareilh I., C. Degryse & P. Pochet, *The platform economy and social law: Key issues in comparative perspective* (ETUI, 2019).

De Stefano V., *The rise of the "just-in-time workforce": on-demand work, crowd work and labour protection in the "gig-economy"* (International Labour Office, 2016).

De Stefano V. & A. Aloisi, *European legal framework for "digital labour platforms"* (European Commission, 2018).

Deakin S., 'The many futures of the contract of employment', in J. Conaghan, R. M. Fischl & K. Klare (eds), *Labour Law in an era of globalization: Transformative practices and possibilities* (Oxford University Press, 2002) 176–196.

Degryse C., *Digitalisation of the economy and its impact on labour markets* (ETUI, 2016).

Delfino M., 'Work in the age of collaborative platforms between innovation and tradition' (2018) 9(4) *European Labour Law Journal* 346–353.

Deva S. & V. Wasza, 'Testbirds. Software testing solutions in the crowdsourcing industry', in W. Wobbe, E. Bova & C. Dragomirescu-Gaina (eds), *The digital economy and the single market – Employment prospects and working conditions in Europe* (Foundation for European Progressive Studies, 2016) 56–64.

Durward D., I. Blohm & J.-M. Leimeister, 'Principal forms of crowdsourcing and crowd work', in W. Wobbe, E. Bova & C. Dragomirescu-Gaina (eds), *The digital economy and the single market – Employment prospects and working conditions in Europe* (Foundation for European Progressive Studies, 2016) 39–55.

Estellés Arolas E. & L.-F. González Guevara, 'Towards an integrated crowdsourcing definition' (2012) 38(2) *Journal of Information Science* 189–200.

Felstiner A., 'Working the crowd: employment and labor law in the crowdsourcing industry' (2011) 32(1) *Berkeley Journal of Employment and Labor Law* 143–204.

Flecker J. & A. Schönauer, 'The production of "placeness": Digital service work in global value chains', in J. Flecker (ed.), *Space, place and global digital work* (Palgrave Macmillan, 2016) 11–31.

Florczak I. & M. Otto, 'Precarious work and labour regulation in the EU: current reality and perspectives', in J. Kenner, I. Florczak & M. Otto (eds), *Precarious work. The challenge for Labour Law in Europe* (Edward Elgar Publishing, 2019) 2–21.

Freedland M. & N. Kountouris, *The legal construction of personal work relations* (Oxford University Press, 2011).

Geerts S.A.M., *Discovering crowdsourcing: theory, classification and directions for use*, (TUE, 2009).

Gillespie T., 'The politics of "platforms"' (2010) 12(3) *New Media & Society* 347–364.

Graham M., I. Hjorth & V. Lehdonvirta, *Digital labour and development: impacts of global digital labour platforms and the gig economy on worker livelihoods* (ETUI, 2017).

Gruber-Risak M. & T. Dullinger, *The concept of "worker" in EU law* (ETUI, 2018).

Harris S. & A. Krueger, *A proposal for modernizing labor laws for twenty-first-century work: The "independent worker"* (The Hamilton Project, 2015).

Hatzopoulos V., *The collaborative economy and EU Law* (Hart Publishing, 2018).

Hill S., *Raw Deal: How the "Uber Economy" and Runaway Capitalism Are Screwing American Workers* (St. Martin's Press, 2015).

Holtgrewe U., *Working in the low-paid service sector: what is to be learned from the analogue world?* (Foundation for European Progressive Studies, 2016).

Houwerzijl M., 'European Union cross-border worker mobility in light of digitalization of labour – More fragmentation underway?', in K. Ahlberg & N. Bruun (eds), *The new foundations of Labour Law* (PL Academic Research, 2017) 233–254.

Howcroft D. & B. Bergvall-Kareborn, 'A typology of crowdwork platforms' (2019) 33(3) *Work Employment and Society* 21–38.

Howe J., 'The rise of crowdsourcing' (2006) 14(6) *Wired Magazine* 176–183.

Huws U., *New forms of platform employment or crowd work* (Foundation for European Progressive Studies, 2016).

Huws U., *The making of a cybertariat: Virtual work in a real world* (Monthly Review Press, 2003).

Kalleberg A., 'Precarious work, insecure workers: employment relations in transition' (2008) 74(1) *American Sociological Review* 1–22.

Katz V., 'Regulating the sharing economy' (2015) 30 *Berkeley Technology Law Journal* 1067–1126.

Kenner J., 'Uber drivers are "workers": the expanding scope of the "worker" concept in the UK's gig economy', in J. Kenner, I. Florczak & M. Otto (eds), *Precarious work. The challenge for Labour Law in Europe* (Edward Elgar Publishing, 2019) 197–216.

Li W., W.-T. Tsai & W. Wu, 'Crowdsourcing for large-scale software development', in W. Li, M. Huhns, W.-T. Tsai & W. Wu (eds), *Crowdsourcing; Cloud-based software development* (Springer, 2015) 3–23.

Lobel O., 'The law of the platform' (2016) 101(1) *Minesota Law Review* 87–166.

Mandl I., *Working conditions in crowd employment and ICT-based mobile work* (Foundation for European Progressive Studies, 2016).

Manyika J. et al., *Independent work: choice, necessity, and the gig economy* (McKinsey & Company, 2016).

McAfee A. & E. Brynjolfsson, *Machine, platform, crowd* (W. W. Norton & Company, 2017).

Messenger J., *Working time and the future of work* (ILO Research paper series, 2018).

Murray-Rust M. & D. Robertson, 'Bootstrapping the next generation of social machines', in W. Li, M. Huhns, W.-T. Tsai & W. Wu (eds), *Crowdsourcing; Cloud-based software development* (Springer, 2015) 53–71.

O'Reilly T., *What's the future of work?* (O'Reilly Media Inc., 2016).

Parker G., M. Van Alstyne & S.-P. Choudary, *Platform revolution* (W.W. Norton & Company, 2017).

Prassl J., *Humans as a service*, (Oxford University Press, 2018).

Prassl J. & M. Gruber-Risak, 'Uber, TaskRabbit & Co: Platforms as employers? Rethinking the Labour Law of Crowdwork', in K. Ahlberg & N. Bruun (eds), *The new foundations of Labour Law* (PL Academic Research, 2017) 97–127.

Rogers B., 'Employment rights in the platform economy: Getting back to basics' (2016) 10(2) *Harvard Law & Policy Review* 479–520.

Rosenblat A. & L. Stark, 'Algorithmic labor and information asymmetries: a case study of Uber's drivers' (2017) 10 *International Journal of Communication* 3758–3784.

Scekic O., H.-L. Truong & S. Dustdar, 'Supporting multilevel incentive mechanisms in crowdsourcing systems: an artifact-centric view', in W. Li, M. Huhns, W.-T. Tsai & W. Wu (eds), *Crowdsourcing; Cloud-based software development* (Springer, 2015) 91–111.

Scholz T., *Uberworked and underpaid: How workers are disrupting the digital economy* (Polity Press, 2017).

Schwab K., *The fourth industrial revolution* (Penguin, 2017).

Smith R. & S. Leberstein, *Rights on demand: Ensuring workplace standards and worker security in the on-demand economy* (National Employment Law Project, 2015).

Strowel A. & W. Vergote, 'Digital platforms: to regulate or not to regulate?', in B. Devolder (ed.), *The platform economy* (Intersentia, 2019) 3–30.

Sundararajan A., *The sharing economy: The end of employment and the rise of crowd-based capitalism* (The MIT Press, 2016).

Unterschütz J., 'Digital work – real bargaining: how can the sustainability of social dialogue be ensured in the digital era?', in J. Kenner, I. Florczak & M. Otto (eds), *Precarious work. The challenge for Labour Law in Europe* (Edward Elgar Publishing, 2019) 222–243.

Van Damme N., 'Law evasion in the platform economy: The Uber case', in B. Devolder (ed.), *The platform economy* (Intersentia, 2019) 89–124.

Verheyen W. & F. Unz, 'And what about the gig worker? Transport law as a lifeline to gig workers falling outside labour law protection', in B. Devolder (ed.), *The platform economy* (Intersentia, 2019) 287–324.

Williams C. & F. Lapeyre, *Dependent self-employment: Trends, challenges and policy responses in the EU* (International Labour Office, 2017).

Wobbe W., *Digitalisation of work and the social impact of the platform economy* (Foundation for European Progressive Studies, 2016).

Wood A., *Algorithmic management: Consequences for work organization and working conditions* (Joint Research Center, 2021).

No Content Without Context: Regulating Hate Speech on Social Media[1]

Giovanni Cinà[2]

Abstract

This paper examines the regulation of hate speech on social media through a comparative and context-sensitive lens. It argues that platform moderation policies—dominated by automated tools and content-based assessments—fail to capture the cultural, political, and historical contexts that shape the meaning and harm of speech. Drawing on empirical evidence linking online hate to offline violece, the study contrasts European and US regulatory models, highlighting the risks of both overreach and under-enforcement in context-blind moderation. It proposes a shift toward context-aware governance combining algorithmic precision with human expertise, locally grounded decision-making, and alignment with international human rights standards. The paper concludes that meaningful regulation of hate speech requires a pluralistic, evidence-based approach attuned to global diversity and power imbalances.

Keywords: hate speech; social media; content moderation; contextual regulation; human rights

Introduction

As Ludwig Wittgenstein famously stated, 'the meaning of a word is its use in the language'.[3] This contextual understanding must be the starting point for any analysis of content moderation. A central aspect of the discourse surrounding hate speech revolves around its alleged contribution to real-world violence. Proponents of stricter regulation argue for a link between hate speech and violence, while others question

[1] This paper is based on a speech delivered by the author at the 'New Technologies in the World of Communication: Legal Issues' conference, held on 31 May 2024, at the University of Nicosia, and it has been revised for publication. This paper does not undertake a comprehensive theoretical analysis of the normative justifications for restricting speech or of the effectiveness of platform self-regulation as a governance model; instead, it focuses on the practical, comparative, and empirical dimensions of content moderation, in line with the policy-oriented conference where the original version of this paper was presented.

[2] Research Fellow of Comparative Law, University of Padua.

[3] Ludwig Wittgenstein, *Philosophical Investigations*, Vol. I (Oxford: Basil Blackwell, 1953) sec. 43.

the validity of such claims. The advent of social media has enabled researchers to determine whether hate speech yields measurable harm by empirically assessing data. This data-driven approach aids in formulating evidence-based policies to address the most significant harms of hate speech while avoiding overly broad regulation.

Social media company policies tend to look for hate speech exclusively in the content of messages. However, the meaning, impact, and repercussions of speech are deeply rooted in context. Grasping the harms of hate speech is impossible without considering political realities and power imbalances. Regulation devoid of contextual understanding inevitably overreaches.

This paper looks at regulatory approaches from both Europe and the US, juxtaposing the 'free speech' approach of the latter against that of the former. Given the universality of issues posed by social media platforms and the global debates surrounding their regulation, the ongoing dialogue between these two Western legal traditions must continue. To explore the tension between global content governance and contextual understanding, this paper begins by examining how social media companies currently approach the moderation of hate speech. Understanding these baseline strategies sets the stage for assessing both their limitations and potential for improvement.

1. Platform Moderation Strategies

Social media platforms combine two content moderation strategies to suppress hate speech: First, they deploy algorithms that scan posts, comments, and messages to automatically detect and remove content violating their hate speech policies.[4] However, the effectiveness of these algorithms can vary by language and context, leading to challenges in accurately identifying hate speech, especially in languages with subtleties or coded language.[5] Second, users can flag potentially abusive content for review by moderators, enabling crowdsourced identification and addressing of hate speech.[6]

Reported content is typically reviewed by human moderators employed by the platforms who make decisions based on community guidelines. These moderators may consider the context, the speaker's intent, and the potential risk of violence as-

[4] David Kaye, 'Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression', § 1, U.N. Doc. A/HRC/38/35 (6 April 2018).

[5] Max Fisher, 'Inside Facebook's Secret Rulebook for Global Political Speech' *The New York Times* (New York, 27 December 2018), available at https://www.nytimes.com/2018/12/27/world/facebook-moderators.html (last accessed 5 November 2025).

[6] Kaye (no 3).

sociated with the speech. Social media companies must comply with different legal requirements regarding hate speech according to the jurisdiction in which they operate.[7] For example, in the EU, they must remove content violating the European Commission's code of conduct on hate speech.[8] They must also continuously refine their moderation processes based on data and feedback—adapting policies to address emerging hate speech forms and improving content removal accuracy.

While platform moderation strategies increasingly rely on automated tools and community guidelines, their design and evaluation often lack a solid empirical foundation. The next section therefore considers the critical role of data-driven research in shaping effective and accountable hate speech policies.

2. Empirical Data and Policy Design

Empirical data is indispensable for developing evidence-based policies to address online hate speech. By analysing data, researchers can establish correlations between online hate speech and offline violence, providing an empirical basis to support regulating hate speech on social media platforms. Data enables policymakers to assess the impact of online hate speech across individuals, communities, and society at large, informing effective mitigation policies.[9]

Moreover, empirical data on user behaviour offers valuable insights into the propagation of hate speech, the actors involved, and its effects on targeted groups[10]—insights crucial for designing interventions and preventive measures. Through collecting and analysing data on the effectiveness of hate speech policies, policymakers can pinpoint areas requiring improvement and refine their approaches grounded in evidence.[11] Crucially, data-driven approaches empower platforms to adapt their con-

[7] Tarleton Gillespie, 'Platforms Are Not Intermediaries', (2018) 2 *Georgetown Law Technology Review* 198–216, 201.

[8] Elizabeth Schulze, 'EU Says Facebook, Google and Twitter Are Getting Faster at Removing Hate Speech Online', *CNBC* (4 February 2019), available at https://www.cnbc.com/2019/02/04/facebook-google-and-twitter-are-getting-faster-at-removing-hate-speech-online-eu-finds--.html (last accessed 5 November 2025).

[9] Richard A. Wilson & Molly K. Land, 'Hate Speech on Social Media: Content Moderation in Context', (2021) 52(3) *Connecticut Law Review* 1029–1076, 1033.

[10] See below § 3.

[11] Mai ElSherief & al., 'Hate Lingo: A Target-based Linguistic Analysis of Hate Speech in Social Media', (2018) *Association for Advancement Artificial Intelligence*, available at https://arxiv.org/abs/1804.04257 (last accessed 5 November 2025).

tent moderation strategies dynamically in response to evolving trends in online hate speech, thereby enhancing their ability to combat such speech effectively.

3. From Online Hate to Offline Harm

Empirical data also enables deeper analysis of the potential real-world consequences of online hate speech. Building on the previous section, this part investigates the growing body of evidence that links digital hate speech to offline violence and discrimination.

Evidence suggests that online hate speech can incite real-world violence.[12] Studies have shown that online hate speech, particularly targeting marginalised groups such as immigrants or religious minorities, is closely associated with an increased risk of real-world violence, as denigrating or discriminatory rhetoric can heighten the likelihood of physical attacks against these communities.[13] This connection underscores the importance of addressing and regulating hate speech on social media platforms to prevent potential harm and promote a safer online environment. Additionally, studies have highlighted instances where online hate speech has been directly linked to physical attacks on specific individuals and their families, indicating the direct impact of harmful online rhetoric on offline behaviour.

One example is the 'Trump effect' study, which identified a correlation between hate crimes and anti-immigrant rhetoric following the 2016 US presidential election.[14] Counties with wide margins voting for Trump reported elevated rates of hate crimes, suggesting a connection between online hate speech and real-world violence.

Another example involves State-sponsored online threats,[15] which can lead to physical harm and violence against human rights defenders or minorities. A stark illustration is Myanmar, where high-ranking military members and Buddhist nationalists used social media platforms, particularly Facebook, to spread anti-Muslim

[12] Front Line Defenders, 'Global Analysis 2018' (2019), available at https://www.frontlinedefenders.org/sites/default/files/global_analysis_2018.pdf (last accessed 5 November 2025), 6.

[13] Josh Jacobs, 'Does Online Hate Drive Anti-Migrant Violence?', *Financial Times* (24 September 2018), available at https://www.ft.com/content/a3d4b800-9bf3-11e8-88de-49c908b1f264 (last accessed 5 November 2025).

[14] Griffin Edwards & Stephen Rushin, 'The Effect of President Trump's Election on Hate Crimes' (2018) *SSRN* 3, available at https://ssrn.com/abstract=3102652 (last accessed 5 November 2025), 1–24.

[15] Carly Nyst & Nick Monaco, 'State-Sponsored Trolling: How Governments Are Deploying Disinformation as Part of Broader Digital Harassment Campaigns' (2018) *Institute for the Future* A-67, 1, available at https://legacy.iftf.org/fileadmin/user_upload/images/DigIntel/IFTF_State_sponsored_trolling_report.pdf (last accessed 5 November 2025).

rhetoric and threats of violence against the Rohingya community.[16] This online hate speech preceded and accompanied violent attacks on Rohingya villages in Rakhine State in 2016 and 2017, leading to a humanitarian crisis and mass displacement of Rohingya civilians.[17] Specifically, over 1,000 posts, comments, images, and videos on Facebook attacking Rohingya or Muslims contributed to a climate of hate speech and violence in Myanmar.[18] The lack of alternative information channels and low digital literacy levels in the country amplified the impact of these derogatory social media messages. Additionally, the population's heavy reliance on Facebook for internet access made it a primary tool for disseminating hate speech and incitement to violence, contributing to the escalation of the Rohingya genocide. The online hate speech targeting the Rohingya community played a significant role in exacerbating tensions, fostering hatred and creating a climate conducive to violence and atrocities against the Rohingya population. The result involved widespread human rights violations, including killings, sexual violence, and forced displacement.[19]

While establishing a direct causal relationship between online hate speech and real-world violence can be challenging, the growing body of evidence suggests a significant association between the two, underscoring the need for effective measures to address and mitigate the harmful effects of hate speech in online spaces to prevent potential escalation to violence in the physical world.

The documented connection between online speech and real-world harm underscores the need for nuanced regulatory frameworks. Yet, many platform policies continue to operate with limited attention to the social, political, and cultural contexts in which speech occurs. The following section examines both the conceptual and operational shortcomings of platform regulations in addressing the complex reality of hate speech.

---

[16] Office of the Prosecutor, 'Situation in the People's Republic of Bangladesh/Republic of the Union of Myanmar' (4 July 2019), ICC-01/19, § 59.

[17] Ibid. §§ 5, 67–68.

[18] Ibid. § 59.

[19] Megha Rajagopalan & al., 'How Facebook Failed the Rohingya in Myanmar', *Buzzfeed News* (27 August 2018), available at https://www.buzzfeednews.com/article/meghara/facebook-myanmar-rohingya-genocide (last accessed 5 November 2025). The assumption of responsibility came from Facebook itself: Alexandra Stevenson, 'Facebook Admits It Was Used to Incite Violence in Myanmar', *The New York Times* (6 November 2018), available at https://www.nytimes.com/2018/11/06/technology/myanmar-facebook.html (last accessed 5 November 2025).

4. The Contextual Blindness of Platform Regulations

Content moderation policies on social media platforms often treat speech as self-contained, relying heavily on predefined keyword lists, rule-based systems, and linguistic indicators to detect violations. While these tools may offer efficiency and consistency, they frequently overlook the broader context in which speech takes place. Context is pivotal when defining and regulating hate speech on social media platforms, as it illuminates the meaning and intent behind speech—words and expressions can carry divergent interpretations based on the context in which they are employed. For instance, in Spain, labelling someone a 'communist' or 'destabiliser' may be perceived as a jest, while in Guatemala, such terms could be construed as incitement to violence, given the contrasting political connotations in each context.[20] Such differences are not merely linguistic—they reflect deep historical and sociopolitical dynamics that define the lived experience of speech recipients. Without accounting for these nuances, regulation risks either overreaching—by suppressing legitimate expression—or underreaching—by failing to intervene when speech is genuinely harmful. In other words, context shapes not only meaning but also the resulting harm.

Furthermore, contextual factors such as power imbalances, historical underpinnings, and social dynamics can profoundly influence the impact of hate speech. Regulating hate speech without accounting for these power dynamics risks engendering overly broad restrictions or ineffective policies.[21] Across cultures and communities, speech can be interpreted differently; thus, considering context ensures that hate speech regulations are culturally sensitive and do not inadvertently suppress legitimate forms of expression.[22]

[20] Wilson & Land (no 8) 1067. Muna Abbas & al., 'Invisible Threats: Mitigating The Risk Of Violence From Online Hate Speech Against Human Rights Defenders In Guatemala' (2019) *American Bar Association* 31–32, available at https://www.americanbar.org/content/dam/aba/administrative/human_rights/invisible-threats-guatemala-may-2019.pdf (last accessed 5 November 2025).

[21] Charles R. Lawrence III, 'If He Hollers Let Him Go: Regulating Racist Speech on Campus' in Mari J. Matsuda & al. (eds), *Words That Wound: Critical Race Theory, Assaultive Speech, And The First Amendment* (London: Routledge, 1993) 53–88, 70.

[22] See X's rule on abuse and harassment: 'We prohibit behavior and content that harasses, shames, or degrades others, as defined below. Please note that, in order to help our teams understand the context, we sometimes need to hear directly from the person being targeted to ensure that we have the necessary information prior to taking appropriate and proportionate enforcement action. […] Some posts may appear to be harmful when viewed in isolation, but may not be when viewed in the context of a larger conversation. For example, friends may consensually use certain terms or phrases to engage with each other that could appear abusive without this context. We also recognise that our platform can be used to call attention to, condemn, or highlight others for harmful behaviors. In such cases, we do not take action where the context

Contextual analysis also enables regulators to assess the proportionality of restrictions on speech. By examining the specific circumstances surrounding speech, policymakers can tailor responses to address harm while preserving freedom of expression. In this, international human rights law also underscores the importance of context in speech regulation.[23] Platforms that align their policies with human rights standards are more likely to comply with legal requirements and promote a safe, inclusive online environment.

While platform regulations expressly acknowledge the importance of context in evaluating content, reality often falls short, leading to challenges in accurately assessing the impact and intent of user-generated content. Platforms frequently provide a narrow definition of context, focusing primarily on the words used in a post or comment.[24] This limited view fails to incorporate external information, events, or regional nuances that may influence the meaning and intent of the content. In striving for a consistent application of content moderation standards, platforms aim for 'objective' rules that can be enforced by any moderator, regardless of their familiarity with the context in which the content was shared.[25] Consequently, what constitutes context is restricted to information readily available to any moderator, potentially overlooking important contextual cues.

The sheer volume of content shared on social media also poses a significant challenge in effectively contextualising posts. Platforms grapple with integrating geographically localised meanings, cultural nuances, and historical contexts into their moderation practices due to the complexity and scale of content moderation tasks.[26] Despite recognising the importance of context, platforms have yet to find systematic ways to integrate contextual factors into their moderation processes.[27] Moderators may lack the tools or resources to consider the broader social, political, and cultural context in which content is shared, leading to a limited understanding of the implications of user-generated content.

---

is clearly non-abusive and intended to counter these kinds of rhetoric'. Available at https://help.x.com/en/rules-and-policies/abusive-behavior (last accessed 30 November 2024).

[23] Frank La Rue, 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression', U.N. Doc. A/HRC/17/27 (16 May 2011), 24.

[24] Wilson & Land (no 8) 1064.

[25] Kate Klonick, 'The New Governors: The People, Rules, and Processes Governing Online Speech' (2018) 131(6) *Harvard Law Review*1635, 1642.

[26] Wilson & Land (no 8) 1032.

[27] Monika Bickert, 'Publishing Our Internal Enforcement Guidelines and Expanding Our Appeals Process', (24 April 2018) available at https://perma.cc/VH7B-QLT9 (last accessed 30 November 2024).

Ultimately, the failure to consider context in platform regulations results in an under-determined assessment of harm caused by content. Without a comprehensive understanding of the context in which speech occurs, moderators may struggle to accurately evaluate the potential impact of content and make informed moderation decisions. To address this disconnect between the importance of context and its neglect in practice, the next section offers concrete recommendations. These proposals aim to reorient platform governance around more locally informed, nuanced, and ethically grounded moderation systems.

5. Toward Context-Aware Moderation

Incorporating context into content moderation practices on social media platforms warrants several recommendations. Platforms should pause the dramatic shift towards artificial intelligence (AI) in content moderation and conduct a detailed and transparent review of the implications of automation.[28] This review should involve independent experts to examine how algorithms may contribute to an increase in hate speech or help perpetuate biases against specific groups. Platforms must disclose the rate of false positives in their algorithms and investigate how coded or implicitly stated hate speech may evade automated detection.

Furthermore, platforms should design an iterative process in which AI and human moderation are integrated. This integration should involve a sufficiently large, trained team of employees knowledgeable about relevant social, political, and cultural contexts.[29] By combining AI capabilities with human expertise, platforms can enhance the accuracy and effectiveness of content evaluation while considering contextual factors. In some situations, the currently protected categories under platform regulations may be insufficient. For instance, human rights defenders, including judges and prosecutors engaged in anti-corruption initiatives, have faced coordinated online hate campaigns in several countries in Central and South America.[30] In Guatemala, Colombia, and Mexico alone, over 200 human rights defenders are killed each year, accounting for about half of the global total. Given the perilous nature of human rights work in such regions, it is reasonable to classify human rights defend-

[28] Wilson & Land (no 8) 1069.

[29] Center for American Progress & al., 'Recommended Internet Company Corporate Policies And Terms Of Service To Reduce Hateful Activities' available at https://www.americanprogress.org/wp-content/uploads/sites/2/2018/10/ModelInternetCompanies-appendix.pdf (last accessed 5 November 2025) 5.

[30] Front Line Defenders (no 11) 16.

ers as a protected group until these attacks diminish. If there are other vulnerable occupational groups, such as social workers or humanitarian aid workers, who are not currently protected, they too should be considered for temporary inclusion in the scope of protection.[31]

To ensure that context is systematically taken into account, platforms should expand the workforce at senior levels of review. Context-based moderation may require higher rates of escalation from frontline moderators to more experienced reviewers, where more nuanced contextual analysis can be conducted. This expansion can help improve the accuracy and depth of content evaluation.[32]

Moreover, platforms should establish offices in each country where they operate, staffed by well-trained local employees, including experienced investigative journalists.[33] These teams would be responsible for designing and implementing social media policies tailored to the context of each region. By leveraging on-the-ground expertise and continually updating lists of harmful terms and expressions, platforms can enhance their understanding of regional contexts and improve moderation practices.[34] A tailored content moderation model could involve moving away from universal lists of hate speech terms implemented by automated tools. Instead, platforms could empower local teams to develop context-specific content moderation procedures related to disinformation, hate speech, and other problematic expressions. This approach aims to democratise content governance by tailoring moderation practices to the unique context of each country or region.

While some of these measures are relatively easy to implement, others would clearly require not only substantial financial resources but also an expansion of the mission of social media companies. These recommendations point toward a broader transformation in how platforms understand their responsibilities. So, we should reflect more on the structural implications of content moderation models rooted in advertising logic and reaffirm the need for context-aware approaches grounded in comparative legal principles.

31 Wilson & Land (no 8) 1071.

32 Ibid. 1069.

33 Ibid. 1074.

34 Jennifer C. Daskal, 'Speech Across Borders', (2019) 105(8) *Virginia Law Review* 1655.

Concluding Remarks

In confronting the global challenges of hate speech on social media, this paper has emphasised the necessity of incorporating contextual understanding into content moderation practices. By analysing empirical data, legal frameworks, and platform governance structures, it has argued that a purely content-based approach is insufficient for capturing the cultural, political, and historical dimensions that give speech its meaning and impact. The comparative analysis highlights how current moderation models—rooted in technical standardisation and influenced by US-centric norms—struggle to address the real-world harms experienced by marginalised communities in diverse global contexts.[35] As Mark Zuckerberg acknowledged during a congressional hearing, social media platforms operate fundamentally as advertising companies.[36] However, the regulatory responsibilities they now bear resemble those of quasi-public institutions. This tension requires a recalibration of their mission: from content policing for brand safety to rights-conscious moderation grounded in human rights law, empirical evidence, and regional expertise. A move toward pluralistic, context-sensitive governance is not only desirable, it is essential for ensuring that the digital public sphere remains inclusive, accountable, and attuned to the lived realities of its global users.

References

Abbas M. & al., Invisible Threats: Mitigating The Risk Of Violence From Online Hate Speech Against Human Rights Defenders In Guatemala' (2019) *American Bar Association,* available at https://www.americanbar.org/content/dam/aba/administrative/human_rights/invisible-threats-guatemala-may-2019.pdf (last accessed 5 November 2025), 1–42.

Bickert M., 'Publishing Our Internal Enforcement Guidelines and Expanding Our Appeals Process', (24 April 2018), available at https://perma.cc/VH7B-QLT9 (last accessed 30 November 2024).

[35] Wilson & Land (no 8) 1073. Molly K. Land, 'The Problem of Platform Law: Pluralistic Legal Ordering and Social Media', in Paul Schiff Berman (ed.), *Oxford Handbook of Global Legal Pluralism* (Oxford: Oxford University Press, 2020) 958.

[36] During Mark Zuckerberg's congressional testimony in April 2018, Senator Orrin Hatch asked him how Facebook could sustain a business model when it does not charge users for its services. Zuckerberg famously replied: 'Senator, we run ads'. Michael Gold, 'Senators Had a Lot to Say About Facebook. That Hasn't Stopped Them From Using It.' *The New York Times* (12 April 2018) available at https://www.nytimes.com/2018/04/12/us/politics/facebook-senators-usage.html (last accessed 30 November 2024). This response became widely quoted and highlighted the advertising-driven revenue model of Facebook.

Center for American Progress & al., 'Recommended Internet Company Corporate Policies And Terms Of Service To Reduce Hateful Activities', available at https://www.americanprogress.org/wp-content/uploads/sites/2/2018/10/ModelInternetCompanies-appendix.pdf (last accessed 5 November 2025), 1–11.

Daskal J.C., 'Speech Across Borders', (2019) 105(8) *Virginia Law Review* 1605–1666.

Edwards G. & S. Rushin, 'The Effect of President Trump's Election on Hate Crimes' (2018) *SSRN* 3, available at https://ssrn.com/abstract=3102652 (last accessed 5 November 2025), 1–24.

ElSherief M. et al., 'Hate Lingo: A Target-based Linguistic Analysis of Hate Speech in Social Media', (2018) *Association for Advancement Artificial Intelligence*, available at https://arxiv.org/abs/1804.04257 (last accessed 5 November 2025).

Fisher M., 'Inside Facebook's Secret Rulebook for Global Political Speech', *The New York Times* (27 December 2018), available at https://www.nytimes.com/2018/12/27/world/facebook-moderators.html (last accessed 5 November 2025).

Front Line Defenders, 'Global Analysis 2018' (2019), available at https://www.frontlinedefenders.org/sites/default/files/global_analysis_2018.pdf (last accessed 5 November 2025), 6.

Gillespie T., 'Platforms Are Not Intermediaries', (2018) 2 *Georgetown Law Technology Review* 198–216.

Gold M., 'Senators Had a Lot to Say About Facebook. That Hasn't Stopped Them From Using It.' *The New York Times* (12 April 2018) available at https://www.nytimes.com/2018/04/12/us/politics/facebook-senators-usage.html (last accessed 30 November 2024).

Jacobs J., 'Does Online Hate Drive Anti-Migrant Violence?', *Financial Times* (24 September 2018), available at https://www.ft.com/content/a3d4b800-9bf3-11e8-88de-49c908b1f264 (last accessed 5 November 2025).

Kaye D., 'Rep. of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression', § 1, U.N. Doc. A/HRC/38/35 (6 April 2018).

Klonick K., 'The New Governors: The People, Rules, and Processes Governing Online Speech', (2018) 131(6) *Harvard Law Review* 1598–1670.

La Rue F., 'Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression', U.N. Doc. A/HRC/17/27 (16 May 2011), 24.

Land M.K., 'The Problem of Platform Law: Pluralistic Legal Ordering and Social Media', in Paul Schiff Berman (ed.), *Oxford Handbook of Global Legal Pluralism* (Oxford: Oxford University Press, 2020) 974–994.

Lawrence C.R. III, 'If He Hollers Let Him Go: Regulating Racist Speech on Campus' in Mari J. Matsuda & al. (eds), *Words That Wound: Critical Race Theory, Assaultive Speech, And The First Amendment* (London: Routledge, 1993), 53–88.

Nyst C. & N. Monaco, 'State-Sponsored Trolling: How Governments Are Deploying Disinformation as Part of Broader Digital Harassment Campaigns' (2018) *Institute for the Future* available at https://legacy.iftf.org/fileadmin/user_upload/images/DigIntel/IFTF_State_sponsored_trolling_report.pdf (last accessed 5 November 2025), A-67.

Office of the Prosecutor, 'Situation in the People's Republic of Bangladesh/Republic of the Union of Myanmar' (4 July 2019), ICC-01/19, § 59.

Rajagopalan M. & al., 'How Facebook Failed the Rohingya in Myanmar', *Buzzfeed News* (27 August 2018), available at https://www.buzzfeednews.com/article/meghara/facebook-myanmar-rohingya-genocide (last accessed 5 November 2025).

Schulze E., 'EU Says Facebook, Google and Twitter Are Getting Faster at Removing Hate Speech Online', *CNBC* (4 February 2019), available at https://www.cnbc.com/2019/02/04/facebook-google-and-twitter-are-getting-faster-at-removing-hate-speech-online-eu-finds--.html (last accessed 5 November 2025).

Stevenson A., 'Facebook Admits It Was Used to Incite Violence in Myanmar', *The New York Times* (6 November 2018), available at https://www.nytimes.com/2018/11/06/technology/myanmarfacebook.html (last accessed 5 November 2025).

Wilson R.A. & M.K. Land, 'Hate Speech on Social Media: Content Moderation in Context' (2021) 52(3) *Connecticut Law Review* 1029–1076.

Wittgenstein L., *Philosophical Investigations* Vol. I (Oxford: Basil Blackwell, 1953).

X, 'Rule on abuse and harassment', available at https://help.x.com/en/rules-and-policies/abusive-behavior (last accessed 30 November 2024).

Rethinking Remedies for Hate Speech: The Case for Apologies in Italian and Comparative Law

Filippo Viglione[1]

Abstract

This paper examines private law remedies for hate speech, with a focus on compelled apologies. After outlining the comparative legal frameworks, it turns to the Italian context, where anti-discrimination law provides a civil law basis for addressing hate speech. Through analysis of key Italian case law, the paper argues that monetary compensation alone often fails to redress the symbolic and relational harm inflicted by hate speech. Apologies, though rare in civil law traditions, may serve as effective tools of corrective justice by acknowledging harm, restoring victims' dignity, and reaffirming societal norms. The paper calls for a greater integration of such remedies into legal practice as part of a plural and proportionate response to hate speech.

Keywords: hate speech; apologies; civil liability

1. Introduction

The issue of hate speech and its legal remedies is well-established in scholarly discourse.[2] It implicates not only the individuals who engage in such speech but also the platforms that, at times, facilitate these communications without sufficient oversight, thereby allowing content with potentially harmful effects to proliferate. Broadly defined, hate speech refers to expressions intended to incite hatred or target individuals by promoting prejudice against specific groups, particularly on grounds such as race, religion, ethnicity, disability, sexual orientation, or gender identity.[3]

---

1 Professor of Comparative Law at the University of Padua.

2 See, among others, Michael Herz & Peter Molnar (eds), *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge University Press, 2012); Aharon Barak, *Proportionality. Constitutional Rights and Their Limitations* (Cambridge University Press, 2012); Jennifer Schweppe & Mark Austin Walters (eds), *The Globalization of Hate: Internationalizing Hate Crime?* (Oxford University Press, 2016); Maxime Lepoutre, 'Hate Speech Laws: Expressive Power is not the Answer' (2020) 25(4) *Legal Theory* 272; Richard Delgado & Jean Stefancic, *Understanding Words that Wound* (Routledge, 2018).

3 More specifically, the UN Rabat Plan of Action on the prohibition of advocacy of national, racial, or religious hatred that constitutes incitement to discrimination, hostility, or violence, proposes that 'a clear distinction should be made between three types of expression: expression that constitutes a criminal of-

The central debate over hate speech is where to draw the line between expression that is merely morally questionable and expression that should be banned.[4] [5] This may be the aim of a constitutional law discourse or of law generally, and distinct approaches to this problem exist within different legal systems.[6]

Pending the adoption of a unified European framework, which has been repeatedly called for through formal communications, comparative analysis has thus far focused on the examination of national legal systems. Comparing the US approach with those taken in Italy, France, Germany, and in various other European countries is a popular academic exercise, even if sometimes national laws are merely summarised, without much explanation of their application. Given that there is no exclusive method of comparison, a proper and correct application—meaning an approach consistent with the most authoritative comparative methodology, which requires contextualisation rather than mechanical juxtaposition[7]—must not be limited to a simple 'comparison by columns'[8] but instead deepen the systematic regulatory context, the historical origin of the rules, and their social impact, so as to make the comparison effective.[9]

Taking a constitutional approach, the US prioritises freedom of speech, which is protected by the First Amendment, while Europe is more willing to criminally sanc-

fence; expression that is not criminally punishable, but may justify a civil suit or administrative sanctions; expression that does not give rise to criminal, civil or administrative sanctions, but still raises concern in terms of tolerance, civility and respect for the rights of others'.

[4] See the Communication from the Commission to the European Parliament and the Council, 'A more inclusive and protective Europe: extending the list of EU crimes to hate speech and hate crime', COM/2021/777 final.

[5] The Committee of Ministers of the European Council, in its Recommendation No. R (1997) 20 provides the following working definition: 'the term "hate speech" shall be understood as covering all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on intolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin'.

[6] See Mark Tushnet, *Advanced Introduction to Freedom of Expression* (Elgar, 2018).

[7] On the need for a context-sensitive methodology in comparative law, see Rodolfo Sacco, 'Legal Formants: A Dynamic Approach to Comparative Law. (Installment I of II)' (1991) 39 *American Journal of Comparative Law* 1–34.

[8] Fernanda Nicola, 'Family Law Exceptionalism in Comparative Law' (2010) 58 *American Journal of Comparative Law* 777, 785.

[9] For a general overview, see, among many others, Mark Van Hoecke, *Epistemology and Methodology of Comparative Law* (Hart Publishing, 2004); Geoffrey Samuel, *Epistemology and Method in Law* (Ashgate, 2003); Pier Giuseppe Monateri (ed.), *Methods of Comparative Law* (Elgar, 2012).

tion hate speech.[10] Civil sanctions are also available, and it is in this arena that much of the discussion on the apology arises. In the attitude towards hate speech, therefore, the very idea of a society is reflected, whether it is characterised by a libertarian culture that accepts the existence of even aberrant opinions—as long as they do not result in a defamation or do not create a present danger—or one which provides some form of sanction for the grimmest and most threatening expressions.[11] More authoritarian cultures may seek to punish as hate speech material which would in other countries be regarded as quite acceptable.

Furthermore, the practical implications of these divergent cultural and legal approaches are magnified by the dynamics of contemporary communication. In today's interconnected world, individuals can easily find others who share and reinforce their views—even those that are discriminatory or hateful. This accelerates the circulation of hate speech, allowing it to transcend geographical boundaries and amplify its social harm.[12] Thus, discriminatory opinions are not relegated to intimate circles but spread fast around the world generating a spiral of hatred. But who decides what constitutes hate speech? When and how are various types of speech protected? And, above all, should people be penalised for the things they say? Upon these premises, this paper presents a brief overview of the problem of private law consequences of hate speech, taking into account the range of remedies that can be used by judges to address the problem. This leads to attributing a particular role to apologies in the context of hate speech case law. An apology may sensitise the speaker to the harm they have caused, make the victim feel better,[13] and establish or strengthen commu-

[10] The distinctiveness of the US model is generally recognised. See Samuel Walker, *Hate Speech: The History of an American Controversy* (Nebraska University Press, 1994) 14–16, 161–62; Kevin Boyle, 'Hate Speech -The United States Versus the Rest of the World' (2001) 53 *Maine Law Review* 487; Guy E. Carmi, 'Dignity versus Liberty: The Two Western Cultures of Free Speech' (2008) 26 *Boston University International Law Journal* 277; Robert Post, 'Hate Speech' in Ivan Hare & James Weinstein (eds), *Extreme Speech and Democracy* (OUP, 2009) 123, 137; Erik Bleich, *The Freedom to be Racist? How the United States and Europe Struggle to Preserve Freedom and Combat Racism* (OUP, 2011) 17–18, 21–22.

[11] Joseph J. Hemmer Jr, 'Hate speech - The egalitarian/libertarian dilemma' (1995) 5 *Howard Journal of Communications*, 307 (proposing a hate speech policy that attempts to balance divergent concerns).

[12] Legal doctrine is generally sceptical of treating online hate speech as legally exceptional. Most scholars advocate for a consistent framework that applies equally to hate speech regardless of the medium through which it is expressed. See Lynn Adelman & Jon Deitrich, 'Extremist Speech and the Internet: The Continuing Importance of Brandenburg' (2001) 4 *Harvard Law and Policy Review* 361.

[13] In the same sense, Maxwell J. Mehlman, 'Promoting Fairness in the Medical Malpractice System' in William M. Sage & Rogan Kersh (eds), *Medical Malpractice and the U.S. Health Care System* (Cambridge University Press, 2009) 152.

nity norms against this type of behaviour in the future—all elements that correspond to the general functions of tort law.

This should not be considered exotic or unusual: even if it is true that most present-day jurisdictions of the civil tradition are not familiar with the remedy of forced apologies, as nowadays is occasionally acknowledged in a number of common law jurisdictions, there are significant examples of similar remedies in the law in past centuries.[14] In the Republic of Venice until the late 18th century, for example, the law classified the following, in decreasing order of importance, as punishments to crime: the death penalty, forced exile, jail sentences, forced labour, financial penalties, and, finally, compelled public apologies ('If the defendant does not apologise within the time established by the judge or if the judge deems that the apology is not properly expressed, a trial will be filed against him for lack of collaboration with the justice').[15]

In some jurisdictions, the analysis of civil remedies shows a clear link between apologies—which may serve purposes and have effects that cannot be attained by mere damages—and expressions that offend honour and dignity (for example, the *amende honorable* action of Roman-Dutch law, which took the form of a declaration by the wrongdoer retracting their words and an apology and prayer for forgiveness).[16] In this sense, apologies appear as secular rituals that may reaffirm a societal order and restore the dignity of the victim.

2. Hate Speech in the Comparative Context

Responses to unlawful hate speech normally take place within the framework of criminal law, minding the constitutional balance between freedom of thought and other important values. Although usually considered of lesser importance, civil remedies may contribute as a legal tool to counteract expressions of hatred. However, remedial and repressive systems are insufficient to eradicate this phenomenon—they must be accompanied by a cultural and educational investment, which is still largely lacking.

[14] See Wannes Vandenbussche, 'Rethinking Non-Pecuniary Remedies for Defamation: The Case for Court-Ordered Apologies' (2020-21) 9 *Journal of International Media & Entertainment Law* 109–170, showing that the ancestors of court-ordered apologies have played a prominent role in the past in many common law and civil law systems.

[15] For a historical and legal overview of the sanctioning system in the Republic of Venice, see Alfredo Viggiano, 'Giustizia, disciplina e ordine pubblico – Storia di Venezia' *Enciclopedia Treccani* (1996).

[16] See Jonathan Burchell 'Retraction, Apology and Reply as Responses to iniuriae' in Eric Descheemaeker & Helen Scott (eds), *Iniuria and the Common Law* (Hart Publishing, 2013) 197–214; Jan Hallebeek & Andrea Zwart-Hink, 'Claiming apologies: a revival of amende honorable?' (2017) 5 *Comparative Legal History* 194–242.

The constitutional framework on hate speech may also be contrasted with other—sometimes very effective—instruments; for example, the sanctioning power provided by labour law or by codes of conduct or sometimes even simple contractual regulations. Italian media recently reported on the following incidents: a) the dismissal of a teacher who, on a social network, wished death on migrants crossing the Mediterranean;[17] b) the suspension of a journalist by his professional association as a consequence of an article that had 'racist' connotations;[18] and c) the termination of the contractual relationship between a host and a homestay company on the basis of homophobic statements by the former.[19]

The plurality of possible remedies is at the heart of the present analysis, which is divided in three parts. After the introduction and some preliminary comparative remarks, the paper narrows down the meaning of 'hate speech' and explores the harm that it creates. The idea of 'harm in hate speech'—which has been broadly investigated, especially in American literature—requires the tracing of the cultural coordinates that can help identify what is allowed. This analysis should consider, among other elements, the prevailing views within a society. While hate speech cannot be reduced to mere dissent from majority opinion, examining how such speech deviates from broadly shared social values may help identify when it crosses the line into legally sanctionable harm. In this, Italian case law offers a particular perspective. Through anti-discrimination legislation, civil judges have tried to build a system that protects victims of hate speech. The Italian legal context and case law will be analysed in the second part of the paper, while the final part considers potential remedies for hate speech. A pragmatic view helps to enhance the role of apologies in mitigating the negative effects of hate speech. In adequately acknowledging the harm of hate speech, apologies may prove compatible with common and civil law systems alike.

Value-based choices must be made to navigate the innumerable paradoxes and short-circuits that hate speech inevitably conveys. In the 1940s, Karl Popper illustrated the most famous one, known as 'the paradox of tolerance':

---

[17] La Repubblica, 'Venezia, condannata la prof delle frasi razziste' (8 November 2018), available at www.repubblica.it/cronaca/2018/11/08/news/condannata_la_prof_delle_frasi_razziste-211115658 (last accessed 28 July 2025).

[18] Il Giornale, 'Facci sospeso perché rivendica il diritto all'odio' (17 June 2017), available at www.ilgiornale.it/news/cronache/facci-sospeso-perch-rivendica-diritto-allodio-1410132.html (last accessed 28 July 2025).

[19] Il Messaggero, '"Niente gay, affitto solo a coppie etero": su Airbnb a Foggia l'annuncio che esclude gli omosessuali' (2 August 2019), available at www.ilmessaggero.it/italia/affitto_gay_foggia_airbnb_omofobia_oggi_ultime_notizie-4655316.html (last accessed 28 July 2025).

> Unlimited tolerance must lead to the disappearance of tolerance. If we extend unlimited tolerance even to those who are intolerant, if we are not prepared to defend a tolerant society against the onslaught of the intolerant, then the tolerant will be destroyed, and tolerance with them.[20]

But more generally, the argument—as in Ernst-Wolfgang Böckenförde's dilemma—that the liberal (*freiheitlich*), secularised State lives by prerequisites which it cannot guarantee itself, alludes to the difficulty of clearly defining the conditions for the restriction of fundamental rights, including freedom of expression.[21] Such difficulties should not prevent the search for appropriate legal tools to eliminate or mitigate the harmful consequences of hate speech.

In this respect, the principle of proportionality offers a valuable framework for balancing competing constitutional rights. It allows for a structured assessment of whether restrictions on expression are suitable, necessary, and proportionate to the legitimate objective pursued, thus helping to reconcile the protection of dignity and equality with the preservation of freedom of speech. Within this broader framework, this paper focuses on the potential role of apologies as a civil remedy. The rationale lies in the symbolic and relational dimension of the harm caused by hate speech: where dignity and equality are undermined, a compelled apology may contribute to restoring the moral status of the victim and reaffirming community norms.

3. Targets and Harms of Hate Speech: Legal Foundations for Civil Remedies

To assess whether apologies serve as an effective remedy in redressing the harm caused by hate speech, it is first necessary to clarify what hate speech entails and to understand the nature of the harm it produces. This section explores how hate speech is addressed in various legal systems, with particular attention to the underlying values at stake—such as human dignity, equality, and social cohesion—that any meaningful remedy, including apologies, ought to reflect. When addressing xenophobic and racist expression, European countries usually act on the criminal level, for

[20] Karl R. Popper, *The Open Society and its Enemies, I, The Spell of Plato* (Routledge, 1945).

[21] Jürgen Habermas & Joseph Ratzinger, *The Dialectics of Secularization: On Reason and Religion* (Brian McNeil tr, Ignatius Press, 2006). Originally in German as *Dialektik der Säkularisierung: Über Vernunft und Religion* (Herder Verlag, 2005).

example, by criminalising negationism (such as Holocaust denial).[22] There are also rules for the electronic dissemination of hate speech, as well as fake or manipulated news;[23] this includes provisions to control third parties who host such online communications.[24] In recent years, however, the trend of using remedial models typical to private law has also emerged, mainly at the EU's urging, specifically to leverage a complex set of provisions that make up the anti-discrimination law.[25] This body of norms—as discussed in the next section—often serves as the legal foundation for civil remedies against hate speech, framing it as a manifestation of discrimination against protected groups.

It is no surprise that the law deals with hatred—conveyed through words or symbols, such as cross burnings by the Ku Klux Klan—as it is a feeling from which further actions may germinate and thus pave the way for dangerous consequences. If hatred is not opposed, but accepted and sometimes even fuelled in the political arena, the law itself can bend before it, as has already tragically happened when the banality of

[22] Indeed, hate speech may represent a violation of memory laws, adopted in many European countries for regulating public memory of historical events that 'preserve the State's version of history'. See Emanuela Fronza, 'The Criminal Protection of Memory: Some Observations about the Offense of Holocaust Denial', in Ludovic Hennebel & Thomas Hochmann (eds), *Genocide Denial and the Law* (Oxford University Press, 2011) 155–181; Robert Kahn, 'Holocaust Denial and Hate Speech', ibid 77–107 (expressing the idea that Holocaust denial is a form of hate speech). See also the Italian decision of the Corte di Cassazione (Cass. Pen., 21.9.2023) where the Supreme Court characterises the display of symbols denigrating the Holocaust, occurring in the context of a public neo-fascist demonstration, as propaganda promoting ideas based on the denial or gross minimisation of the Holocaust (in Giurisprudenza Italiana, 2024, 1142).

[23] See, in addition to the German legislation cited in the following note, also the EU level: Digital Services Act (Regulation (EU) 2022/2065), in force since November 2022, introducing due diligence obligations for online platforms, including measures against hate speech and disinformation.

[24] The most relevant example is the German Network Enforcement Act (*NetzDG, Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken*), which aims to combat hate speech and fake news in social networks. The law refers to social networks with at least 2 million members, not to journalistically and editorially designed services. Providers are obliged to establish a transparent procedure for dealing with complaints about illegal content (§ 3 NetzDG) and are subject to a reporting and documentation obligations.

[25] According to Article 2 of the Treaty of European Union, non-discrimination is one of the fundamental principles of the EU. Article 10 of the Treaty on the Functioning of the European Union requires the EU to combat discrimination based on sex, racial or ethnic origin, religion or belief, disability, age, or sexual orientation, when defining and implementing its policies and activities. In 2000, two directives were adopted: the Employment Equality Directive (2000/78/EC), which prohibited discrimination on the basis of sexual orientation, religion or belief, age, and disability, in the area of employment; and the Racial Equality Directive (2000/43/EC), which introduced prohibition of discrimination on the basis of race or ethnicity in the context of employment, but also in accessing the welfare system and social security, as well as goods and services.

evil, as Hannah Arendt explained, took over.[26] Moreover, the harm caused by hate speech is never insignificant and requires an adequate response in terms of relief for the victim.

Among European countries, which largely adopt similar regulatory instruments, there is no consensus on appropriate sanctions on authors of hate speech; it may depend on the difficulty of interrelating the very idea of honour or dignity with large groups of people, gathered on the basis of their ethnicity, language, sexual orientation, or religion.[27] The search for philosophical foundations of criminal prohibitions (attacking honour and dignity, or public and social order) could be of some help, even when one looks at the civil remedies that legal systems tend to use in countering hate speech.

This is what happens in Germany, where human dignity plays a pivotal role in constitutional discourse and represents the highest value of the German Basic Law (*Grundgesetz* or *GG)*, informing the substance and spirit of the entire document; tellingly, human dignity appears in Article 1 of the GG ('Human dignity shall be inviolable. To respect and protect it shall be duty of all state authority'). In this sense, the crime of incitement to hatred (*Volksverhetzung*) may consist of an attack on human dignity by parts of the population, through conduct that disturbs public peace.[28] In this context, the interest compromised by hate speech is identified as human dignity. Regarding Holocaust denial, the case law of the Federal Constitutional Court (*Bundesverfassungsgericht*) draws a clear distinction between statements of fact and expressions of opinion. The Court has held that Article 5 of the GG, which guarantees freedom of expression, does not protect statements of fact that have been proven false.

The prevailing French understanding of hate speech regulation is quite different. Damage originating from hate speech is conceived as an injury to groups' honour, both in the Marchandeau Decree of 1939 and in the more recent memory laws (*lois*

[26] Hannah Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil* (New York: Viking Press, 1963).

[27] For more details on the penal responses to hate speech in European legal systems, see I. Hare & J. Weinstein (eds), *Extreme Speech and Democracy* (OUP, 2009).

[28] Section 130 of Germany's Penal Code, on group defamation, prohibits 'attacks on human dignity by insulting, maliciously maligning, or defaming part of the population'. The broad reference to 'part of the population' helps protect all vulnerable groups, and not only those based on race, sexual orientation, religion, or other specified and nominated characteristics.

mémorielles).[29] The function is to establish the criminal repression of expressions that offend particular groups of people by inciting hatred among citizens or that produce the same effect of injury to honour by denying the genocides listed by law.[30]

Similar difficulties in identifying the basis for criminal sanctions in compliance with the principle of harmfulness[31] are found in Italian law, where doctrine and jurisprudence have manifested many uncertainties not only on the compatibility of the incriminating cases with Article 21, Section 1 of the Constitution ('Everyone shall have the right to freely express their thoughts in speech, writing, or any other form of communication'), but also on the evaluation of the interest protected by the criminal law. Nonetheless, the judicial decisions of the last 30 years have framed the anti-racist objectives in the spirit of the Constitution, thereby admitting the possibility for the legal system to punish even crimes of mere conduct and abstract danger, as well as the irrelevancy for the incitement to hatred to achieve concrete effects. In the background, however, the alternative remains between a foundation that lies in the 'personalistic' protection of the dignity of individuals and the hypothesis in which the incriminating norm is aimed at protecting the peaceful coexistence of the various groups within society.[32]

Finally, the criminalisation of hate speech in the UK has its own approach, where—as indicated by the title of the Public Order Act 1986—a sanctioning require-

[29] Through various memory laws, France criminalises denial, minimising, and grossly abating Holocaust and other genocides recognised by international tribunals. In particular, the Gayssot Act of July 13, 1990 extended the 1881 Freedom of Press Law by introducing a new article that punishes the denial of the Jewish Holocaust as defined by the Nuremberg trials. In 2001, on the grounds of Taubira Act, France also officially condemned the historical slave trade as a crime against humanity. In counterpoint, the country then adopted the 2005 Mekacher Act, which mandated the inclusion of positive elements of French presence in the colonies, stating that educational programmes must teach the 'positive role' of France's presence abroad.

[30] Cfr. Johann Michel, *Gouverner les mémoires – Les politiques mémorielles en France* (Presses Universitaires de France, 2010); Marc Frangi, 'Les lois mémorielles: de l'expression de la volonté générale au législateur historien' (2005) *RD publ* 241; Patrick Fraisseix, 'Le Droit mémoriel' (2006) *Revue française de droit constitutionnel* 483–508.

[31] According to this principle, there cannot be any crime without harm (which includes actual damage and even potential danger). This means that harm is a constitutive element of any crime (*nullum crimen sine iniuria*). Two consequences can be drawn from this premise: a) the legislator cannot criminalise harmless behaviours; and b) the judge cannot punish an act that, even if it corresponds to the abstract legislative provision, was harmless in that specific case.

[32] Jeremy Waldron, 'Dignity and Defamation: The Visibility of Hate' (2010) 123 *Harvard Law Review* 1596, at 1609: 'it is not the group as such that we are ultimately concerned about [...] The concern in the end is individualistic'.

ment connected to the protection of public order seems to prevail, without direct implications for personal dignity or for the honour of individuals or groups. This Act prohibits speech that is threatening, abusive, or insulting if the author intends to stir up racial hatred but also if, having regard to all the circumstances, racial hatred is likely to be stirred up thereby.[33]

For the purpose of analysing civil remedies, this very brief reconstruction of the justification of criminal laws contributes to the impression that the real difficulty in providing remedies for hate speech lies not so much in its compatibility with the constitutional principles of freedom of expression—which requires a careful drawing of the boundaries of the freedom itself—but rather in the correct delimitation of the interests that hate speech is capable of damaging. In the wide spectrum of civil remedies, the proper detection of the harm generated by hate speech may help foreground apologies as a tool for restoring the victim's dignity through a symbolic transfer of humiliation to the offender.

The harms that such regulations are designed to remedy include those expressed by critical race theorists, who argue, *inter alia*, that being subjected to hate speech is analogous to 'spirit murder', meaning that hate speech enacts 'disregard for others whose lives qualitatively depend on our regard'.[34]

In this debate, one of the most notable contributions has been that of Jeremy Waldron in his book, *The Harm in Hate Speech*.[35] His idea, which may have some implications for the remedies that can be used and on apologies in particular, is that the harm is not so much a result of the speech targeted at, say, racial groups or homosexuals, but rather in the speech itself. In other words, the harm is not caused by hate speech—the speech itself constitutes the harm;[36] restrictions on hate speech therefore become a tool for securing a 'dignitarian social order', bestowing liberal democracies with a positive duty to protect the atmosphere of mutual respect against certain forms of vicious attacks. Logically, then, this view excludes hate speech from the freedom of speech principle, while protecting members of vulnerable minorities.

In practice, however, hate speech often produces both collective and individual consequences. While its primary function may be to degrade the social standing of

[33] Public Order Act 1986 § 18(1).

[34] Patricia J. Williams, 'Spirit-Murdering the Messenger: The Discourse of Fingerpointing as the Law's Response to Racism', (1987) 42 *University of Miami Law Review* 127, at 132.

[35] Jeremy Waldron, *The Harm in Hate Speech* (Harvard University Press, 2012).

[36] For a similar distinction, see Ishani Maitra & Mary Kate McGowan, 'Introduction and Overview', in Ishani Maitra & Mary Kate McGowan (eds), *Speech and Harm* (Oxford University Press, 2012) 4–8.

the targeted group, individual victims may also experience concrete psychological or social harm, even when the attack is framed in general terms. The dual nature of this harm—symbolic and personal—complicates the landscape, and suggests the need for remedies that can operate on both levels. It is more a damage caused to society in general[37] and this circumstance compels the interpreter to discover remedies that are different from pecuniary compensation and that can rebuild social harmony in the community, even when meant to restore the single individual victim of a hate speech (one such remedy being a court-ordered apology, which will be further discussed in Section 5, below).

A more common approach considers hate speech as an attack on human dignity or on the honour of the people who are part of a discriminated group.[38] Some authors have emphasised that 'direct, immediate, and substantial injury' may be caused, whether or not there were 'fighting words' or the risk of immediate public disorder involved.[39] In this perspective, it is a commonly made mistake to define hate speech as only that which is likely to lead to public disorder.

Any racist speech—which entails deeper implications than at first impression—would then increase the difficulty to redeem from a situation of social inferiority; thus, a correct policy would impose a corrective sanction, both on the criminal law side and on the civil liability front. This theory, elaborated in the US legal system, implies a consequent overcoming of the consolidated interpretation of the First Amendment of the US Constitution.

Beyond the problem occupying a peculiar position, in terms of whether the harm is caused by hate speech or the hate speech itself is what constitutes the harm (arguing from a performative and not purely expressive nature of hate speech), the alternatives informing the European debate appear to be the same. If we reflect on discrimination, which is the trait the topic of hate speech most frequently deals with,

[37] See Waldron's peculiar idea of dignity as a social and legal status that demands corresponding recognition and treatment (Waldron, no 32, at 1612).

[38] Steven J Heyman, 'Hate Speech, Public Discourse, and the First Amendment', in Ishani Maitra & Mary Kate McGowan (eds), *Speech and Harm* (no 18) 158–181; Alexander Tsesis, 'Dignity and Speech: The Regulation of Hate Speech in a Democracy' (2009) 44 *Wake Forrest Law Review* 497–532.

[39] For an early approach to these theses, which later grew into critical race theory, a doctrinal movement which experienced its greatest fortunes in the late 1980s, see Richard Delgado, 'Words that Wound: A Tort Action for Racial Insults, Epithets, and Name Calling' (1982) 17 *Harvard Civil Rights-Civil Liberties Review* 133 (supporting a justification for the use of civil liability remedies); Mari J. Matsuda & al., , *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment* (Westview Press, 1993); Catharine MacKinnon, *Only Words* (Harvard University Press, 1993).

the principle of equality inevitably comes to the fore, thus establishing the right not to be discriminated against even through expressions of thought.

This argument raises more than one systematic problem, especially related to the very configurability of the hypothesis that equality can be violated not only by discriminatory treatments—which are easily recognised in every field, starting from labour law—but also by manifestations of thought. If hate speech is indeed framed as a form of discrimination, then it is true that an effective remedial framework must go beyond symbolic or expressive remedies. In such cases, mere apologies may not suffice—but they can nonetheless play a complementary role. While structural remedies and compensation may be required to address the material effects of discriminatory conduct, apologies may still serve to restore the symbolic dimension of equality, reaffirm the victim's social standing, and promote public repudiation of the discriminatory message.

4. Examples from Italian Case Law

Since the definition of hate speech is contestable, it is common and useful in academic scholarship to lead with examples taken from case law. In the Italian legal system, the connection between hate speech and anti-discrimination law is not merely instrumental but also conceptual.[40] Hate speech is addressed, in many cases, as a form of discriminatory conduct, insofar as it reinforces exclusion, denigration, or unequal treatment of persons based on protected characteristics such as ethnicity, religion, or national origin. This approach allows civil judges to apply anti-discrimination norms as a legal basis for granting remedies—even in cases where criminal law is not directly invoked—thus expanding the scope of protection beyond interpersonal insults to include broader forms of symbolic or group-based harm. The Italian legislative system does not offer a wide spectrum of judicial decisions and that may arise from the lack of a national human rights authority (for example, an anti-discrimination board or a human rights commission—bodies that would arguably be beneficial in promoting access to remedies and supporting victims of hate speech). However, according to Article 74 of the Italian Code of Criminal Procedure, the alleged victims of a crime are allowed to start civil proceedings within the criminal trial to claim compensation

[40] The choice to focus on the Italian legal context is motivated by its distinctive reliance on civil remedies grounded in anti-discrimination law, rather than on criminal repression, which remains more limited than in other European jurisdictions. This makes Italy a suitable case study for exploring the potential of private law responses—especially apologies—in addressing the harm caused by hate speech.

for monetary and non-monetary damage caused by the commission of said crime.[41] Although it is possible to pursue a separate civil lawsuit, the higher speed and lower costs make the choice to bring the claim for compensation within the criminal proceedings much more common.[42]

Article 5 of Legislative Decree no. 215/2003 entitles associations and legal persons to act in support of or on behalf of victims of racial and ethnic discrimination. Legal standing is granted to organisations and associations specialised in the promotion of the right to equality and registered with the National Office Against Racial Discrimination (known locally as UNAR) to initiate civil cases either within the criminal trial or independently of it.[43] Only these organisations are entitled to stand on behalf of, or to intervene in support of, the alleged victims of racial, ethnic, national, or religious discriminatory acts (including hate speech) in a judicial proceeding. This specificity of the Italian legal system is rooted in the idea that hate speech requires targeting a group, or an individual but only if considered as a member of a group. These organisations are also entitled to bring claims in cases of collective discrimination against a

[41] Although the victim and the civil plaintiff are often the same person, a civil claim for damage may be filed by anyone who has been only indirectly damaged by the commission of the crime.

[42] In a broader perspective, it is also necessary to consider the influence of the jurisprudence of the European Court of Human Rights. The Court has followed two different approaches: the first, based on Article 17 of the Convention (prohibition of abuse of rights) excludes hate speech from any protection where comments negate the fundamental values of the Convention. A second approach sets restrictions on the protection afforded by Article 10 (freedom of expression), where the speech in question, while being hate speech, is not apt to destroy the fundamental values of the Convention. While freedom of expression is subject to exceptions in Paragraph 2 of Article 10, these exceptions 'must, however, be construed strictly, and the need for any restrictions must be established convincingly'. (Şener, App. No. 26680/95; accord Thoma, 2001-III Eur. Ct. H.R. at 84). Any distinction between persons holding one opinion or another would be contrary to the principles of democracy, which is strongly protected by the Court. Applying its usual technique, the Court has approached hate speech on a case-by-case basis, considering mainly the nature of the offensive or hateful speech in question and the context in which it was made. For a recent decision, see the European Court of Human Rights, 15.5.2023, *Sanchez v. France*: the applicant, a local councillor and parliamentary candidate, was fined for inciting hatred or violence against Muslims by failing to promptly delete inflammatory comments from his public Facebook page used during his campaign. The comment authors were also convicted as accomplices. The case arose from a complaint by the partner of a political opponent, who felt insulted and confronted one author, prompting him to delete his comment and inform the applicant. Although the applicant advised caution in future comments, he did not moderate the existing ones and no violation of Article 10 was found.

[43] See Mathias Möschel, 'Italy's (Surprising) Use of Racial Harassment Provisions as a Means of Fighting Discrimination', in Barbara Havelková & Mathias Möschel (eds), *Anti-Discrimination Law in Civil Law Jurisdictions* (OUP, 2019) 279.

group, in the absence of individuals specifically targeted or immediately identified in the offensive expression.[44]

When the plaintiff is looking for civil relief (damages or other forms of civil remedy) in a case of hate speech, they are required to provide unequivocal evidence that the statement made by the defendant reaches the level of wrongfulness of a discriminatory act; on the other side, the defendant is allowed to prove that their statement was not in fact discriminatory, for example highlighting the political connotation of the speech.

Before going to court, parties could ask UNAR to mediate; the body would then investigate the complaint to ascertain whether hate speech has occurred and seek to facilitate a confidential settlement between the complainant and the respondent. During this step, it should be possible to offer some remedies, which could include an agreement to desist, to apologise, to publish a retraction, or to conduct an educational campaign in the workplace. While this is possible, and desirable in a context of alternative dispute resolution, it is much more difficult for a civil law judge to order a remedy other than pecuniary compensation.

[44] The Italian Supreme Court (Corte di Cassazione) requested a preliminary ruling to the Court of Justice of the European Union (CJEU) regarding the right of associations to bring legal proceedings without acting in the name of a specific compliant or in the absence of an identifiable compliant (Cass., 20.07.2018, no. 19443). The case—where the question arose—was related to the application of Directive 2000/78/CE for discrimination on the grounds of sexual orientation and access to work, and may be considered as another hate speech example. In July 2013, Mr. Taormina, a lawyer and former member of the Italian Parliament, declared during a radio programme that he would never hire a gay lawyer at his law firm and made other offensive comments about homosexual people. An association composed by lawyers, promoting and defending LGBTI people's rights, brought an action before a local tribunal, claiming that those statements constituted discrimination prohibited by Legislative Decree no. 216/2003 implementing Directive 2000/78/EC on equal treatment in employment and occupation. Both the Tribunal of Bergamo (06.08.2014) and, then, the Court of Appeal (App. Brescia, 11.12.2014) ruled in favour of the association and penalised Mr. Taormina with a fine of €10,000, as well as ordering the publication of the decision in a national newspaper at his expense (which is a very common method of non-pecuniary relief). Seven years following Mr. Taormina's comments, the controversy continues. Indeed, the Italian Corte di Cassazione asked to the CJEU if Article 9 of the Directive no. 2000/78/CE must be interpreted as meaning that an association, which in its statute declares the aim of promoting the culture and the respect of LGBTI people's rights, should be automatically considered as having a legitimate interest in the compliance of the Directive. In the decision of 23rd April 2020, the CJEU concluded that national legislation can include profit or non-profit organisations among the subjects with legal standing to bring actions for the enforcement of obligations under the Directive. Those organisations may also obtain damages, in circumstances that are capable of constituting discrimination, within the meaning of that Directive, even if it is not possible to identify an injured party.

Three examples will be provided below. They are all cases of civil proceedings brought within a criminal trial for hate speech, both by a representative association of a protected group and by individuals not directly targeted in the offensive expression but identifiable members of the targeted, protected group. The three decisions perfectly fit in the general framework of hate speech case law, where hate speech is defined as an expression that promotes hatred against traditionally disadvantaged groups.[45]

All the examined decisions may be significant to the discussion in this paper for three reasons. First, they indicate a clear link between anti-discrimination law and hate speech cases, which may help detect the area of law most appropriate to supply adequate remedies. Second, the traditional remedy of damages may result in a sort of dissatisfaction, because of an inability to fully realise the corrective justice function of civil liability. Third, the analysis of the judicial decisions shows that the subsequent conduct of the wrongdoer has consequences in the adjudication process, and that suggests some conclusion regarding the suitability of remedies, such as apologies, for torts that protect the dignity of the victim.

a) The first case is related to a petition organised by a regional councillor of the Veneto Region 'to send gypsies away from our city', intended to evict a Roma camp from the city of Verona.[46] Aside from the criminal sanctions imposed (the authors were initially condemned for instigating others to commit an act of discrimination for racial, ethnic, national, or religious motives), the civil parties in the judicial proceedings (seven members of the Sinti community and Opera Nomadi, an association for the integration of Roma minorities in Italy) were awarded approximately €50,000 for non-monetary damages. This was a very controversial case (and a very complex judicial affair, concluded with the decision of the Cassazione penale no. 41819, 10 July 2009), especially because the defendants asserted that the petition constituted political speech, which is strongly protected in liberal democracies.

b) The second example is a 2015 decision of the Civil Court of Rome. In the context of discrimination and cases of hate speech towards the Roma communities, this decision seems an important signal to contrast the stereotypes used against these

[45] For a similar definition of hate speech see Mayo Moran, 'Talking about Hate Speech: A Rhetorical Analysis of American and Canadian Approaches to the Regulation of Hate Speech' (1994) *Wisconsin Law Review* 1425, at 1430.

[46] Italian Supreme Court of Cassation, Criminal Section I, judgment no. 41819 of 10 July 2009 (Tosi and others), C.E.D. Cassazione, rv. 244812.

communities, damaging the personality rights of their members.[47] The decision concerns a textbook for the preparation for the Bar exam, containing legal opinions. One concerned the offenses of receiving and careless purchase pursuant to Article 712 of the Criminal Code (entitled 'Purchase of things of suspected origin'). In the examples of subjective circumstances that should give rise to the suspicion that the thing proceeds from a crime, the author of the opinion indicated, in particular, the purchase from 'a beggar, from a gypsy or from a convicted person'. With the first instance decision, the Civil Court of Rome accepted the initial request, in which a woman of Roma ethnicity, the anti-discrimination group Associazione 21 luglio, and the Association for Juridical Studies on Immigration (ASGI) asked to declare discriminatory the reference to gypsies. The Court also ordered the Publishing Group Simone and the author of the publication:

> to cease the discriminatory behaviour, providing for the withdrawal from the market of the publication or subsequent editions bearing the same content and, in the case of subsequent publications, to the elimination of the expression 'when the thing, despite its considerable value, is offered for sale by a gypsy' in the treatment of the circumstances of the criminal origin of the property as a constitutive element of the crime referred to in art. 712 Criminal Code.[48]

Finally, the Court ordered the publishing house to pay a financial compensation for non-monetary damages of €1,000 to the Roma woman.

c) The Court of Milan was called to adjudicate on the request for compensation submitted by two associations—both active in the field of anti-discrimination—against the mayor of a small town in the province of Veneto, who, during a radio broadcast, made heavily offensive statements about Roma people as well as migrants of African origin.[49] The use of offensive expressions, which exclude the merely 'political' value of the manifestation of thought, in the opinion of the judge of Milan constituted discriminatory behaviour, suitable to create a hostile environment and to degrade the dignity of the social groups involved. The manifestation of thought, in this case, determined an injury to equal social dignity and to the right to non-discrimination. The anti-discrimination approach also sanctions behaviours that are

[47] Civil Court of Rome, First Instance Decision, 6 March 2015, RG n. 53220/2014.

[48] See ASGI, 'Casa editrice condannata per pubblicazione che discrimina i rom' (10 March 2015), available at www.asgi.it/antidiscriminazione/casa-editrice-condannata-per-pubblicazione-che-discrimina-i-rom/ (last accessed 28 July 2025) (in Italian).

[49] Civil Court of Milan, First Instance Decision, 9 February 2016, RG n. 30595/2014.

only potentially harmful, but that may still create an intimidating, hostile, degrading, humiliating and offensive environment. This circumstance may move the frontier of civil remedies from compensation to actually repairing the damage done. The Court did not attribute any weight to subsequent declarations by the mayor, in which—though not apologising—he seemed to deny his own statements. The Court of Milan therefore ordered the compensation of non-monetary damages in favour of the two associations, for €6,000 each, taking into account the seriousness of the harassment, its reiteration, and the institutional role of the injurer.

5. Traditional and New Remedies: The Role of Apologies

In all the above cases, a sense of disappointment or at least of incomplete satisfaction arises, assuming that monetary damages cannot be considered a sufficient remedy for any kind of harm. Drawing a clear line between what is allowed and what causes an unlawful harm is never easy (and it is impossible here to consider the constitutional implications of this discourse). To give an example, if someone communicates to their virtual friends on the internet that, in their opinion, homosexual conduct is reprehensible and homosexuality is a disease, we may reckon that they are legitimately manifesting their own thoughts, however retrograde, unscientific, and culturally backward; in the same way, manifestations of hostility towards migrants could certainly not be punished as such, but only when they turn into rough insults, utterly intolerable in a civilised community.

The line between what legitimises a compensation claim and what remains guaranteed by freedom of expression is, however, subtle and there are many cases that resist a clear framework. One example is white supremacy, an opinion tolerated in the US even when it takes the form of the Ku Klux Klan's cross burnings, as long as members of that group do not express any threat towards determined subjects.

This is the broadest problem of hate speech. But even if we manage to provide a solution, the issue of appropriate civil remedies remains. When a court states that the author of a case of hate speech is liable, the aggrieved party may not primarily be seeking damages,[50] but instead a restoration of their group's reputation. If the harm of hate speech is in itself, this becomes even more plausible, and targeted persons

[50] In his seminal work, 'Words that Wound' (no 39, at 179), Richard Delgado suggests the opportunity of a specific tort for racist speech. The plaintiff should prove: a) the language was intended to demean through reference to race; b) the plaintiff understood it as intended, to demean through reference to race; and c) a reasonable person would recognise the speech as a racial insult.

may just want the court to re-establish their social standing in a way other than a mere award of damages. In any case, the very idea of a civil remedy may operate as a sort of assurance offered by a society to all its members:[51] every instance of abuse or defamation or humiliation on grounds of race or ethnicity, religion, or sexual orientation will be taken seriously by the legal system.

In countries that give standing to organisations and associations specialised in the promotion of the right to equality, another problem arises: it is difficult to prove these organisations suffer any authentic non-pecuniary loss;[52] the aim of all these organisations could be more effectively achieved with different remedies and in particular with those remedies that restore the image of the group within the society. Here, again, the complex nature of hate speech laws requires a careful detection of which vulnerable minorities—and, consequently, their representative organisations—deserve protection.[53]

Hence, various types of non-pecuniary relief can be considered: apologies, publication of court decisions, the right of reply, or retractions. The latter two seem to suit defamation cases, while apologies and publication of court decisions could serve hate speech.[54] In choosing among these remedies, the proportionality principle can be a guiding criterion. It ensures that the remedy adopted is not only appropriate to

[51] See Waldron (no 35, at 85).

[52] The idea of 'non-pecuniary loss' means any damage causing mental distress, grief, or psychosomatic complaints that may affect the victim of a tort (so-called *pretium doloris*). The Italian civil liability system is complicated by the particular provision of Article 2059 of the Italian Civil Code, which sets forth that compensation for any non-pecuniary loss is admissible only in cases laid down by law. From a substantive viewpoint, such cases are those where the injurious conduct causing such damage is a criminal offense (cf. Article 185 of the Italian Criminal Code) or where an injury to a constitutionally protected interest occurs. Indeed, the scope of non-pecuniary damage now encompasses any violation of a constitutional interest. For example, a physical or psychological injury does not cause a loss of wealth per se but, since it represents a violation of the right to health protected by Article 32 of the Constitution the claimant is now able to recover damages for non-pecuniary losses as well.

[53] In the absence of agreement on what hate speech is, some authors use a narrow definition: messages of racial inferiority directed against historically oppressed and subordinated groups. See Mari J. Matsuda, 'Public Response to Racist Speech: Considering the Victim's Story' (1989) 87 *Michigan Law Review* 2320, at 2323.

[54] For some examples of court-ordered apologies in a South African hate speech case, see Marelize Marais, 'A constitutional perspective on the Sparrow judgements' (2017) 42 *Journal for Juridical Science* 25–64.

In Australia, in *Burns v. Radio 2UE Sydney Pty Ltd (No 2)* [2005] NSWADT 24, the NSW Anti-Discrimination Tribunal defined a court-ordered apology as an acknowledgement of 'wrongdoing' that is distinguished from a personal apology which is 'sincere and which is incapable of being achieved by a court order'. The legitimacy of apologies made under compulsion has been frequently questioned in the legal doctrine: see Robyn Carroll, 'You Can't Order Sorriness. So is there Any Value in an Ordered Apology? An

redress the harm suffered, but also the least intrusive means to protect the dignity of the victim while respecting the broader framework of fundamental rights.

In the civil liability domain, part of the legal doctrine considers that the only function of compensation of non-patrimonial damage lies in the contrast of pain and suffering due to the wrongful behaviour. A different and quite common opinion considers the main function of non-pecuniary damage being deterrence and punishment, rather than compensation. Hate speech case law can be of some help in shaping the boundaries of compensation of non-patrimonial damage, especially because judges take into consideration many different elements, which are strictly linked to the function of civil liability: a) the seriousness of the behaviour; b) the psychological element (negligence, gross negligence, intention); c) the eventual enrichment obtained as a consequence of the wrongful behaviour; and d) the financial status of the wrongdoer. All these elements seem to reveal another function of non-patrimonial damages, which is its punitive scope.[55]

At the same time, the remedies traditionally used clearly show a very limited expressive or restorative power. Restoring the targeted group's reputation (or dignity, or honour) could instead be achieved by compelling the wrongdoer to take back their words and apologise for spreading them. Of course, such a remedy is not intended to replace compensatory mechanisms where hate speech leads to physical violence or provable psychological injury. Rather, apologies may serve a distinct function in cases where the harm is primarily symbolic, affecting the victim's status or social standing, and where the wrong consists in the degradation of equality and dignity rather than in personal injury in the strict sense. In Italian case law, the importance of apologies is limited, since they can be considered as a mitigating tool in the assessment of non-pecuniary damage;[56] conversely, if the wrongdoer, instead of apologising, repeats the offensive conduct (again, as in the third case considered in the previous section), the damages granted by the court may be augmented as a consequence of an additional harm.[57]

Analysis of Ordered Apologies in Anti-Discrimination Cases' (2010) 33 *UNSW Law Journal* 360; Gijs van Dijck, 'The Ordered Apology' (2017) 37 *Oxford Journal of Legal Studies* 562–587.

[55] According to Katherine Gelber & Luke McNamara, 'The Effects of Civil Hate Speech Laws: Lessons from Australia' (2015) 49 *Law & Society Review* 631, at 637, in case of unlawful hate speech, in the Australian experience, judicial orders of compensation are typically about 10,000 AUD.

[56] See Nicola Brutti, 'Legal Narratives and Compensation Trends in Tort Law: The Case of Public Apology' (2013) 24 *European Business Law Review* 127.

[57] This very limited interest in apologies may be misleading if we consider the function of the tort law system and the specific nature of the harm in hate speech cases. These features may support the case for

An apology for hate speech should comprise an expression of regret and a promise to change behaviour in the future. In this sense, an apology is different from a rectification in a defamation case:[58] an instruction to retract a defamatory or false statement is the most common method to deal with injurious falsehoods, but in hate speech cases there is no room for rectification, only for regret and apology.[59]

Criticisms have been raised regarding the appropriateness of an apology as a remedy. Richard Delgado considers that:

> the mild nature of an apology renders it almost wholly inadequate to redress and deter hate crimes and hate speech. It atomises the harm, fails to consider its power dimension, and leaves the perpetrator intact and the victim un-redressed. It cuts off consideration of broader issues and encourages all concerned to go on with their lives as if nothing very serious has happened.[60]

However, these criticisms do not consider that it is often the very victim of hate speech who wants the author's apology, which then meets the plaintiff's psychological needs. This explains why an apology may well be an adequate legal remedy when it is expressly sought (judicially or at an extrajudicial stage) by the offended party. In the Italian judicial system, which has long been excessively clogged, one must not forget the importance of the deflationary function of the apology, which may reduce litigation in tortious liability.[61]

---

introducing an apology-protecting provision into Italian tort law. Although this may appear counterintuitive from a common law perspective, it should be recalled that the development of Italian tort law is mainly case-law driven, making statutory innovation in this area relatively rare. The debate is still in its early stages, and no empirical evidence is currently available.

[58] The use of apology as a means for re-establishing the victim's social standing after a defamation has been suggested in the legal doctrine. See Vandenbussche (no 14); Robyn Carroll & Jeffrey Berryman, 'Making Amends by Apologising for Defamatory Publications: Developments in the Twenty-First Century', in Kit Barker, Karen Fairweather & Ross Granthama (eds), *Private law in the 21st century* (Hart Publishing, 2006) 480–481.

[59] Since an apology conveys a particular message of regret, it may be possible to consider it as a sort of 'more speech', that actively contributes to the 'marketplace of ideas'.

[60] Richard Delgado & Jean Stefancic, 'Apologize and Move On? Finding a Remedy for Pornography, Insult, and Hate Speech' (1996) 67 *University of Colorado Law Review* 93, at 99.

[61] A new generation of scholars correctly tend to defend the role that litigation can play in promoting democratic values (among others, see Alexandra Lahav, *In Praise of Litigation* (Oxford University Press, 2017)). Notwithstanding that lawsuits are the central mechanism for enforcing and protecting rights, this has to deal not only with practical difficulties arising from inefficiencies of the system, but also with the inevitable cost of any litigation.

Apologies are also a fundamental part of remedies as they diminish the sense of humiliation created,[62] redress humiliation for the victim, shame the offender, and help to heal the emotional wounds associated with a wrong (as Professor Prue Vines has argued, apologies are 'a tool of communication and of emotion'[63]). Indeed, apologies transfer the sense of humiliation to the author of the hate speech, while remaining aware of Martin Luther King's lesson that winning the enemy's friendship and understanding is better than defeating and humiliating them. An apology can thus play a more general role, not only limited to the wallet of the representative association or the individual belonging to the group; corrective justice acquires a wider dimension, potentially reaching the whole audience of the addressees of the speech[64] and even reducing the propensity to sue in a hate speech context.[65]

Sometimes an order to apologise can be issued within alternative dispute resolution (ADR) mechanisms, expressly constructed for issues related to discrimination.[66] In many jurisdictions, there are specific commissions or anti-discrimination boards set up to protect people from discriminatory conduct. Once hate speech has been determined, they seek conciliation through orders for cessation of conduct, or public apologies, or retractions. In particular, the involvement of such organisations may be of great help in the process of offer and acceptance of an apology, making mediation and ADR ideal fields for meaningful apologies.[67] In the Italian legal system, informal

[62] This circumstance makes clear the link between apology and hate speech. Indeed, as has been noted, 'apologies help victims regain their sense of self-worth and dignity, both of which are essential aspects of psychological well-being': Brent T. White, 'Say You're Sorry: Court-Ordered Apologies as a Civil Rights Remedy' (2006) 91 *Cornell Law Review* 1261, at 1274. In the same sense, Richard Abel, *Speech and Respect* (Sweet & Maxwell, 1994) 146.

[63] Prue Vines, 'The Power of Apology: Mercy, Forgiveness or Corrective Justice in the Civil Liability Arena?' (2007) 1 *Public Space* 1, at 15.

[64] Examples of court orders to publish rectifications containing apologies in the same social media where a defamatory statement was made are provided by Andrea Zwart-Hink, Arno Akkermans & Kiliaan Van Wees, 'Compelled Apologies as a Legal Remedy: Some Thoughts from a Civil Law Jurisdiction' (2014) 38 *University of Western Australia Law Review* 100, at 112.

[65] In this perspective, apologies may contribute in avoiding litigation and promoting settlement. See Jennifer K. Robbennolt, 'Apologies and Legal Settlement: An Empirical Examination' (2003) 102 *Michigan Law Review* 460, at 463.

[66] An apology order is available as a remedy for discrimination under Australian legislation. See Robyn Carroll, 'The Ordered "Apology" as a Remedy Under Anti-Discrimination Legislation in Australia: An Exercise in Futility?', in Russell Weaver & François Lichère (eds), *Recognition and Enforcement of Judgements* (Presses Universitaires d'Aix-Marseille, 2010) 362.

[67] For a discussion on the role of apologies in the context of mediation see Deborah L. Levi, 'The Role of Apology in Mediation' (1997) 72 *NYU Law Review* 1165; Lee Taft, 'Apology Subverted: The Commodification of Apology' (2000) 109 *Yale Law Journal* 1135 (arguing that apologies are ineffective if they are not

conciliation mechanisms, with general reference to discriminatory practices, are also included among the competences of the UNAR. This equality body is tasked with receiving reports and complaints from victims of discriminatory conduct, including those who may be victims of hate speech, and acts as an informal mediator. In this role, UNAR may solicit a settlement of the conflict that can also occur through the request for the removal of discriminatory behaviour and even by means of a public apology from its author or an agreement that provides for a sum in compensation. But if this phase ends without a satisfactory agreement, similar mechanisms could also be envisaged in court, which represent a form of relief for hate speech victims and which can be added to the provision of compensation for the non-pecuniary damage suffered.

In conclusion, I believe that the availability of a judicial decision mandating an apology holds intrinsic importance, as it provides targeted communities with assurance that the law can serve as a safeguard against discrimination—a form of legal vindication. However, I disagree with scholars who argue that apologies carry an educational function. While some suggest that apologies, through judicial precedents, can dissuade hate speakers, the broader educational role of judicial decisions appears highly problematic. Such a role risks fostering the indoctrination of prevailing ideologies, particularly when freedom of expression is at stake. Nonetheless, scepticism about the educative value of apologies does not preclude their potential to prompt wrongdoers to reflect on the harm caused or to reinforce a collective rejection of such conduct.

At the same time, monetary compensation alone seems insufficient as a remedy, given the challenges of quantifying the harm caused and its limited impact on the public sphere—especially when hate speech is understood as constituting harm rather than merely causing it. In this regard, apologies may be viewed as a form of corrective justice, functioning as a means to redress and restore balance between the perpetrator and the victim of hate speech. Furthermore, apologies exhibit a clearer restorative purpose. In the context of hate speech, this corrective justice framework addresses the status injury inflicted by the speech, wherein the targeted group is degraded and the offender elevates themselves at the victim's expense, thus working to restore equality and dignity.

sincere); Robyn Carroll, Alfred Allan & Margaret Halsmith, 'Apologies, Mediation and the Law: Resolution of Civil Disputes' (2017) 7 *Oñati Socio-legal Series* 569 (examining possible concerns associated with apologies offered in mediation).

Against this background, a few concluding observations may be drawn. Civil remedies—particularly apologies—should not be viewed as marginal or symbolic responses, but as tools with the potential to restore dignity and reaffirm community standards. While further development is needed, their integration into legal practice may enrich the repertoire of responses to hate speech beyond the limits of monetary compensation alone.

References

Abel R., *Speech and Respect* (Sweet & Maxwell, 1994)

Adelman L. & J. Deitrich, 'Extremist Speech and the Internet: The Continuing Importance of Brandenburg' (2001) 4 *Harvard Law and Policy Review* 361–375.

Arendt H., *Eichmann in Jerusalem: A Report on the Banality of Evil* (New York: Viking Press, 1963).

Barak A., *Proportionality. Constitutional Rights and Their Limitations* (Cambridge University Press, 2012).

Bleich E., *The Freedom to be Racist? How the United States and Europe Struggle to Preserve Freedom and Combat Racism* (Oxford University Press, 2011).

Boyle K., 'Hate Speech -The United States Versus the Rest of the World' (2001) 53(2) *Maine Law Review* 487–502.

Brutti N., 'Legal Narratives and Compensation Trends in Tort Law: The Case of Public Apology' (2013) 24 *European Business Law Review* 127–148.

Burchell J. 'Retraction, Apology and Reply as Responses to iniuriae' in E. Descheemaeker & H. Scott (eds), *Iniuria and the Common Law* (Hart Publishing, 2013) 197–214.

Carmi G.E., 'Dignity versus Liberty: The Two Western Cultures of Free Speech' (2008) 26 *Boston University International Law Journal* 277–374.

Carroll R., 'The Ordered 'Apology' as a Remedy Under Anti-Discrimination Legislation in Australia: An Exercise in Futility?', in R. Weaver & F. Lichère, *Recognition and Enforcement of Judgements* (Presses Universitaires d'Aix-Marseille 2010) 362–435

Carroll R., 'You Can't Order Sorriness. So is there Any Value in an Ordered Apology? An Analysis of Ordered Apologies in Anti-Discrimination Cases' (2010) 33 *UNSW Law Journal* 360–385.

Carroll R., A. Allan & M. Halsmith, 'Apologies, Mediation and the Law: Resolution of Civil Disputes' (2017) 7 *Oñati Socio-legal Series* 569–600.

Carroll R. & J. Berryman, 'Making Amends by Apologising for Defamatory Publications: Developments in the Twenty-First Century', in K. Barker, K. Fairweather & R. Granthama (eds), *Private law in the 21st century* (Hart Publishing, 2006) 479–499.

Delgado R. & J. Stefancic, 'Apologize and Move On? Finding a Remedy for Pornography, Insult, and Hate Speech' (1996) 67 *University of Colorado Law Review* 93–111.

Delgado R. & J. Stefancic, *Understanding Words that Wound* (Routledge, 2018).

Delgado R., 'Words that Wound: A Tort Action for Racial Insults, Epithets, and Name Calling' (1982) 17 *Harvard Civil Rights - Civil Liberties Law Review* 133–181.

Fraisseix P., 'Le Droit mémoriel' (2006) *Revue française de droit constitutionnel* 483–508.

Frangi M., 'Les lois mémorielles: de l'expression de la volonté générale au législateur historien' (2005) *RD publ* 241–266.

Fronza E., 'The Criminal Protection of Memory: Some Observations about the Offense of Holocaust Denial', in L. Hennebel & T. Hochmann, *Genocide Denial and the Law* (Oxford University Press, 2011) 155–181.

Gelber K. & L. McNamara, 'The Effects of Civil Hate Speech Laws: Lessons from Australia' (2015) 49 *Law & Society Review* 631–664.

Habermas J. & J. Ratzinger, *The Dialectics of Secularization: On Reason and Religion* (Brian McNeil tr, Ignatius Press, 2006)

Hallebeek J. & A. Zwart-Hink, 'Claiming apologies: a revival of amende honorable?' (2017) 5 *Comparative Legal History* 194–242.

Hare I. & J. Weinstein (eds), *Extreme Speech and Democracy* (OUP, 2009).

Hemmer Jr J., 'Hate speech - The egalitarian/libertarian dilemma' (1995) 5 *Howard Journal of Communications*, 307–330.

Herz M. & P. Molnar (eds), *The Content and Context of Hate Speech: Rethinking Regulation and Responses* (Cambridge University Press, 2012).

Heyman S.J., 'Hate Speech, Public Discourse, and the First Amendment', in I. Maitra & M. K. McGowan (eds), *Speech and Harm* (Oxford University Press, 2012) 158–181.

Kahn R., 'Holocaust Denial and Hate Speech', in L. Hennebel & T. Hochmann (eds), *Genocide Denial and the Law* (Oxford University Press, 2011) 77–107.

Lahav A., *In Praise of Litigation* (Oxford University Press, 2017).

Lepoutre M., 'Hate Speech Laws: Expressive Power is not the Answer' (2020) 25(4) *Legal Theory* 272–296.

Levi D. L., 'The Role of Apology in Mediation' (1997) 72 *NYU Law Review* 1165–1210.

MacKinnon C., *Only Words* (Harvard University Press, 1993)

Maitra I. & M. K. McGowan, 'Introduction and Overview', in I. Maitra & M. K. McGowan (eds), *Speech and Harm* (Oxford University Press, 2012) 4–8.

Marais M., 'A constitutional perspective on the Sparrow judgements' (2017) 42 *Journal for Juridical Science* 25–64.

Matsuda M.J., 'Public Response to Racist Speech: Considering the Victim's Story' (1989) 87 *Michigan Law Review* 2320–2381.

Matsuda M.J., Ch. R. Lawrence, R. Delgado, K. Williams Crenshaw, *Words That Wound: Critical Race Theory, Assaultive Speech, and the First Amendment* (Westview Press, 1993).

Mehlman M. J., 'Promoting Fairness in the Medical Malpractice System' in W. M. Sage & R. Kersh (eds), *Medical Malpractice and the U.S. Health Care System* (Cambridge University Press, 2009) 137–153.

Michel J., *Gouverner les mémoires – Les politiques mémorielles en France* (Presses Universitaires de France, 2010).

Monateri P.G. (ed.), *Methods of Comparative Law* (Elgar, 2012).

Moran M., 'Talking about Hate Speech: A Rhetorical Analysis of American and Canadian Approaches to the Regulation of Hate Speech' (1994) *Wisconsin Law Review* 1425–1514.

Möschel M., 'Italy's (Surprising) Use of Racial Harassment Provisions as a Means of Fighting Discrimination', in B. Havelková & M. Möschel (eds), *Anti-Discrimination Law in Civil Law Jurisdictions* (Oxford University Press, 2019) 279–308.

Nicola F., 'Family Law Exceptionalism in Comparative Law' (2010) 58 *American Journal of Comparative Law* 777–810

Popper K.R., *The Open Society and its Enemies, I, The Spell of Plato* (Routledge, 1945).

Post R., 'Hate Speech' in I. Hare & J. Weinstein (eds), *Extreme Speech and Democracy* (Oxford University Press, 2009) 123–138.

Robbennolt J.K., 'Apologies and Legal Settlement: An Empirical Examination' (2003) 102 *Michigan Law Review* 460–516.

Sacco R., 'Legal Formants: A Dynamic Approach to Comparative Law. (Installment I of II)' (1991) 39 *American Journal of Comparative Law* 1–34.

Samuel G., *Epistemology and Method in Law* (Ashgate, 2003).

Schweppe J. & M. Austin Walters (eds), *The Globalization of Hate: Internationalizing Hate Crime?* (Oxford University Press, 2016).

Taft L., 'Apology Subverted: The Commodification of Apology' (2000) 109 *Yale Law Journal* 1135–1160.

Tsesis A., 'Dignity and Speech: The Regulation of Hate Speech in a Democracy' (2009) 44 *Wake Forrest Law Review* 497–532.

Tushnet M., *Advanced Introduction to Freedom of Expression* (Elgar, 2018).

van Dijck G., 'The Ordered Apology' (2017) 37 *Oxford Journal of Legal Studies* 562–587.

Van Hoecke M., *Epistemology and Methodology of Comparative Law* (Hart Publishing, 2004).

Vandenbussche W., 'Rethinking Non-Pecuniary Remedies for Defamation: The Case for Court-Ordered Apologies', (2020-21) 9 *Journal of International Media & Entertainment Law* 109–170.

Viggiano A., 'Giustizia, disciplina e ordine pubblico – Storia di Venezia', Enciclopedia Treccani, 1996.

Vines P., 'The Power of Apology: Mercy, Forgiveness or Corrective Justice in the Civil Liability Arena?' (2007) 1 *Public Space* 1–29.

Waldron J., 'Dignity and Defamation: The Visibility of Hate' (2010) 123 *Harvard Law Review* 1596–1657.

Waldron J., *The Harm in Hate Speech* (Harvard University Press, 2012).

Walker S., *Hate Speech: The History of an American Controversy* (Nebraska University Press, 1994)

White B. T., 'Say You're Sorry: Court-Ordered Apologies as a Civil Rights Remedy' (2006) 91 *Cornell Law Review* 1261–1311.

Williams P.J., 'Spirit-Murdering the Messenger: The Discourse of Fingerpointing as the Law's Response to Racism', (1987) 42 *University of Miami Law Review* 127–157.

Zwart-Hink A., A. Akkermans & K. Van Wees, 'Compelled Apologies as a Legal Remedy: Some Thoughts from a Civil Law Jurisdiction' (2014) 38 *University of Western Australia Law Review* 100–122.

Internet Service Providers: Key Players in the New e-Evidence Paradigm?

Vagia Polyzoidou[1]

Abstract

This paper examines the European Union (EU) e-Evidence package—a set of instruments comprising a new Directive and Regulation on 'European Production and Preservation Orders for electronic evidence in criminal proceedings and for the execution of custodial sentences following criminal proceedings'—which is expected to enter into force in 2026. It analyses the package's provisions and likely effects on cross-border criminal investigations and prosecutions in the EU. The e-Evidence package significantly transforms European criminal procedural law, building on the 2018 framework through new tools that employ the same technologies often used to plan and commit offences. The analysis traces the development of this new era of mutual trust and assistance, focusing on the expanding role of private entities—especially internet service providers (ISPs) —in obtaining and transmitting electronic evidence. While the new measures may enhance prosecutorial efficiency and cross-border cooperation, such progress must not come at the expense of the protection of fundamental human rights.

Keywords: electronic evidence (e-Evidence); internet service providers (ISPs); European Production Order (EPOC); European Preservation Order (EPOC-PR); cybercrime

1. Introduction

The rapid rise[2] of new technologies—in combination with the recent unprecedented crisis and dramatic political changes[3]—has exposed the legal shortcomings in addressing cybercrime.[4] In response, the European Union (EU) has sought to make

[1] Associate Professor in Criminal Law and Criminal Procedure, University of Nicosia.

[2] According to SWD (2018) 118 final on the Impact Assessment accompanying the e-Evidence package (17.4.2018), 85% of criminal investigations involve digital data, available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52018SC0118 (last accessed 6 December 2025).

[3] Valsamis Mitsilegas, Elspeth Guild 'Police and criminal justice co-operation after Brexit' (2023) 30 (11) *Journal of European Public Policy* 251.

[4] See, inter alia, Jonathan Clough, *Principles of Cybercrime* (Cambridge University Press: Cambridge,

its legal framework more comprehensive, building on the mechanisms established in 2018—international judicial cooperation and mutual legal assistance (MLA). The updated framework introduces new tools, including the 2023 e-Evidence package.[5] These tools aim to address challenges faced by investigators, such as the permanent storage of data, data location shifts, and how to deal with non-EU service providers. A key question remains: What is the cost to these legal advancements?

2. The e-Evidence Context[6]

For decades, the EU lacked a common, binding definition of e-Evidence,[7] despite efforts to define it.[8] This gap in definition extended to the most complex category of digital data used in criminal investigations:[9] content data.[10] It was also an ongoing

2015) 9; Majid Yar, Kevi Steinmetz, *Cybercrime and Society* (Sage: London, 2019) 3; Matthew Williams, *Virtually Criminal - Crime, deviance and regulation online* (Oxon: Routledge, 2006) 70; Ulrich Sieber, 'Computer Crimes, Cyber – Terrorism, Child Pornography and Financial Crimes (General Report)', in D. Spinellis (ed.), *Computer Crimes, Cyber – Terrorism, Child Pornography and Financial Crimes*: *Reports presented to the Preparatory Colloquium for the Round Table II of the 17th International Congress of Penal Law (Beijing, 2004)* (Athens: Antonis Sakkoulas, 2004) 16; David Speer, 'Redefining Borders: The challenges of Cybercrime' (2000) 34 *Crime, Law and Social Change* 259; Nadine Strossen, 'Cybercrimes vs Cyberliberties'(2000) 14(1) *International Review of Law Computers and Technology* 11.

[5] European Commission, 'cross border access to electronic evidence' available at https://commission.europa.eu/law/cross-border-cases/judicial-cooperation/types-judicial-cooperation/e-evidence-cross-border-access-electronic-evidence_en (last accessed 6 December 2025).

[6] Technological developments and the resulting updates to EU instruments and tools has created a terminological black hole in the field of criminal procedure, as Audrito & al. pointed out, trying to help the legal experts via a statistical research and data analysis. See Davide Audrito & al., 'Analogical lightweight ontology of EU criminal procedural rights in judicial cooperation' (2023) 31 *Artificial Intelligence and Law* 640.

[7] As a result, electronic evidence was not necessarily tied to electronic crime, cybercrime, or e-crime, nor was it aimed at proving the existence of cybercrime, see 2001 Convention on Cybercrime adopted within the framework of the CoE.

[8] Compare to the Electronic Evidence Guide (which resulted from a joint CoE and EU project), in which it is defined as 'any information generated, stored or transmitted in digital form that may later be needed to prove or disprove a fact disputed in legal proceedings' in Nigel Jones & al., 'Electronic Evidence Guide. A Basic Guide for Police Officers, Prosecutors and Judges' (Council of the EU 2020), available at https://rm.coe.int/c-proc-electronic-evidence-guide-2-1-en-june-2020-web2/16809ed4b4 (last accessed 6 December 2025).

[9] Sofie Depauw, 'Electronic evidence in criminal matters: how about e-Evidence instruments 2.0?', *EuCLR* (2018) 8(1), available at https://doi.org/10.5771/2193-5505-2018-1-62, 66–68. For the discussion between Member States and ECtHR see ibid, fn 33–37.

[10] Following the 2023 regulatory updates, Article 3.12 of the Regulation now provides a clear definition of content data, specifying that it can be used to identify individuals or gather further information about their activities.

challenge during the negotiations for the new e-Evidence framework, where EU institutions could not agree on much of the terminology;[11] ultimately a definition of electronic evidence was reached and can be found in Article 3.8 of the e-Evidence Regulation: 'subscriber data, traffic data or content data stored by or on behalf of a ISP, in an electronic form, at the time of the receipt of a European Production Order Certificate (EPOC) or of a European Preservation Order Certificate (EPOC-PR)'.[12]

3. Internet Service Providers (ISPs)' role

3.1. Past Insufficient Attempts

EU criminal procedural law has long addressed cross-border cooperation in an effort to overcome challenges posed by territorial borders, differing jurisdictions, and the inevitable globalisation of electronic information.[13] From the Schengen Implementation Convention (1990) to the Amsterdam Treaty (1997) and the Lisbon Treaty (2009), judicial cooperation and MLA have continuously evolved.[14]

[11] Theodore Christakis, 'From Mutual Trust to the Gordian Knot of Notifications The EU e-Evidence Legislative Project in Vanessa Franssen, Stanislaw Tosza (eds), *The Cambridge Handbook of Digital Evidence in Criminal Matters* (Cambridge, UK: Cambridge University Press, 2023) https://ssrn.com/abstract=4306874 (in press) 9 and fn 37–40 where the author explains how the Commission's definition in the initial draft ('evidence stored in electronic form by or on behalf of a ISP at the time of receipt of a production or preservation order certificate, consisting in stored subscriber data, access data, transactional data and content data') was declined by Rapporteur Birgit Sippel as 'it could automatically imply that the data gathered is admissible as evidence in a criminal proceeding', suggesting replacing the term with 'a more neutral terminology, namely "electronic information"'. The European Parliament ultimately accepted this proposal, transforming the title (see 'European Production and Preservation Orders for electronic *information* in criminal *proceedings*' (emphasis author's own)). See also, Sippel's report (2019), https://www.europarl.europa.eu/doceo/document/LIBE-PR-642987_EN.pdf (last accessed 6 December 2025).

[12] Regulation (EU) 2023/1543 of the European Parliament and of the Council of 12 July 2023 on European Production Orders and European Preservation Orders for electronic evidence in criminal proceedings and for the execution of custodial sentences following criminal proceedings available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32023R1543 (last accessed 23 January 2026).

[13] Evidence includes information but information is not necessarily evidence, See Christakis (no 11).

[14] For the digitalisation of judicial cooperation, the EU introduced two legal instruments in December 2023: a) Directive (EU) 2023/2843 amending Directives 2011/99/EU and 2014/41/EU of the European Parliament and of the Council, Council Directive 2003/8/EC and Council Framework Decisions 2002/584/JHA, 2003/577/JHA, 2005/214/JHA, 2006/783/JHA, 2008/909/JHA, 2008/947/JHA, 2009/829/JHA and 2009/948/JHA, as regards digitalisation of judicial cooperation; and b) Regulation (EU) 2023/2844 on the digitalisation of judicial cooperation and access to justice in cross-border civil, commercial and criminal matters, and amending certain acts in the field of judicial cooperation.

Despite these developments, the EU's legal framework for gathering electronic evidence and MLA in criminal investigations were outdated mechanisms,[15] even after the adoption of Directive 2014/41/EU, which introduced the European Investigation Order (hereinafter EIO) in criminal matters.[16] Although authorities were granted access to electronic data, provided appropriate procedural safeguards were in place, complications arose when the internet service provider (ISP) was located outside the country or the data was stored abroad. In such cases, law enforcement had to rely on cross-border cooperation tools, particularly the EIO within the EU's area of freedom, security, and justice, and MLA outside this area—especially for content data held by US companies.[17]

The main challenge in regulating this framework lay in whether to opt for direct cooperation with ISPs or indirect cooperation between judicial authorities. Accessing e-Evidence is often a confusing process for authorities, as data may be stored in another country, with ISPs maintaining servers in multiple locations, both within and outside the EU. This forces judicial authorities, who need to collect e-Evidence, to navigate lengthy and complex procedures to gain access. As a result, the European Commission proposed new rules to streamline and expedite access to e-Evidence. Nearly five years later, these proposed rules were set to be adopted, albeit following significant objections and numerous amendments to the final text.[18]

The urgent need for new, more flexible tools[19]—those allowing judicial authorities in a Member State (MS) to directly issue judicial orders for electronic evidence to

[15] Depauw (no 9) 73.

[16] On 3 April 2014. Until the 2023 e-Evidence package, the European guidance for effective cross-border collection of electronic evidence included: a) cross-border jurisdiction: delineating investigative powers in cyberspace; b) MLA between judicial authorities; and c) direct cooperation with ISPs.

[17] Stanislaw Tosza, 'Gathering Electronic Evidence for Administrative Investigations - Exploring an Under-the-Radar Area' (2023) 2 *EUCRIM*, 216–222, available at https://eucrim.eu/articles/gathering-electronic-evidence-for-administrative-investigations/ (last accessed 23 January 2026).

[18] In the spring of 2018, the European Commission published a final preparatory paper stating that, in cross-border cases, authorities must rely on one of three channels: a) judicial cooperation between public authorities (often too slow); b) direct cooperation between public authorities and ISPs (often cumbersome and not transparent); or c) direct access to e-Evidence (where legal frameworks remain fragmented), doubting, in essence, the EIO. Some days later, it proposed new rules to better equip law enforcers and judicial authorities without abrogating EIO and MLA procedures, at least explicitly.

[19] Gunnar Strömmer (Swedish Minister of Justice) stated that the basic argument in favour of new legal tools was justified on the grounds that new rules would provide judges and prosecutors with quick access to evidence they need, regardless of where it is stored, before it disappears; Consilium Council adopts EU laws on better access to electronic evidence, 27.6.2023, available at https://www.consilium.europa.eu/en/

ISPs in another MS—led to the adoption of the 'e-Evidence package'.[20] Under this package, judicial authorities from any MS can request access to e-Evidence from ISPs offering services in the EU or those established or represented in another MS.

3.1.1. A Lack of Mutual Trust

The relationship between the 'issuing authority', the 'enforcing authority', and the 'targeted/suspected person' has sparked a debate about the effectiveness of evidence searches within this trilateral framework. Efforts to optimise cooperation between the authorities have raised a fundamental question: To what extent should the judicial authority in the ISP's State be actively involved? To answer this, one must inevitably consider the concept of 'mutual trust', as either 'blind faith' or 'reasoned trust',[21] and recognise the need for a more direct relationship,[22] even in light of the existing legal instruments.[23] A need for enhanced direct cooperation was also highlighted by the European Court of Human Rights (ECtHR) in the 2008 case of *K.U. v. Finland*,[24] where Finnish law was found inadequate under Article 8 of the European Convention of Human Rights (the right to respect for private life).[25]

press/press-releases/2023/06/27/council-adopts-eu-laws-on-better-access-to-electronic-evidence/ (last accessed 23 January 2026).

[20] Laviero Buono, 'The genesis of the European Union's new proposed legal instrument(s) on e-evidence. Towards the EU Production and Preservation Orders', *ERA Forum 2019*, 30.

[21] Gianluca Forlani, 'The E-evidence Package - The Happy Ending of a Long Negotiation Saga', (2023) 2 *EUCRIM* 176.

[22] Direct cooperation has become the preferred approach on both a transnational and transatlantic level (as exemplified by the CLOUD Act in the US, which shares similar principles with the e-Evidence package to ensure prompt access to data relevant to criminal investigations). According to Rojszczak, there is a transition from a mediated model (based on MLA) to an unmediated model (based on the Budapest Convention on Cybercrime); see Marcin Rojszczak, 'e-Evidence Cooperation in Criminal Matters from an EU Perspective' (2022) 85(4) *The Modern Law Review*, available at https://doi.org/10.1111/1468-2230.12749 (last accessed 23 January 2026), 1019.

[23] Directive 2014/41/EU on the European Investigation Order in criminal matters (EIO), The European Convention on Mutual Assistance in Criminal Matters between the Member States of the European Union, Regulation (EU) 2018/1727 of 14 November 2018 on the European Union Agency for Criminal Justice Cooperation (Eurojust), Regulation (EU) 2016/794 on Europol, Council Framework Decision 2002/465/JHA on joint investigation teams, The Council of Europe Convention on Cybercrime (the Budapest Convention), and bilateral agreements between the EU and third parties.

[24] *K.U. v. Finland* [2008], Application no. 2872/02, Judgment 2 December 2008, margin no. 46.

[25] The law, as it stood, did not compel the operator of the internet server to provide information about the other user. Instead, it only provided the option to bring criminal charges or a civil action for damages against the server operator for malicious misrepresentation, which the ECtHR deemed unacceptable.

3.1.2. 'Trust'-ing the Process

The delay in launching the project stemmed not only from the initial lack of consensus on mandatory terms and on the type of legal instruments needed to cover e-Evidence provisions[26] but also from a fundamental desire to address the limitations of traditional MLA instruments—such as inter-State requests under the previous EIO—which often proved inefficient because authorities could not access data directly from service providers. Instead, they had to rely on time-consuming, multi-layered channels between States, a structural feature that hindered swift and effective evidence gathering. As a result, the e-Evidence package was designed with a fundamentally different approach, enabling law enforcement authorities to request data relevant to criminal investigations directly from ISPs. This shift toward direct cooperation between States represented a qualitative leap in mutual trust, moving from a general form of cooperation to what can be described as 'absolute mutual trust'.[27]

However, this new system raised various questions and challenges during the trilogue discussions,[28] especially concerning the notification system—the well-known 'notification knot'.[29] The scope and volume of notifications, coupled with concerns about potential infringement of fundamental rights, became the central points of contention between the Council of the EU, the European Commission, and the European Parliament.

[26] According to Sippel's report, the provisions of the Directive should be incorporated into the Regulation in order to avoid the circumvention of the legal basis of the Directive, see Sippel's report (no 11) and Christakis (no 11), 21–23 and fn 95, where author presents how the EC, the Council and the EP have disagreed on several other issues, beginning with 'the title of the legislative package ("electronic evidence" or "electronic information"?) and the legal instruments that it should contain, to important substantive issues such as how human rights will be protected, the role of ISPs, how conflicts of law problems should be addressed, and which categories of data should be covered by the legislative package and their corresponding legal regime'.

[27] Christakis (no 11).

[28] The e-Evidence package comes from the compromise between the Council of the EU, European Commission and European Parliament, the so-called trilogue. They agreed on the following essentials: a) application of the residence criterion to exclude notification to the executing State; b) refusal of orders as a consequence of a pending rule-of-law procedure under Art. 7(1) and (2) of the Treaty on European Union (TEU); c) optional nature of the grounds for refusal and limited role of the ISP in non-execution of orders; d) the distinction between the ISP and the data controller; and e) deletion of data.

[29] The 'copyright' of the phrase belongs to Christakis (no 11) as well as the answers on some of the crucial questions that were imposed. Also, Rojszczak (no 22), 1010–1014.

3.2. Re-shaping the Present

The regulation framework concerning e-Evidence was finally adopted by the Council of the EU and the European Parliament after five years[30] of amended proposals, objections, and negotiations in late July of 2023.[31] It took the form of two new legal instruments: a) Regulation (EU) 2023/1543[32] on European Production Orders (EPOC) and European Preservation Orders (EPOC-PR) for electronic evidence in criminal proceedings and for the execution of custodial sentences following criminal proceedings (the e-Evidence Regulation);[33] and b) Directive (EU) 2023/1544[34] laying down harmonised rules on the designation of establishments and the appointment of legal representatives for the purpose of gathering electronic evidence in criminal proceedings (the e-Evidence Directive).[35] As the new system for gathering electronic evidence, the 'e-Evidence package'[36] aimed to address both the volatile nature of e-Evidence and the challenge of locating stored data.

[30] The agreement on new rules between the Council of the EU and the European Parliament took place in January 2023 while legislative proposals on e-Evidence were presented more than five years prior (European Commission published two proposals on relevant legal instruments on 17 April 2018), with COM (2018) 225 final and COM (2018) 226 final. For a summary of the initial proposal, see Thomas Wahl, 'Commission Proposes Legislative Framework for E-evidence' (2018) 1 *EUCRIM* 35.

[31] The e-Evidence Regulation must be implemented as of 18 August 2026 while EU MS should transpose the e-Evidence Directive into national law by 18 February 2026.

[32] The Regulation lays down the rules under which an MS judicial authority, in criminal proceedings, may issue an EPOC or EPOC-PR and thereby (directly) order an ISP offering services in the EU and established in another MS, or, if not established, represented by a legal representative in another MS, to produce or to preserve electronic evidence regardless of the location of the data.

[33] (no 12).

[34] The Directive lays down the rules on the designation of establishments and the appointment of legal representatives of certain ISPs that offer services in the EU, for the receipt of, compliance with, and enforcement of decisions and orders issued by competent authorities of the MS, for the purposes of gathering electronic evidence in criminal proceedings.

[35] Directive (EU) 2023/1544 of the European Parliament and of the Council of 12 July 2023 laying down harmonised rules on the designation of designated establishments and the appointment of legal representatives for the purpose of gathering electronic evidence in criminal proceedings, *O.J.* L 191, 28.7.2023, 181–190.

[36] The package includes the main legal tools referred to in this paper, as well as the EU Digital Services Act (DSA), which introduced responsibilities and a system of accountability and transparency for providers of intermediary services, and Regulation 2021/784, which regulates the duties of care that hosting ISPs must apply to remove or disable access to terrorist content online. It also encompasses the bilateral negotiations on an EU–US e-Evidence agreement, the initial negotiations on a United Nations legal instrument on cybercrime, and, most importantly, the Council of Europe's Second Additional Protocol to the Cybercrime Convention, which provides tools to strengthen cooperation and dissemination of electronic evidence (like direct cooperation between competent authorities of CoE member states and ISPs of anoth-

3.2.1. How the e-Evidence Package Applies to ISPs[37]

The e-Evidence Regulation introduces an alternative mechanism to the existing international cooperation and MLA tools, addressing the problems stemming from the volatile nature of e-Evidence and the 'loss of location' issue by establishing new procedures for quick, efficient, and effective cross-border access.[38] The e-Evidence Regulation enables authorities to access stored data, no matter where it is located. The EPOC allows a MS judicial authority to request access to e-Evidence directly from an ISP established or represented in another MS immediately or at a later stage. The ISP has 10 days to respond, or 6 hours if it is an emergency. Meanwhile, the EPOC-PR prevents the ISP from deleting or altering e-Evidence while the order is being processed. The e-Evidence Regulation thus lays down: a) the conditions for issuing these instruments; b) the rules governing their execution by ISPs; c) notification of and grounds of refusal for the executing MS; d) the enforcement and penalties (pecuniary sanctions) procedure; e) rights of the persons whose data is sought; and f) a review procedure in case there is conflict of laws with a third state.[39] Penalties are also envisioned for non-complying ISPs: they may face penalties of up to 2% of their total worldwide annual turnover for the preceding financial year. A notification system is also foreseen, particularly for urgent cases.[40]

er State party and effective means of obtaining subscriber information and traffic data); see more in Vagia Polyzoidou, 'Combatting the Cybercrime: Thoughts Based on the Second Additional Protocol (Draft) to the Budapest Convention on Cybercrime' in Tatiana-Eleni Synodinou & al. (eds) *EU Internet Law in the Digital Single Market* (New York: Springer International Publishing, 2021) 355 ff.

[37] Analytically in Adam Juszczak, Elisa Sason, 'The Use of Electronic Evidence in the European Area of Freedom, Security, and Justice - An Introduction to the New EU Package on e-Evidence' (2023) 2 *EUCRIM* 182 ff.

[38] The e-Evidence Regulation creates EPOCs and EPOC-PRs that can be issued by judicial authorities to obtain or preserve e-Evidence regardless of the location of the data. These orders may cover any category of data, including subscriber, traffic, and content data. A threshold has been established for traffic data (except for data requested for the sole purpose of identifying the user) and for content data. These can only be requested for crimes punishable in the issuing country by a maximum custodial sentence of at least three years, or for specific offences relating to cybercrime, child pornography, counterfeiting of non-cash means of payment, or terrorism.

[39] Alexandru Frunza-Nicolescu, 'Electronic Evidence Collection in Cases of the European Public Prosecutor's Office Legal Framework, Procedures, and Specifics' (2023) 2 *EUCRIM* 211 ff.

[40] Unless the issuing authority believes that the offense has been or is likely to be committed within its own territory, or that the individual whose data is being requested resides within its jurisdiction, a notification system will be established for traffic data (excluding data requested solely for user identification purposes) and content data. The purpose of this notification is to inform the enforcing State and provide it with an opportunity to evaluate and, if necessary, invoke any of the refusal grounds specified in the legislation, such as the protection of the requested data. The enforcing State has 10 days, or 96 hours in emer-

An essential instrument in applying the e-Evidence Regulation, the e-Evidence Directive laying down harmonised rules on the designation of establishments and the appointment of legal representatives for the purpose of gathering electronic evidence in criminal proceedings establishes the rules for ISPs' legal representatives—a key step due to the lack of a general legal requirement for non-EU ISPs to be physically present in the EU. The e-Evidence Directive ensures that there is at least one potential addressee for the newly created orders for each ISP falling within the scope of the e-Evidence Regulation, the premise of which is that competent authorities are entitled to issue binding requests to ISPs offering services within the EU regardless of their place of establishment or the physical location of the data. Law enforcement authorities in one MS may now issue orders that are transmitted directly to private actors in a different MS and that have to be executed without the involvement of the authorities of that MS.[41]

Ultimately, the e-Evidence Directive ensures that ISPs across the EU are subject to the same obligations and that any ISP offering services in the EU has designated a legal representative in at least one MS.[42]

3.2.2. The new role of ISPs

The refusal of ISPs to provide data[43] has been one of the primary challenges in the field of e-Evidence;[44] it arises from the jurisdictional differences between legal sys-

gency situations, to raise any grounds for refusal. If it does, the ISP is required to halt the execution of the order and refrain from transferring the data. The issuing authority must then withdraw the order. In emergency situations where data has already been transferred, the issuing authority can either delete or impose restrictions on the data, or comply with the conditions set forth for its use.

[41] Tosza (no 17).

[42] ISPs established outside the EU but offering services within it are obliged to appoint a legal representative for receiving, complying with, and enforcing decisions and orders. See Forlani (no 21) 176.

[43] Either due to jurisdictional conflicts or because the affected individual's nationality complicates the request.

[44] According to SWD (2018) 118 final on the Impact Assessment accompanying the e-Evidence package (17.4.2018), in almost two-thirds (65%) of criminal investigations where e-Evidence was relevant, a cross-border request to service providers was needed while judicial authorities almost universally reported that the length of MLA procedures was a major issue. Geographical conditions and localisation indisputably play a significant role (see *Yahoo! v. Belgium* related to disclosure of technical assistance to get to other than content data), until 'cloud computing' showed that data is less held on a specific device or in closed networks, but rather distributed over multiple providers and locations. Cloud evidence can be related to remote email services (such as Gmail) or to remote data storage and sharing systems (such as Google Docs). Therefore, there is sometimes talk of 'loss of (knowledge of) location' according to Depauw (no 9) 70, fn 49–53; Christakis (no 11) ibid.

tems. The new e-Evidence package, however, allows MS, through their investigating authorities, to directly request electronic evidence from foreign ISPs, including those based in other MS. This represents the first time such direct requests can be made, through EPOCs and EPOC-PRs. As a result, cross-border criminal investigations are expected to become more efficient, as many of the problems associated with accessing digital data across borders will be addressed.

Nevertheless, ISPs now find themselves in a conflicting position, balancing three competing interests: their own business and profit motives, the law enforcement authorities' need for information in criminal investigations, and their customers' privacy and data protection rights. How a private,[45] non-judicial entity[46] can effectively manage these new responsibilities remains a significant and unresolved question, particularly given the extremely complex compliance requirements involved.

3.3. The Alarming Future

The e-Evidence package is an ambitious direct cooperation mechanism that raises a lot of questions related to the legal basis of e-Evidence and the need to ensure the fundamental principles of international cooperation and protection of human rights, all while balancing the facilitation of prosecution by judicial authorities. So, what does the future hold for ISPs and their role in complying with the new rules?

[45] Valsamis Mitsilegas, 'The privatisation of mutual trust in Europe's area of criminal justice: The case of e-Evidence' (2018) 25(3) *Maastricht Journal of European and Comparative Law* 263 ff.

[46] Anna Vilà Cuñat, 'e-Evidence: ¿una evidencia? e-Evidence: an evidence?' (2023) 4 *Revista Sistema Penal Crítico ARTÍCULOS,* available at https://doi.org/10.14201/rspc.31480 (last accessed 23 January 2026), 11.

First,[47] 'mutual cooperation and recognition,[48] as we have known it until now,[49] refers to a relationship between EU MS (an issuing and an executing State) and not one between an issuing State and an individual or a legal entity with financial interests (ISPs). In other words, it is not just a shift in the roles of the MS, but also a challenge to the guarantees provided by the enforcing State's control (and its judicial authorities). Moreover, the potential for financial penalties imposed on ISPs that fail to comply with the e-Evidence Regulation's provisions significantly lowers the likelihood that they would refuse to comply with an order.[50]

[47] This paper does not compare the e-Evidence Regulation with fundamental European principles and rights, crucial issues which will no doubt emerge once the former is implemented in 2026 (e.g., the 'sacrifice' of dual criminality/*ne bis in idem* issues or—even more—problematics related to the rule of law). See more in ELI, 'a Proposal for a Directive of the European Parliament and of the Council on Mutual Admissibility of Evidence and Electronic Evidence in Criminal Proceedings' available at https://www.europeanlawinstitute.eu/fileadmin/user_upload/p_eli/Publications/ELI_Proposal_for_a_Directive_on_Mutual_Admissibility_of_Evidence_and_Electronic_Evidence_in_Criminal_Proceedings_in_the_EU.pdf (last accessed 23 January 2026). See analytically Lorena Bachmaier Winter, 'Mutual Admissibility of Evidence and Electronic Evidence in the EU - A New Try for European Minimum Rules in Criminal Proceedings?' in (2023) 2 *EUCRIM* 223 ff. Not to mention that electronic evidence motivated many 'legal brainstormings' even before it became the e-Evidence package, see also the 2016 'Road Map for a Common European legal framework for electronic evidence' available at http://www.evidenceproject.eu/the-activities/deliverables.html, (last accessed 25 January 2025).

[48] Sarkowicz questions and denounces the adopted method as a violation of the principle of mutual recognition of judicial decisions (as in those cases a judicial authority should be established for all categories of electronic evidence as competent to recognise the order, without any exceptions); but the e-Evidence Regulation requires blind action in European criminal proceedings that 'has no rational justification in the context of trust as several threats resulting from the current regulations to the rights of suspects and accused persons'. See Justyna Sarkowicz, 'The Principle of Mutual Recognition of Judicial Decisions in the Process of Collective Evidence: European Production Order A Step Forward or Step Back?' (2023) 9 *Zbornik Radova Pravnog Fakulteta* 87; Contra, Stanislaw Tosza, 'All evidence is equal, but electronic evidence is more equal than any other: The relationship between the European Investigation Order and the European Production Order' (2020) 11(2) *New Journal of European Criminal Law* 161 ff.

[49] Art. 82 (1) and (2) of the Treaty on the Functioning of the EU (TFEU) establishes the competence of the EU to adopt measures and minimum rules in order to implement this principle, while respecting the possibility for MS to maintain a higher level of protection of fundamental rights; that's how the EU has already adopted Directives and Regulations on procedural rights, conflicts of jurisdiction, or victims' rights etc. The recognised importance of Article 82 TFEU will be relevant as the EPPO Regulation (transfers on the basis of appropriate safeguards in a legally binding instrument) will gain importance when the Second Additional Protocol to the Cybercrime Convention and a modernised Council of Europe Data Protection Convention (Convention 108+) enter into force and are ratified by several third countries. See Rojszczak (no 22) 1014.

[50] For the possibility to refuse only with regard to data relating to traffic and content (Article 8 e-Evidence Regulation) and for very limited reasons provided exhaustively in Article 12 of the e-Evidence Reg-

Second, it is individuals' privacy and personal data[51] that may be infringed upon by the provisions and obligations imposed on ISPs. However, this intrusion seems inevitable when crucial legal interests are at stake, particularly when there is a risk of losing important evidence; and in any case, the debate around privacy as a fundamental human right that must be balanced against other rights is widely known. That said, one might argue that privacy and the legal interests derived from it become even more vulnerable due to its inclusion in (or exclusion from) the e-Evidence package. For example, professional secrecy,[52] confidentiality, and ethics could be compromised when vast amounts of data are accessed without sufficient restrictions or limitations—even when there is a legitimate reason for the investigation. Thus, free access to a vast array of information may not only be excessive and a violation of sensitive personal data but also complicate the investigation itself. There is also the risk of creating 'fishing expeditions';[53] therefore, ISPs should know precisely what information authorities need for their investigations. They can achieve this by demanding a detailed description of the offense, specifying the relevant data, the timeframe, and the number of people affected for every order. They should also only be required to disclose sensitive data, such as medical records subject to prior review and approval by a competent authority.[54] Furthermore, there must be a mechanism allowing the refusal of an order in cases of clear abuse, as confirmed by accredited legal experts.

Moreover, the 'incomplete' notification system in e-Evidence provisions creates many problems regarding the right to a fair trial, which entails the obligation to promptly inform the suspected individual about the process whenever an order is issued, effectively requiring a mandatory notification system. With the EPOC and EPOC-PR, a different approach is applied, leading to the undermining of rights in case of the EPOC-PR, since the accused rights are diminished due to time restrictions and

ulation see also Maria Ludwiczak Glassey, 'Preuves électroniques: état de la situation en Suisse face à l'avancée majeure du droit européen' (2023) 2 *EUCRIM* 207.

[51] Christakis (no 11) 25, Vilà Cuñat (no 46) 8.

[52] EDRi & al., 'Plenary vote on the "e-Evidence package" Regulation and Directive on "European production and preservation orders for electronic evidence in criminal matters" and "legal representatives"', Brussels, 12 June 2023 https://www.cpme.eu/api/documents/adopted/2023/06/Joint-Letter-Plenary-Vote-E-evidence-13-June-2023.pdf (last accessed 23 January 2026), 2.

[53] Both for a) data on various individuals; and b) extensive data on a single individual by collecting unnecessary information. EDRi, 'Demonstrating gaps in the e-Evidence Regulation' (2021) https://www.cpme.eu/api/documents/adopted/2021/9/cpme.2021-116.eEvidence-Scenarios.pdf (last accessed 23 January 2026), 14.

[54] A court or an independent administrative authority.

lack of early access to the case file. This undermines the principle of equality of arms, as the accused lacks the corresponding opportunity to prepare and present their case, leaving the prosecution with an unfair advantage. There is no doubt that the accused's participation in the defence process could offer an alternative legal perspective and, in some cases, prevent errors, minor or significant. Although implementing changes could undermine the e-Evidence package's goals of speeding up procedures and enhancing effectiveness, the principle of equality of arms requires a stricter obligation both on MS and ISPs receiving such requests to process any request as urgently as requests from law enforcement authorities; that's why earlier access to the case file for the data subject should be prioritised.

4. Concluding Remarks

Gathering evidence, particularly e-Evidence, is inherently prone to infringing on human rights such as secrecy and confidentiality; and this risk increases when the responsibility is delegated to third parties like ISPs. The purpose of collecting e-Evidence should not be to prevent future crimes that could create panoptic systems that 'predict' behaviours, which would undermine the possibility of rehabilitation and create a pervasive fear within society. This could ultimately erode the rule of law.

Furthermore, the 'effective remedies'[55] that are included in the e-Evidence Regulation are not as effective as they are portrayed,[56] given that the notification system covers only two types of data, while under the EPOC-PR data can be compulsorily retained by ISPs for a period of 60 to 90 days, and even longer until the EPOC is either issued or revoked.[57]

The challenge lies in ensuring access to the necessary data while simultaneously preventing its abuse or unlawful retention and transfer via an adequate notification system.[58] Involvement from validation authorities, as well as both issuing and enforcing State authorities, could be crucial in this regard. However, if the executing

[55] Article 8 of the e-Evidence Regulation includes auxiliary (to any national ones) remedies for any individual whose data is collected, the suspect, or accused person during the criminal proceedings in which the data was being used, describing inter alia 'the possibility to challenge the legality of the measure, including its necessity and proportionality, without prejudice to the guarantees of fundamental rights in the enforcing State' and obliging the issuing State and any other MS to which the electronic evidence has been transmitted 'to ensure that the rights of the defence and the fairness of the proceedings are respected when assessing evidence obtained through the European Production Order'.

[56] EDRi & al. (no 52) 2.

[57] Article 11 para 1 of the e-Evidence Regulation.

[58] Christakis (no 11), 25 ff.

MS were required to validate each order before data could be transferred, this would increase bureaucracy and render the idea of a 'win-win' scenario unrealistic.

References

Audrito D. & al., 'Analogical lightweight ontology of EU criminal procedural rights in judicial cooperation', Artificial Intelligence and Law (2023) 31 https://doi.org/10.1007/s10506-022-09332-9 (last accessed 25 January 2025), 629–652.

Bachmaier Winter L. 'Mutual Admissibility of Evidence and Electronic Evidence in the EU - A New Try for European Minimum Rules in Criminal Proceedings?' (2023) 3 *EUCRIM* 223–229.

Buono L., 'The genesis of the European Union's new proposed legal instrument(s) on e-evidence. Towards the EU Production and Preservation Orders' *ERA Forum 2019*, 307–312.

Christakis T., 'From Mutual Trust to the Gordian Knot of Notifications The EU e-Evidence Legislative Project' in Franssen V., Tosza S. (eds), *The Cambridge Handbook of Digital Evidence in Criminal Matters* (Cambridge, UK: Cambridge University Press, 2023).

Clough J., *Principles of Cybercrime* (Cambridge, UK: Cambridge University Press, 2015).

Consilium, Europa, Council adopts EU laws on better access to electronic evidence, 27.6.2023, https://www.consilium.europa.eu/en/press/press-releases/2023/06/27/council-adopts-eu-laws-on-better-access-to-electronic-evidence/ (last accessed 25 January 2024).

Depauw S., 'Electronic evidence in criminal matters: how about e-evidence instruments 2.0?' 2018 1 *EuCLR* 63–82, available at https://doi.org/10.5771/2193-5505-2018-1-62 (last accessed 25 January 2025).

EDRi & al., 'Plenary vote - on the "e-evidence package" Regulation and Directive on "European production and preservation orders for electronic evidence in criminal matters" and "legal representatives"', Brussels, 12 June 2023, available at https://www.cpme.eu/api/documents/adopted/2023/06/Joint-Letter-Plenary-Vote-E-evidence-13-June-2023.pdf (last accessed 25 January 2025).

EDRi & al., 'Political compromise on the e-evidence proposal. European media and journalists, civil society groups and technology companies call on decision-makers to improve fundamental rights protections', Brussels 22 November 2022, available at https://www.cpme.eu/api/documents/adopted/2022/11/E-evidence-Letter---Political-Trilogue-22112022.pdf (last accessed 25 January 2025).

EDRi, 'Demonstrating gaps in the e-Evidence Regulation' (2021), available at

https://www.cpme.eu/api/documents/adopted/2021/9/cpme.2021-116.eEvidence-Scenarios.pdf (last accessed 25 January 2025).

Forlani G., 'The E-evidence Package - The Happy Ending of a Long Negotiation Saga', 2023 2 *EUCRIM* 174–181.

Frunza-Nicolescu A., 'Electronic Evidence Collection in Cases of the European Public Prosecutor's Office Legal Framework, Procedures, and Specifics', 2023 2 *EUCRIM* 210– 216

Jones N. & al., 'Electronic Evidence Guide. A Basic Guide for Police Officers, Prosecutors and Judges' (the Council of the EU 2020) available at https://rm.coe.int/c-proc-electronic-evidence-guide-2-1-en-june-2020-web2/16809ed4b4 (last accessed 25 January 2025).

Juszczak Ad., E. Sason, 'The Use of Electronic Evidence in the European Area of Freedom, Security, and Justice - An Introduction to the New EU Package on E-evidence', 2023 2 *EUCRIM* 182–197.

Ludwiczak Glassey M., 'Preuves électroniques : état de la situation en Suisse face à l'avancée majeure du droit européen' 2023 2 *EUCRIM* 204–209.

Mitsilegas V., E. Guild, 'Police and criminal justice co-operation after Brexit' (2023) 30(11) *Journal of European Public Policy* 2519–2539.

Mitsilegas V., 'The privatisation of mutual trust in Europe's area of criminal justice: The case of e-evidence' (2018) 25(3) *Maastricht Journal of European and Comparative Law* 263–265.

Polyzoidou V., 'Combatting the Cybercrime: Thoughts Based on the Second Additional Protocol (Draft) to the Budapest Convention on Cybercrime' in T-E. Synodinou & al. (eds), EU *Internet Law in the Digital Single Market* (New York: Springer International Publishing, 2021) 355–375.

'Road Map for a Common European legal framework for electronic evidence' (2016) available at http://www.evidenceproject.eu/the-activities/deliverables.html, (last accessed 25 January 2025).

Rojszczak M, 'e-Evidence Cooperation in Criminal Matters from an EU Perspective' (2022) 85(4) *The Modern Law Review* 997–1028, available at https://doi.org/10.1111/1468-2230.12749 (last accessed 25 January 2025).

Sarkowicz J., 'The Principle of Mutual Recognition of Judicial Decisions in the Process of Collective Evidence: European Production Order A Step Forward or Step Back?' (2023) 9 *Zbornik Radova Pravnog Fakulteta* 74–92.

Sieber U., 'Computer Crimes, Cyber – Terrorism, Child Pornography and Financial Crimes (General Report)' in D. Spinellis (ed.) *Computer Crimes, Cyber – Terrorism, Child Pornography and Financial Crimes, Reports presented to the Preparatory Colloquium for the Round Table II of the 17th International Congress of Penal Law (Beijing, 2004)* (Antonis Sakkoulas: Athens, 2004) 11–50.

Sippel's Report (2019) available at https://www.europarl.europa.eu/doceo/document/LIBE-PR-642987_EN.pdf (last accessed 25 January 2025).

Speer D., 'Redefining Borders: 'Redefining Borders: The challenges of Cybercrime' (2000) 34 *Crime, Law and Social Change* 259–273.

Strossen N., 'Cybercrimes vs Cyberliberties'(2000) *International Review of Law Computers and Technology* 11–24.

SWD 118 final on the Impact Assessment accompanying the e-evidence package (17.4.2018), available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52018SC0118 (last accessed 23 September 2025).

Tosza S., 'Gathering Electronic Evidence for Administrative Investigations - Exploring an Under-the-Radar Area' (2023) 2 *EUCRIM*, available at https://eucrim.eu/articles/gathering-electronic-evidence-for-administrative-investigations/ (last accessed 25 January 2025), 216–222.

Tosza S., 'All evidence is equal, but electronic evidence is more equal than any other: The relationship between the European Investigation Order and the European Production Order' (2020) 11(2) *New Journal of European Criminal Law* 161–183.

Vilà Cuñat A. 'E-evidence: ¿una evidencia? E-evidence: an evidence?' (2023) 4 *Revista Sistema Penal Crítico ARTÍCULOS*, available at https://doi.org/10.14201/rspc.31480 (last accessed 23 January 2026), 1–14.

Wahl T., 'Commission Proposes Legislative Framework for E-evidence' (2018) 1 *EUCRIM* 35– 36.

Williams M., *Virtually Criminal - Crime, deviance and regulation online* (Oxon: Routledge, 2006).

Yar M., K. Steinmetz, *Cybercrime and Society* (London, Sage: 2019).

Artificial Intelligence, Fundamental Rights, and Digital Constitutionalism from a Comparative Law Perspective: The EU, US, and China Regulatory Models

Dimitrios I. Koukiadis[1]

Abstract

The interplay among Artificial Intelligence (AI), fundamental rights, and digital constitutionalism introduces innovative challenges from a comparative law perspective. Disruptive technologies are reshaping society and challenging traditional legal frameworks designed for analogue communities. The discussion introduces digital constitutionalism as a new doctrinal framework that extends beyond nation-State regulation, incorporating supranational, international, and transnational dimensions. This paper contrasts regulatory models in the EU, US, and China, highlighting the EU's comprehensive, risk-based approach, the US' decentralised, innovation-favouring approach, and China's sector-specific, iterative regulatory measures emphasising content moderation, cybersecurity, and ideological alignment with socialist core values. Despite key differences, all three models share a commitment to transparency, accountability, and protection of fundamental rights. A comparative analysis reveals the need for harmonisation, multi-stakeholder engagement, and adaptable legal frameworks. Ultimately, AI's normative order calls for a balanced approach that promotes innovation while safeguarding human rights in the digital era.

Keywords: digital constitutionalism; nation-State regulation; risk-based approach; innovation; fundamental rights protection

[1] Assistant Professor, School of Law, University of Nicosia; Adjunct Professor, School of Law, Goethe University Frankfurt/Main. AI's large language models have contributed to the formulation of the following text. The final text has been reviewed, verified, and edited by the author.

Prologue

Disruptive technologies[2] are fundamentally and invariably changing the function of 'Society 5.0'.[3] Accordingly, the State, at all levels—national, supranational, international, transnational—is called to manage, govern, and regulate them. The State must re-evaluate the normative principles of lawmaking and the construction of regulatory regimes and establish laws, rules, and policies to navigate the challenges posed by these technologies.

Some of the key challenges for Society 5.0 are: a) management of the digitisation process; b) control of enabling technologies such as big data, artificial intelligence (AI), and bio-nanotechnology; c) reorganisation of digital skills to enhance human potential; d) social acceptance of the multiplicity of new technologies' applications; and e) reform of the 'constitutional' legal framework. While presenting great opportunities for economic growth and development, emerging disruptive technologies also generate threats and risks to citizens' fundamental rights, impartial decision-making, and 'Statal' regulatory and governance capacity.

In an environment of normative uncertainty, the law must find a balance between technological innovation and protection of fundamental rights. After all, that is what

[2] The term has met various definitions and descriptions since its first appearance in 1997 by Harvard Business School Professor Clayton Christensen. Christensen found that disruptive technologies may enter and expand emerging market niches, improving with time and ultimately attacking established products in their traditional markets. Later, disruptive technologies described a process by which a product or service takes root in simple applications at the bottom of the market, and then relentlessly moves upmarket, eventually displacing established competitors. The implication is therefore one of discontinuity. Historical and modern examples of disruptive technologies include the wheel, the steam-engine, electricity, the telephone, the computer, the internet, 3D printing, and artificial intelligence. These technologies have triggered changes that profoundly changed socio-economic systems. For an introduction to disruptive technologies, see WTO, *Introduction: 'Disruptive…' or (just) 'Emerging' technologies?*, 2022, available at https://www.wto.org/english/res_e/booksp_e/wco_wto_introduction.pdf; For further analysis, see Christensen Institute, *Disruptive Innovation Theory: The theory that influenced Steve Jobs, Jeff Bezos, and Andy Grove*, 2024, available at https://www.christenseninstitute.org/theory/disruptive-innovation/; J.M.Utterback/H.J. Acee, Disrupted Technologies: An Expanded View, 2005, 9(1) International Journal of Innovation Management, pp. 1–17; C.Millar/M.Lockett/T.Ladd, *Disruption: Technology, innovation and society*, Special Issue of Technological Forecasting and Social Change an international journal, pp. 254-260, 2017 available at https://www.researchgate.net/publication/321015457_Disruption_Technology_innovation_and_society#

[3] The term was first presented in the 5th Science and Technology Basic Plan, and adopted by the Japanese government in January 2016. See A.Deguchi/C.Hirai/H.Matsuoka/T.Nakano/K.Oshima/M.Tai/S. Tani, What is Society 5.0?, available at https://www.researchgate.net/publication/341746797_What_Is_Society_50.

the law is about—protection of human rights and values, and the rule of law in an age where technology causes wonderment and disruption in equally profound measure.

When it comes to the fast-mutating applications of the AI technology—for example, deciding our access to education and employment; predicting who will require healthcare or social assistance, and our likelihood of future criminality; analysing our emotions; and monitoring the movement of citizens in public spaces and the behaviour of employees in the workplace—AI may consistently undermine the fundamental rights of the most marginalised in society. As it continues to evolve through new applications, such as large language models (LLMs), it is essential to strike a balance between its full effect on socio-techno-economic development and the respect of some key fundamental rights and principles, such as:

1. Equality
2. Transparency and explainability
3. Accountability and responsibility
4. Ethics and fairness
5. Privacy
6. Human rights, diversity, and inclusivity

The regulatory and normative challenges facing the AI environment require the development not so much of general principles and guidelines, but of a holistic portfolio of substantive and procedural tools—beyond the level of nation-State's regulatory sovereignty—that would serve both the growth of digital economy and the protection of citizens' fundamental rights.

1. AI and the Collingridge Dilemma

Among the various definitions given for AI, one of the most widely accepted, which has also been adopted by the EU Artificial Intelligence Act (AIA), is:

> a machine-based system 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. Different AI systems vary in their levels of autonomy and adaptiveness after deployment.[4]

[4] This definition was first released by the OECD. Early drafts of the AIA had chosen another definition of AI. After the various negotiation rounds between the participating parties, the AIA ultimately adopted the OECD definition. See Article 3 of the Regulation (EU) 2024/1689 of the European Parliament and of the

The key regulatory challenge of AI is how to best evaluate and manage technological progress and its social control, and the endemic 'pacing' problem between new technologies and the law. Technological innovation has long surpassed the ability of law to keep up, an issue that runs through not only the ordinary laws passed by national or supra-national parliaments, but also the constitutional level of lawmaking. At the heart of the AI regulatory challenges lies the viability and capacity of constitutional law itself to deal with the normative issues posed by the AI applications.

The above pacing issue, already addressed through the Collingridge dilemma,[5] runs through the broader regulatory debate on disruptive technologies. Suffice to say that in the three earlier industrial revolutions, the question of whether the regulatory, normative, societal needs could keep in step with the occasional technological innovations was an omnipresent one.

Indeed, in all the three earlier industrial revolutions, technology has always played a leading role in changing the power dynamics within a society between technology innovation on the one hand, and law and social relations on the other. At the dawn of the internet, about 30 years ago, a key argument was that, depending on the level of control and oversight, the internet would either reinforce existing social structures and traditional legal doctrines, or turn them upside down. The internet era was perhaps the first to see the appearance of 'constitutional' law apparatuses in the normative discourse of technology regulation.[6] Prior to this, the normative debate had been

Council of 13 June 2024 laying down harmonised rules on AI and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139, and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797, and (EU) 2020/1828 (AIA), available at http://data.europa.eu/eli/reg/2024/1689/oj; For the OECD approach to AI, see OECD, Artificial Intelligence & Responsible Business conduct, 2018, available at https://mneguidelines.oecd.org/RBC-and-artificial-intelligence.pdf.

[5] One of the most influential writers on the pacing issue between technology and society has been David Collingridge, who argued that technology plays a crucial role in shaping social relations and power dynamics. Technology can reinforce existing social structures or turn them upside down depending on its level of control and oversight. See D. Collingridge, *The Social Control of Technology,* Frances Pinter, 1980. For a critical and more updated approach to Collingridge's argument, see A.Genus/A.Stirtling, *Collingridge and the dilemma of control: Towards responsible and accountable innovation,* 47(1) Research Policy, 2018, pp. 61-69, available at https://doi.org/10.1016/j.respol.2017.09.012

[6] From the early years of the global introduction of internet—in the early 90s—the primary doctrinal arguments and vocabulary at the heart of the legal discourse were of constitutional law origin., e.g. a 'new digital Constitution', a 'Declaration of Independence' for the cyberspace community, 'new Constitutional rights', a 'Constitution for the internet', 'new digital fundamental rights', or an 'internet Bill of Rights'. Perhaps, because it was evident from the start of the internet that the concept of 'fundamental rights' would be at the epicentre of internet's normative order, the relevant global normative discussion has focused on 'constitutional law' terminology. On this issue, see B.-J. Koops/R. Leenes/P. De Hert (eds.), *Constitutional Rights and New*

on the level of ordinary legislative process, not touching the realm of constitutional law.

A series of ongoing transformations in contemporary society are now challenging existing constitutional law apparatuses. The changes prompted by the digital revolution in ourselves, in our relationships with other individuals and, ultimately, in society at large, are fermenting under a vault of constitutional norms that have been shaped for 'analogue' communities. The 'Silicon giants'[7] have now gained more control than ever over the lives of global citizens. By managing AI and its various applications and services, these multinational tech companies de facto influence how and to what extent people enjoy their fundamental rights.

A multiplicity of normative counteractions is emerging to address the constitutional challenges of the digital revolution. These counteractions, serving a legal pluralistic concept of constitutional lawmaking, attempt to reaffirm our core fundamental rights in the digital context and rebalance the new asymmetries of power created by AI's rapid and profound development.

In order to confront these power asymmetries created by AI, and prior to designing a normative and regulatory framework for it, some crucial questions should be asked on the nature of forthcoming AI regulation:[8] What are the preferred types of regulation? At what level should regulation exist or be defined? Can AI regulation enable both digital innovation and the protection of fundamental rights? How does the prevalence of a technology affect how it is regulated? And finally, how should public trust be maintained in AI regulatory efforts?

The normative and regulatory question is therefore not only what the content of AI regulation should be, but also who should create it. The existing structure of multilevel constitutionalism, i.e. nation-State constitutionalism, supra-national con-

Technologies: A Comparative Study Covering Belgium, Canada, France, Germany, Sweden, and the United States, Tilburg University- Institute for Law, Technology, and Society (2007), available at www.uvt.nl/tilt; S. Rodotà, *A Constitution for the Internet* (2006), available at G. Teubner, Globale Zivilverfassungen: Alternativen zur staatszentrierten Verfassungstheorie, in 63 Zeitschrift für ausländisches öffentliches Recht und Völlkerrecht, pp. 1-28 (2003); G. Teubner, Verfassungsfragmente: Gesellschaftlicher Konstitutionalisierung in der Globalisierung, Suhrkamp, Berlin (2012); J. Boyle, Foucault in Cyberspace: Surveillance, Sovereignty, and Hardwired Censors, in 66 University of Cincinnati Law Review, pp. 177-205 (1998); G.Ziccardi, *Resistance, Liberation Technology and Human Rights in the Digital Age*, pp. 152-159, Springer (2013).

[7] For the most recent version of the Silicon giants, see E.Tarver, *The Biggest Companies in Silicon Valley*, November 23, 2024, available at https://www.investopedia.com/articles/markets/103015/biggest-companies-silicon-valley.asp.

[8] Regulatory Horizons Council, *The Future of Technological Innovations and the role of Regulation,* pp. 6-7, 2021

stitutionalism (such as EU constitutionalism), and international constitutionalism (such as the Council of Europe, the United Nations, or World Trade Organization regulatory initiatives),[9] is now supplemented by a fourth level of constitutionalism, i.e. transnational constitutionalism. This kind of constitutionalism, which is based on the concept of 'multi-stakeholderism', aspires to involve States and global society, along with the big tech companies, as leading players in the production of norms and laws for the AI ecosystem.

2. Digital Constitutionalism for the AI Ecosystem

A logical question raised within the AI regulation discourse is why a suitable regulatory and legal framework for AI could not be an ordinary secondary law/regulation/statute, but, instead, must follow the semiology and semantics of a 'constitutional law'. A first response would be that, during the last three decades of the global debate on the normative and regulatory framework of the internet, the relevant argumentation has had a very strong 'constitutional law' connotation and content. The initial normative point of reference at the centre of this discourse has not been the tradition of civil law, commercial law, or intellectual property law, but the constitutional law dimension of forthcoming law. Even the language used to describe the new normative framework is based on the constitutional law discipline.[10]

Secondly, the extent to which the information society and digital innovation have exploded in the last decade has resulted in a profound influence on citizens' core fundamental rights. Contrary to the digital reality 30 years ago, today, the continuous and speedy development of disruptive technologies—especially AI—coincides with an ongoing risk of violation of fundamental rights, such as freedom of expression, personality rights, privacy, free will, and personal data, and a simultaneous weakening of public interest and the democratic principle.

[9] For the Council of Europe's regulatory approach to AI, see Council of Europe, *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, September 5, 2024, available at https://rm.coe.int/1680afae3c. This is the first-ever legally binding international treaty in this field. For the UN AI governance model, see United Nations, *The Global Digital Compact*, September 2024, available at ttps://www.un.org/techenvoy/global-digital-compact. For the WTO AI regulatory standpoint, see WTO, *The Policies of AI and Trade*, November 2024, available at https://www.wto.org/english/res_e/booksp_e/trading_intelligence_ch3_e.pdf.

[10] See supra note 5 and, also, J.P. Barlow, *A Declaration of Independence of Cyberspace,* Electronic Frontier Foundation, February 8, 1996, available at https://www.eff.org/cyberspace-independence.

The 'constitutionalisation' of the discourse on an adequate and effective normative framework for AI—including calls for a constitutional law level of regulation[11]—is thus a response from stakeholders in the AI ecosystem against the proliferation of AI applications: LLMs, generative AI, machine learning (ML), general purpose AI, deep learning AI, AI neural networks, predictive AI, etc. The high status and historical role of a 'constitution' provides us with a sense of a stronger protection of our fundamental rights against the accumulated power and influence of the contemporary technology companies. As Edoardo Celeste puts it:

> Digital constitutionalism represents the conceptual lymph of the current constitutional moment. It normatively advocates a reconfiguration of the constitutional framework. Analogue norms are no longer able to address the full range of complexities of the virtual environment. A series of normative counteractions are emerging to implement the principles of a constitutionalism rethought for the digital age [...] Constitutionalism is translated in a language that speaks to the actors of the virtual environment. In this way, old principles become more easily applicable in new societal contexts. Further corollaries, and even novel norms emerge to express foundational constitutional values in the digital society.[12]

So far, however, the concept of digital constitutionalism remains more a theoretical and normative construction than a solid doctrine with overall legal recognition and practical application within national, supranational, or international legal orders. Save for the supranational legislation of the EU on the digital economy through its digital legislative 'package'[13] which, despite its global impact and effect, typically binds only the EU Member States, worldwide, there is no regulation/legislation transcending national borders, backed by a strong adjudication system and enforcement mechanisms.

---

[11] See A. Bradford, *Europe's Digital Constitution,* VerfBlog, 2023/9/06, available at https://verfassungsblog.de/europes-digital-constitution; B.J. Koops/R. Leenes/P. De Hert (eds.), *Constitutional Rights and New Technologies: A Comparative Study Covering Belgium, Canada, France, Germany, Sweden, and the United States*, Tilburg University- Institute for Law, Technology, and Society (2007), available at www.uvt.nl/tilt.

[12] E.Celeste, *Digital Constitutionalism. The Role of Internet Bills of Rights*, pp. 84,85, Routledge (2023). Celeste argues that including 'digital' before the term 'constitutionalism' refers to a strand of constitutional theory that seeks to articulate principles for digital society. What is common among the terms 'digital', 'global', and 'international' is the transcendence of State constitutionalism.

[13] The EU digital legal 'package' includes the General Data Protection Regulation (GDPR), the Digital Services Act (DSA), the Digital Markets Act (DMA), the Data Act, the Data Governance Act (DGA), the Artificial Intelligence Act (AIA), and the forthcoming e-privacy regulation.

2.1 Multilevel Constitutionalism

The above argumentation is based on the fact that a constitution: a) is a source of law enjoying supremacy over other sources of law; b) constitutes a superior legal order; and c) organises legal systems producing secondary laws. As mentioned, modern constitutions are defined by the concept of multilevel constitutionalism, which encompasses national, supranational, international, and transnational constitutionalism.

Traditional constitutionalism, as in the Westphalian notion of the nation-State, has been relativised in parallel with the relativisation of the sovereignty of the traditional nation-State, as well as of the post-Westphalian Statal authority.[14] This relativised constitution can, in various ways, incorporate the problems created by the internationalisation of the constitution, by the constitutionalisation of international law, by the international judicial review of national constitutions, and by the international judicial review of the decisions of national, supreme, and constitutional courts. Then, it can assimilate them, reverse them, and make them elements of its own expanded normativity.

In this context, like the nation-State, international organisations and hybrid supranational entities such as the EU cannot follow, much less control and subordinate to their legal order, AI development, which is linked to elements that are not spatial and have no local/national characteristics. We are thus experiencing new forms of internationalisation and privatisation of very critical sectors related to disruptive technologies and AI.

These sectors are outside the control of the State, and also outside the control of the international legal order and, to a very large extent, of the EU legal order, despite its efforts to produce directives, regulations, and international conventions, or to interpret a classic text such as the European Convention on Human Rights, which evolves through the case law of the European Court of Human Rights, and tries to capture such phenomena as the internet, which is not controlled and legally organised like conventional space.[15] Besides, the concept of sovereignty over the internet is highly contested,[16] fluid, and undefined; this, of course, affects both traditional constitutionalism as well as multilevel constitutionalism.

---

[14] For an accurate analysis of both the Westphalian and the post-Westphalian international order, see C.Vergerio, *Beyond the Nation-State,* 2021, available at https://www.bostonreview.net/articles/beyond-the-nation-state/

[15] N.Tsagourias, 'The legal status of cyberspace', in N. Tsagourias/R.Buchan (eds), *Research Handbook on International Law and Cyberspace*, pp 13-29, Edward Elgan Publishing (2015).

[16] Id., at p. 16 ff.

2.2. The Value of the Comparative Law Perspective

In such a global and pluralistic normative environment, the comparative law perspective is of the highest importance, considering the need for harmonisation of law, i.e. making the legal rules of two or more legal systems more alike, and, occasionally, for the unification of law (the introduction of identical legal rules into two or more legal systems).[17]

In the field of disruptive technologies, and especially AI regulation, where the national, supranational, and international levels of lawmaking interconnect, and where the question arises of whether and to what extent a harmonisation or unification of law can be achieved, comparative law can help us choose between the lowest common denominator, a compromise solution with a synthesis of legal rules, or a general solution that comprises existing normative models.[18] This solution could function as the pillar for national legislative reform, an international treaty, a supranational legislative act, such as EU directives and regulations, or a regulatory model based on 'soft law' and the notion of self-regulation.[19] Alternatively, a comparative law analysis, which is based not only on the study of various legal orders and legal systems but also of their legal traditions and legal cultures, may lead us to reject the harmonisation or unification of laws due to irreconcilable differences.

Additionally, given the dominant position of the tech giants in the development of AI programmes and, therefore, in the dictating of some AI regulatory and governance rules, comparative law can bring the global AI business community closer to the legal community at all levels, i.e. national, supranational, international, and develop similar regulation and governance responses that could facilitate digital innovation and meet the needs of the digital economy, as well as the need for fundamental rights protection. Thus, the alternative of a transnational regulation and governance of AI, i.e. a regulation and governance based both on State and private actors, respecting the rights of AI's multi-stakeholder community, could also be a normative option, demonstrating that there could be more regulatory options to consider besides the traditional harmonisation or unification options.

Regardless of whether a comparative law analysis leads us to a harmonisation or unification of law, a convergence of legal systems, or a transnational regulatory alter-

[17] See the excellent analysis on the need for the comparative law perspective in M. Bogdan, *Concise Introduction to Comparative Law,* p. 18, Europa Law Publishing (2013).

[18] See M.Siems, *Comparative Law*, p. 4, Cambridge University Press (2014).

[19] Id.

native, it should also lead us to a normative outcome with practical advantages—legal and non-legal. The AI economy has become so vast and influential on a daily and practical level[20] that for a normative solution to be viable it must bring substantive practical advantages to all the stakeholders involved, eschewing normative solutions of a purely declaratory nature that lack solid governance infrastructure.

Globally, as this paper is written, besides the EU legislative initiatives, and the recent Council of Europe AI treaty, approximately 31 national laws on AI have been introduced and 13 are currently in draft status. For reasons of space, this paper will mainly focus on the EU Artificial Intelligence Act, the key US regulatory initiatives, and the most important Chinese regulatory measures on AI.

3. The EU Artificial Intelligence Act (AIA)

Following the regulatory footsteps of the General Data Protection Regulation (GDPR), and aiming to produce the world's first holistic regulation of AI technology, the EU chose two long-term, strategic objectives: to make the EU a world-class hub for AI regulation, and to enable citizens and businesses to enjoy the benefits of a rule-abiding AI while feeling safe and secure.

To achieve these objectives, the European Commission in April 2021 proposed the Artificial Intelligence Act (AIA), the most ambitious and holistic legislative initiative of the EU's AI regulatory agenda. After many rounds of negotiations and amendments within the EU legislative process, the European Parliament voted on and adopted the AIA on 13 March 2024, and it came into force in August 2024.[21]

[20] For the impact of AI on society in general, see P. Salokannel, *The Impact of AI: How Artificial Intelligence is Transforming Society*, March 7, 2024, available at https://www.3dbear.io/blog/the-impact-of-ai-how-artificial-intelligence-is-transforming-society; For AI's impact on economy, see K. Georgieva, *AI Will Transform the Global Economy,* January 14, 2024, available at https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity; Tony Blair Institute for Global Change, The Impact of AI on the Labour Market, November 2924, available at https://assets.ctfassets.net/75ila1cntaeh/5yiS3TBdF96J4Kt0tFmj3k/c4b260667ed70fc791020268cfed443b/1mABBNuF6Anp7zTv1ZeIxe--155801112024.

[21] See Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on AI and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance), available at https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689.

The AIA entered into force on 1 August 2024, but it will only be fully applicable on 2 August 2026, with some exceptions, such as the prohibitions and AI literacy obligations, which entered into force on 2 February 2025.

The AIA aims to address the risks arising from any potential application of AI, to protect the fundamental rights affected by the use of AI, to build an ecosystem of excellence in the field of AI by strengthening research, digital innovation, and industrial development,[22] and to regulate, through a set of proportionate and flexible rules, the large-scale economic and social benefits of AI across entrepreneurial and social activities.

The European Commission itself has also stressed that the AIA aims to enable the EU to play a global regulatory role at least for the coming decade.[23] It is this holistic nature of the AIA that differentiates the EU initiative from other AI national and international regulatory attempts,[24] and reveals the EU's willingness to play a leading role not only in the regulation but also in the governance of AI.[25]

3.1 The AIA's 'Risk-based' Approach

The AIA seeks to ensure a high level of protection for the fundamental rights enshrined in the EU treaties and the EU Charter of Fundamental Rights through addressing the different levels of risk posed by AI applications. The AIA has a regulatory approach based on four different levels of risk: minimal, limited, high, and unacceptable. It establishes a reliable risk methodology, especially with regards to AI applications that pose the most common risks to the health, safety, and fundamental rights of individuals. [26]

[22] Id., recital 1.

[23] European Commission, *Shaping Europe's digital future-Policy and Legislation*, 2 June 2024, available at https://digital-strategy.ec.europa.eu.

[24] So far, worldwide, approximately 31 national laws have been put into force regulating AI technology. These include the US regulatory initiatives, including President Joe Biden's Executive Order and the AI Bill of Rights; China, at the level of AI's basic ethical principles; the UK at the level of guiding ethical standards; and Brazil, whose legislation is deeply influenced by the EU AIA, and Japan. All the above initiatives, differing only in scope of influence, have opted for sectoral, as opposed to holistic, AI regulation. They all follow, *mutandis mutandis*, as was the case with the GDPR, the logic and philosophy of relevant EU legislative initiatives.

[25] The term 'governance' is broader than both 'regulation' and 'legislation'. It encompasses the latter two concepts, but is also about enacting specific policies, and making key decisions about how a technology operates overall in a specific territory. Key components of the term 'governance' include the content of a legislation, how it is produced, the decision-makers of a regulatory framework, the legislation's geographical limits, the penalties and their enforcement, the coercive nature of regulation, and the administrative oversight structure of compliance with the legislation.

[26] See Article 1 of the AIA, available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689.

To explain the logic of the AIA, a 'risk-based' approach strives to achieve a balance between the benefits of innovation and protection of individuals' rights; it also aims to define what risk is acceptable and to what extent. The AIA adopts an *ex ante* approach that helps prevent harmful applications from going to market, rather than assessing *ex post* the behaviour of manufacturers and deployers in complying with general principles.[27] The AIA states:

> In order to introduce a proportionate and effective set of binding rules for AI systems, a clearly defined risk-based approach should be followed. That approach should tailor the type and content of such rules to the intensity and scope of the risks that AI systems can generate. It is therefore necessary to prohibit certain unacceptable AI practices, to lay down requirements for high-risk AI systems and obligations for the relevant operators, and to lay down transparency obligations for certain AI systems.[28]

As the AIA risk pyramid demonstrates (see Figure 1 below), the AIA's risk classification is as follows: At the top of AI's risk pyramid lies unacceptable risk, for applications which are prohibited, e.g. social scoring systems and manipulative AI. Below the unacceptable risk, we find the high-risk AI applications, such as those impacting biometrics, education, employment, access to essential services, law enforcement, immigration, and administration of justice and democratic processes.[29] The next level contains limited risk AI systems, which are subject to lighter transparency obligations. In this category, developers and deployers must ensure that end-users are aware that they are interacting with AI, e.g. chatbots and deepfakes. Finally, at the bottom of the pyramid lie the minimal risk AI applications, which are unregulated.

---

[27] See the relevant analysis of A. Mantelero, *The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template*, 54 Computer Law & Security Review: The International Journal of Technology Law and Practice, pp. 1-31, at 3 (2024), available at https://sciencedirect.com/sciene/article/pii/S0267364924000864. Mantelero also argues that the AIA, viewed from the technology innovation perspective, is limited in its protection of fundamental rights because is primary goal is a product safety regulation. See id., at 4.

[28] AIA, recital 26.

[29] For the AI risk pyramid Figure, see L. Witzel, *5 Things You Must Know Now About the Coming EU AI Regulation*, 2021, available at https://medium.com/@loriaustex/5-things-you-must-know-now-about-the-coming-eu-ai-regulation-d2f8b4b2a4a9. For high-risk AI systems, see the concise but very illustrative analysis in A.L. Kempf/N. Rauer, *A guide to high-risk AI systems under the EU AI Act,* 13 February 2024, available at https://www.pinsentmasons.com/out-law/guides/guide-to-high-risk-ai-systems-under-the-eu-ai-act.

These include the majority of AI applications currently available on the EU single market, such as AI-enabled video games and spam filters.

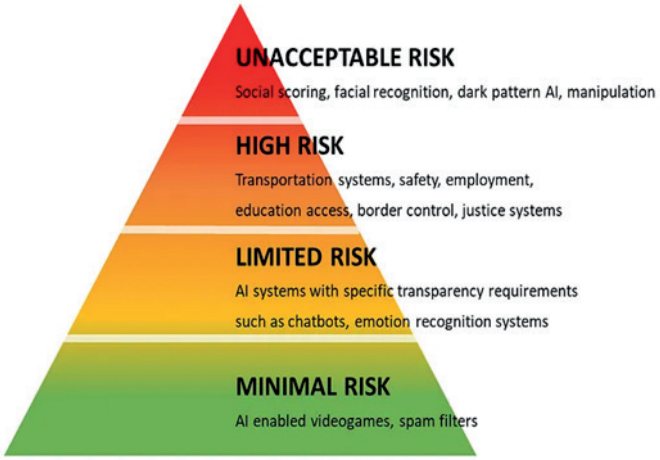

Figure 1: the AIA's risk classification

Within its risk-based approach, the AIA seeks to strengthen the protection of rights already covered by the EU Charter of Fundamental Rights, such as the right to human dignity (Article 1), respect for privacy and protection of personal data (Articles 7 and 8), non-discrimination (Article 21) and equality between women and men (Article 23), freedom of expression (Article 11) and freedom of assembly (Article 12), fair trial and defence (Articles 47 and 48), and the general principle of good administration.[30]

The AIA's risk-based approach,[31] while commendable in its goals and its balancing approach, has also received normative criticism. The AIA's risk-based framework subordinates fundamental rights to a balancing exercise that offsets potential infringements against anticipated economic or technological benefits.[32] By permitting regulators to assign variable weights to EU fundamental rights, and to excuse their

[30] AIA, recital 48, available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689.

[31] For an overall critical analysis of the content of AI Act and specifically its risk-based approach, see M. Veale/F.Z. Borgesius, *'Demystifying the Draft EU Artificial Intelligence Act. Analysing the good, the bad, and the unclear elements of the proposed approach'*, 4 Computer Law Review International, pp. 97-112, at 99 ff. (2021)

[32] A. Mantelero, supra note 26, at 3.

limitation whenever an innovation dividend is invoked, the risk-based framework may normalise unnecessary or disproportionate restrictions of the aforementioned rights. In so doing, it erodes the absolute character of certain rights, generates inconsistent levels of protection, and ultimately undermines the EU's commitment to a rights-first legal order. In brief, risks can be weighed against the potential benefits of innovation, leading to an uncertain balancing outcome in terms of the acceptance of risks.[33]

Furthermore, the legal vagueness of the term 'high-risk systems' as defined in the AIA and their categorisation rely on a speculative *ex ante* assessment of a potential harm without a consistent methodological framework for risk quantification or qualification. The criteria 'significant risk' to health or 'adverse impact' on rights are not very clearly explained and are thus susceptible to various interpretations by national supervisory authorities. Certain uses of AI are categorically labelled as high-risk regardless of contextual nuances, which engenders both over- and under-inclusiveness. This vagueness may also have a significant impact on the sense of legal certainty that the AIA wishes to establish, and may also challenge the principle of proportionality, which demands that legislative measures be necessary and proportionate to their objective. A risk-based categorisation of AI applications that fails to clearly delimit the scope of its application risks violating this principle.

The AIA also allocates considerable discretion to national authorities to interpret and apply the risk taxonomy, thereby raising the fear of regulatory fragmentation across EU Member States. The harmonising function of EU law may thus be compromised, especially in areas where national conceptions of risk tolerance or rights sensitivity diverge. This is particularly concerning given the open-ended nature of Article 7 of the Act, which allows for the expansion of Annex III, and the potential for divergent implementations under the subsidiarity principle.

The above points of criticism notwithstanding, in the AIA Act, the EU legislator opted for an 'acceptable' risk approach, restricting "high risk" AI technology to specific AI applications. This pro-innovation view in framing the risk-based logic is consistent with a first-generation law such as the AIA. As at the beginning of the industrial revolution and, later, at the dawn of the information society era, the initial regulatory approach balances risk mitigation with its cost in terms of a chilling effect on investment in innovation.

[33] Id.

3.2 The AIA's Fundamental Rights Impact Assessment (FRIA)

Serving the spirit of the risk-based model, and following, to a certain extent, the logic of the earlier Human Rights Impact Assessment (HRIA)[34] and the Data Protection Impact Assessment (DPIA),[35] the AIA's Article 27 establishes the Fundamental Rights Impact Assessment (FRIA). FRIA is an assessment that deployers of high-risk AI systems (as per Article 6(2) of the AIA) must perform to assess the impact of high-risk AI systems on fundamental rights.

Like its counterpart GDPR's DPIA, FRIA features an *ex ante* approach, a rights-based focus on risk assessment, a circular iterative structure that follows the product/service throughout its lifecycle, and an expert-based nature.[36] The main difference between DPIA and FRIA is the fact that the former focuses only on personal data protection, while the latter considers all potentially affected rights and freedoms.

While it represents a laudable attempt at ensuring regulatory accountability and rights integration, FRIA's predominantly technocratic methodology may risk reducing substantive normative considerations to procedural formalities. Such a reliance on *ex ante* risk quantification and categorisation, without a sufficiently dynamic mechanism to capture evolving harms, invites critiques that its current formulation could inadvertently undermine the very fundamental rights it seeks to protect.[37]

Undoubtedly, given the already global impact of the GDPR, the AIA is also expected to have far-reaching regulatory implications for AI applications not only in the EU, but also worldwide. For example, high-risk systems, as defined by the AIA, necessitate compliance with stringent pre-market obligations, ranging from risk management and risk impact assessment to various control measures and transparency requirements. Moreover, providers of generative AI models, based on risk assessment, will have to meet additional obligations.

Even more than the GDPR paradigm, the AIA is closer to the concept of the risk society. For the new digital society, which is already being hatched, and in which all

[34] The Human Rights Impact Assessment is a process that assesses the adverse impacts of projects and activities on the enjoyment of human rights in relation to potentially affected rights-holders. See United Nations - Human Rights Council, *Guiding Principles on Business and Human Rights: Implementing the United Nations 'Protect, Respect and Remedy' Framework'*, Resolution 17/4 of 16 June 2011.

[35] The Data Protection Impact Assessment was established by the GDPR, in Article 35, and itself also follows the doctrinal paradigm of the Privacy Impact Assessment (PIA), which has been used globally since the 90s. For a further analysis on the logic and function of PIAs, see D. Wright/P. De Hert (eds), *Privacy Impact Assessment*, Springer 2012.

[36] See A. Mantelero, supra note 26, at 5.

[37] See Veale/Borgesius, supra note 30.

the applications of AI to date have already created a 'postmodern asymmetric threat society',[38] EU AI legislation is, so far, the most comprehensive, credible, institutionally strong, and human-centred response.

Like the GDPR before it, the AIA is another expression of a modern, flexible supranational digital constitutionalism, which balances the promotion of the full digital transformation of human society and the protection of fundamental human rights and democratic values. Against the tendency to privatise the AI normative order, mainly through the self-regulatory choices of the big tech companies, and the deconstruction of the nation-State as a reliable and effective producer of AI's digital normativity, the EU legislative package for the digital economy is, perhaps, the only holistic and enforceable AI law we have within the current global normative order.

4. The US' Normative Approach to AI and Fundamental Rights

In the US legal order, Congress and the White House alike have been getting to grips with AI technology and its impact on their economy and their citizens' fundamental rights. US policymakers have struggled to keep pace with the development of AI, with Congress often trying to determine which AI field to legislate. Nevertheless, either because of the EU regulatory initiatives, or of the global calls for a sound, comprehensive, effective AI regulatory and governance framework, the US is navigating a rapidly evolving AI regulatory and policy landscape through legislative efforts outlined in federal strategic plans and regulatory developments. At the same time, dozens of

[38] Introduced by sociologists Urlich Beck and Anthony Giddens, the 'risk society' is a theory that concerns the transition from modern industrial society to a new era distinguished by technological dangers. The risk society is defined not just by the distribution of goods, but by the distribution of 'bads', such as informational asymmetries, technological hazards, pollution, contamination, and other by-products of production. For an overall analysis of the concept of risk society, see the classic work of Urlich Beck, *Risk Society: Towards a Newest Modernity,* 1992, available at http://www.riversimulator.org/Resources/Anthropology/RiskSociety/RiskSocietyTowardsAnewModernity1992Beck.pdf; For AI asymmetries, see J. Nosta, *The Precarious Asymmetries of Human-AI Relationships. Bridging the communication chasm between humans and AI,* 1 April 2024, available at https://www.psychologytoday.com/us/blog/the-digital-self/202404/the-precarious-asymmetries-of-human-ai-relationships; R.Hacker, *The Asymmetry of Knowledge: How AI Unveils the Invisible Realms of Science,* 7 June 2024, available at https://roberthhacker.medium.com/the-asymmetry-of-knowledge-how-ai-unveils-the-invisible-realms-of-science-ca0499847514; L. Eliot, *AI Ethics Wary About Worsening Of AI Asymmetry Amid Humans Getting The Short End Of The Stick,* 19 August 2022, available at https://www.forbes.com/sites/lanceeliot/2022/08/19/ai-ethics-wary-about-worsening-of-ai-asymmetry-amid-humans-getting-the-short-end-of-the-stick/; T. Marwala/E. Hurwitz, Artificial Intelligence and Asymmetric Information Theory, October 2015, available at https://www.researchgate.net/publication/282906709_Artificial_Intelligence_and_Asymmetric_Information_Theory.

(Silicon Valley) tech companies are working to influence the eventual normative order of AI within the US, and the eventual policies that will govern the use of AI.

Till the end of 2024, over 200 AI laws have been introduced, and States like Colorado, California, Connecticut, Texas, and New York have implemented their own AI laws.[39] The interplay between State and federal measures poses challenges for uniform compliance, with organisations needing to navigate varying requirements.

Compared to the EU paradigm, US legislators (federal and State) have been less keen on AI regulation, with the thinking that any kind or regulation of such a rapidly developing disruptive technology would hinder its positive effects on the digital economy and innovation. The US approach likens the current stage of AI to the early days of cyberspace, when the key argument was that no Statal authority should be in charge, but the participating stakeholders should enjoy a regulatory autonomy, and, thus, be in charge of producing the norms and rules of the game. US lawmakers are wary of repeating the mistakes made with attempts to regulate social media, which proved futile.[40]

Federally, significant developments include the US President's Executive Order, the US AI Bill of Rights, and the National Artificial Intelligence Research and Development Strategic Plan (the NAIRDSP). These are some key US AI regulatory and governance initiatives that focus on innovation and governance.

4.1. The US President's Executive Order on AI

In October 2023, the White House issued an Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (the EO).[41] The EO states

[39] For the progress and the pacing of the individual US States passing AI laws, see Comply Advantage, *AI Regulation in the US — A Guide for Compliance Officers*, 2023, available at www.complyadvantage.com.

[40] See E.Chemerinski, *States' Misguided Attempts to Regulate Social Media*, March 1, 2024, available at https://www.governing.com/policy/states-misguided-attempts-to-regulate-social-media.

[41] See *Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence*, October 30, 2023, available at https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30 /.

However, President Trump, on 23 January 2025, issued another executive order titled *Removing Barriers to American Leadership in Artificial Intelligence* (Trump AI Executive Order). The Trump AI Executive Order revoked the Biden administration's AI Executive Order. Nevertheless, the Trump AI Executive Order is not finalised as a text, is much shorter (only two pages long), has an action plan of 180 days, and does not repeal the whole content of the Biden AI Executive Order. On the contrary, many of the policy directives included in the Biden AI Executive Order, such as promoting competition and innovation in the AI sector and maintaining the US' role as a global leader in AI, have been kept in the new Trump AI Executive Order. What will change with the final new AI executive order remains to be seen. For the text of the

that responsible AI use has the potential to solve urgent challenges while making our world more prosperous, productive, innovative, and secure. Conversely, irresponsible use could exacerbate societal harms such as fraud, discrimination, bias, and disinformation.[42] The EO establishes a government-wide effort to guide responsible AI development and deployment through federal agency leadership, regulation of industry, and engagement with international partners. There are eight AI guiding principles:[43]

1. Safety and security
2. Innovation and competition
3. Worker support
4. Consideration of AI bias and civil rights
5. Consumer protection
6. Privacy
7. Federal use of AI
8. International leadership

The EO recognises that AI models may perpetuate biases and that their implementation may lead to civil rights violations. Therefore, it includes a focus on equity and civil rights when AI is used in the justice system and the administration of federal government programmes and benefits.[44]

Moreover, the EO demonstrates that privacy and civil liberties must be protected as AI advances. AI is making it easier to extract, re-identify, link, infer, and act on sensitive information about people's identities, locations, habits, and desires, and AI's capabilities in these areas can increase the risk of personal data being exploited and exposed. To combat this risk, the federal government demands that the collection, use, and retention of data is lawful, secure, and mitigates privacy and confidentiality risks.[45]

The EO also makes specific reference to the risks produced by AI applications towards the American workforce.[46] A responsible development and use of AI requires

Trump AI Executive Order, see https://www.federalregister.gov/documents/2025/01/31/2025-02172/removing-barriers-to-american-leadership-in-artificial-intelligence

[42] Id., section 1.

[43] Id., section 2

[44] Id., section 2 (d).

[45] Id., section 2 (f).

[46] Id., section 2 (c). For an overall evaluation of the E.O. on AI, see also M.Sobel, *Demystifying the White House Executive Order on AI. What enterprise leaders need to know,* available at https://writer.com/

a commitment to supporting American workers, and as AI creates new jobs and industries, all workers need a seat at the table, including through collective bargaining, to ensure that they benefit from these opportunities. Within an enterprise, AI should not be deployed in ways that undermine rights, worsen job quality, encourage undue worker surveillance, lessen market competition, introduce new health and safety risks, or cause harmful labour-force disruptions.

4.2. The AI Bill of Rights

The EO outlines the administration's efforts to:

> build on the important steps that have already been taken — such as issuing the Blueprint for an AI Bill of Rights, the AI Risk Management Framework, and Executive Order 14091 of February 16, 2023 [...] in seeking to ensure that AI complies with all Federal laws and to promote robust technical evaluations, careful oversight, engagement with affected communities, and rigorous regulation.[47]

The Blueprint for an AI Bill of Rights comprises five principles and associated practices to guide the design, use, and deployment of automated systems to protect the rights of the American public:[48] safety, fairness, privacy, transparency, and human oversight. These principles apply to automated systems that can impact people's rights, opportunities, or access to critical resources.

The key idea behind the AI Bill of Rights[49] is that while AI has already brought extraordinary benefits in almost all sectors of Americans' daily lives—health, education, working conditions, agriculture, communication, energy, and transportation—and will keep redefining every part of American society and economy, progress must not come at the price of civil rights, democratic values, or foundational American principles.

---

blog/executive-order-ai/; For the impact of Joe Biden's EO on US citizens as well as on the US economy, see S. Kiosse/K.Fath/K.Harbaugh, *300 Days Since Biden's AI Executive Order: What have Federal Agencies Accomplished and What is on the Horizon?,* September 11, 2024, available at https://www.privacyworld.blog/2024/09/300-days-since-bidens-ai-executive-order-what-have-federal-agencies-accomplished-and-what-is-on-the-horizon/.

[47] US President's Executive Order, section 2 (d).

[48] White House, *What is the Blueprint for an AI Bill of Rights?,* available at https://www.whitehouse.gov/ostp/ai-bill-of-rights/what-is-the-blueprint-for-an-ai-bill-of-rights/

[49] White House, *Blueprint for an AI Bill of Rights. Making automated systems work for the american people,* available at https://www.whitehouse.gov/ostp/ai-bill-of-rights/

The normative framework set by the AI Bill of Rights is scheduled to apply to automated systems that have the potential to meaningfully impact the American public's rights, opportunities, or access to critical resources or services.[50] These should be enjoyed equally and be fully protected, regardless of the changing role that automated systems may play in their lives.[51] However, unlike the ten amendments that constitute the US Constitution's Bill of Rights, the AI Bill of Rights is not a binding document. It aims to support the development of policies and practices that protect civil rights and promote democratic values in the building, deployment, and governance of automated systems.[52]

All five principles and associated practices of the AI Bill of Rights form an overlapping set of backstops against potential harms that are expected to appear through AI applications. This purposefully overlapping framework, when taken as a whole, forms a blueprint to help protect the public from harm. The measures taken to realise the vision set forward in this framework should be proportionate with the extent and nature of the harm, or risk of harm, to people's rights, opportunities, and access.[53]

4.3. The US National AI Research and Development Strategic Plan (NAIRDSP)

The 2023 update of the NAIRDSP emphasises maintaining US leadership in AI through coordinated investments in fundamental and responsible AI research.[54] Strategies include fostering collaboration between human and AI systems, addressing ethical and societal implications, and improving datasets and computational resources. Particularly notable is the addition of a ninth strategy in this update, which advocates for international collaboration to address global challenges like sustainability and public health. This plan lays the groundwork for technological and scientific advancements, emphasising sustainable and equitable development.

This plan, which follows NAIRDSPs issued in 2016 and 2019, reaffirms eight strategies: a) seeking to make long-term investments in fundamental and responsible AI

[50] Id.

[51] Id.

[52] Id.

[53] White House, *Applying the Blueprint for an AI Bill of Rights*, available at https://www.whitehouse.gov/ostp/ai-bill-of-rights/applying-the-blueprint-for-an-ai-bill-of-rights/

[54] White House, *National artificial intelligence Research and development Strategic plan 2023 update,* available at https://www.whitehouse.gov/wp-content/uploads/2023/05/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf.

research; b) developing effective methods for human–AI collaboration; c) understanding and addressing the ethical, legal, and societal implications of AI; d) ensuring the safety and security of AI systems; e) advancing knowledge of how to design AI; f) developing shared public datasets and environments for AI training and testing; g) better understanding the national AI research and development workforce needs; and h) establishing a principled and coordinated approach to international collaboration in AI research. As mentioned above, it adds a ninth strategy: underscoring a principled and coordinated approach to international collaboration in AI research.[55]

This strategic approach reflects the federal government's recognition of AI's transformative potential and the need for a coordinated approach to mitigate risk. By emphasising principles such as equity, transparency, and safety, the document aligns with broader national priorities, including those outlined in the Biden administration's initiatives. However, it leaves governance and regulatory specifics to companion documents like the AI Risk Management Framework.[56]

AI risk management is a key component of responsible development and use of AI systems. Responsible AI practices can help align the decisions about AI system design, development, and uses with intended aims and values. Core concepts in responsible AI emphasise human centricity, social responsibility, and sustainability. AI risk management can drive responsible uses and practices by prompting organisations and their internal teams who design, develop, and deploy AI to think more critically about context and potential or unexpected negative and positive impacts. Understanding and managing the risks of AI systems help enhance trustworthiness, and in turn, cultivate public trust.[57]

5. The Normative Order of AI in China

In China, laws on data protection, cybersecurity, unfair competition, intellectual property, and e-commerce may be applied to the use of AI, like in other jurisdictions.

[55] Id., p. viii

[56] US Department of Commerce/National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, January 2023, available at https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf. The AI RMF refers to an AI system as an engineered or machine-based system that can, for a given set of objectives, generate outputs such as predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.

[57] Id., p.1.

However, the Cyberspace Administration of China (CAC)[58] has been world-first in introducing new laws specifically to regulate AI. In 2023, the CAC launched a series of measures targeted at the production of deepfakes, making China the first country to try control the explosive growth of this area of AI technology.[59]

For Chinese legislators, it has been increasingly clear that a sound and effective normative framework is a *sine qua non* for AI to succeed; specifically, to prevent monopolistic firms from restricting broader access to data and innovations, and to exclude those firms selling poor quality or harmful AI services. Allowing the latter to proliferate would disrupt markets and reduce trust towards the AI industry.[60]

China's recent legislative efforts have sought to regulate AI through a series of regulations like the Interim Measures for the Management of Generative AI Services,[61] the Algorithm Recommendation Regulation,[62] the Deep Synthesis Regulation,[63] the

[58] The Cyberspace Administration of China (CAC) is the country's top internet 'watchdog'

[59] For a comprehensive analysis of China's regulatory efforts of AI, from a jurisprudential, normative, and competition's approach, see A.H. Zhang, *The Promise and Perils of China's Regulation of Artificial Intelligence*, 63 (1) Columbia Journal of Transnational Law, pp. 101-157, at 103 (2024), available at https://ssrn.com/abstract=4708976.

[60] Ada Lovelace Institute, *Regulate to innovate: A route to regulation that reflects the ambition of the UK AI Strategy,* 2021, available at https://www.adalovelaceinstitute.org/report/regulate-innovate/; For a critical and comparative analysis of China's regulations on algorithms, see Friedrich Ebert Stiftung, China's Regulations on Algorithms, January 2023, available at https://library.fes.de/pdf-files/bueros/bruessel/19904.pdf; for China's broader approach to regulating the digital economy, see the analysis of Y.-W. Lei, *Delivering Solidarity: Platform Architecture and Collective Contention in China's Platform Economy*', 86 (2) American Sociological Review, pp. 279–309, 2021.

[61] The Interim Measures on Generative AI demonstrate the evolution of China's attempt to adopt tailored rules for generative AI—they are a series of administrative guidelines with legally binding effect that regulate several aspects of the generative AI products and services offered to consumers and the general public within China. The significance of the interim measures lies in the additional requirements they carry, and in the establishment of specific mechanisms for enforcement. Although an integral part of the Chinese legislative package on AI, and despite its importance for a comparative law analysis, the interim measures will not be further analysed in this paper. For a deeper analysis of the interim measures and their normative background, as well as for a comparative law approach, see S. Migliorini, *China's Interim Measures on generative AI: Origin, content and significance*, 53 Computer Law & Security Review, pp. 1-7 (2024), available at https://www.sciencedirect.com/science/article/abs/pii/S0267364924000529

[62] See China: Regulations on the Management of Algorithm Recommendation for Internet Information Services ([Unofficial translation by Digital Policy Alert), available at https://clairk.digitalpolicyalert.org/documents/china-regulations-on-the-management-of-algorithm-recommendation-for-internet-information-services/raw.

[63] See China: Regulations on the Management of Algorithm Recommendation for Internet Information Services ([Unofficial translation by Digital Policy Alert), available at https://clairk.digitalpolicyalert.org/documents/china-regulations-on-the-management-of-algorithm-recommendation-for-internet-information-services/raw.

Generative AI Regulation,[64] and the Draft Ethical Review Measures.[65] Together, these policies form a framework aimed at mitigating risks associated with AI technologies while fostering ethical and secure innovation. China's regulatory efforts cover an extensive range of obligations imposed on tech firms—ranging from content moderation to data protection, intellectual property, and ethical issues. China's regulatory approach, as noted, places it ahead of many jurisdictions, reflecting a decisive strategy to address technological challenges and opportunities presented by AI.

In this paper, we will focus on China's three most impactful regulations on AI to date: the Algorithm Recommendation Regulation, the Deep Synthesis Regulation, and the Generative AI Regulation.

5.1. The 2022 Algorithm Recommendation Regulation

The Algorithm Recommendation Regulation[66] applies to algorithm recommendation technologies—like generative and synthetic algorithms—that provide internet information services in China.[67] The regulation is the latest in a five-year arc of policymaking aimed at developing, regulating, and redefining operations of power in Chinese cyberspace, which began with the 2017 Cybersecurity Law.[68]

Focusing on algorithmic technologies such as personalised recommendations and filtering, the regulation aims to ensure transparency, fairness, and compliance with social ethics while mitigating risks like manipulation and addiction. For example, the provision of algorithm recommendation services must abide by laws and regulations, respect social morality and ethics, abide by business and professional ethics, and follow the principles of impartiality, fairness, openness and transparency, scientific rationality, and honesty.[69]

---

[64] See China Law Translate, *Interim Measures for the Management of Generative Artificial Intelligence Services,* July 13, 2023, available at https://www.chinalawtranslate.com/en/generative-ai-interim/

[65] For an overview of the Review, see S. Yang/C. Fung/B. Zhou, *AI Ethics: Overview (China),* January 2025, available at https://www.chinalawvision.com/2025/01/digital-economy-ai/ai-ethics-overview-china/.

[66] For the full text of the Regulation, see China: Regulations on the Management of Algorithm Recommendation for Internet Information Services ([Unofficial translation by Digital Policy Alert), available at https://clairk.digitalpolicyalert.org/documents/china-regulations-on-the-management-of-algorithm-recommendation-for-internet-information-services/raw.

[67] Latham&Watkins, *China's New AI Regulations*, p. 1, August 13, 2023, available at https://www.lw.com/.

[68] S.Stacks, *Testimony: Hearing on Promoting Competition, Growth, and Privacy Protection in the Technology Sector Senate Finance Subcommittee on Fiscal Responsibility and Economic Growth,* 2022, available at https://www.finance.senate.gov/imo/media/doc/Samm%20Sacks%20Testimony%20-%20Senate%20Finance%20-%20December%207 %202021.pdf

[69] See article 4 of the Algorithm Recommendation Regulation.

The Algorithm Recommendation Regulation provides that if any service provider violates the regulation, the penalties in accordance with other applicable laws and regulations shall apply, and in the absence thereof, the CAC and relevant authorities may issue warnings or public denouncements, or order rectifications within a time limit. If the violation is not rectified in time or involves aggravating circumstances, the authorities may order suspension of information updates and impose a fine of CNY 10,000–100,000.[70]

5.2. The 2023 Deep Synthesis Regulation

The Deep Synthesis Regulation defines deep synthesis technologies as 'technologies that use generative sequencing algorithms, such as deep learning and virtual reality, to create text, images, audio, video, virtual scenes, or other information'.[71] However, the technology is commonly known for its deepfake products (the term is a combination of 'deep learning' and 'fake')—using powerful ML and AI techniques to edit or synthesise visual and audio information.[72]

Other than providing a definition of deepfake content, the regulation clarifies the duties and responsibilities of providers and reflects a multi-stakeholder approach to supervision. It imposes requirements for algorithmic transparency, content labelling, and user identity verification. The emphasis on synthetic content labelling, especially in contexts like biometric manipulation, demonstrates a proactive stance against misuse.[73]

Under Article 4, deep synthesis services must comply with laws and regulations, respect social morality and ethics, adhere to the approved political direction, public opinion and value orientation, and promote positive and upstanding deep synthesis services. The same provision requires deep synthesis service providers and technical

[70] Id., article 31

[71] L. Zhang /Library of Congress, *China: Provisions on Deep Synthesis Technology Enter into Effect*, April 26, 2023, available at https://www.loc.gov/item/global-legal-monitor/2023-04-25/china-provisions-on-deep-synthesis-technology-enter-into-effect/

[72] G. Interesse, *China to Regulate Deep Synthesis (Deepfake) Technology Starting 2023,* December 20, 2022, available at https://www.china-briefing.com/news/china-to-regulate-deep-synthesis-deep-fake-technology-starting-january-2023/

[73] As far as the 'labelling' policy is concerned, both the EU AIA and the Deep Synthesis Regulation adopt the importance of clear labelling of AI-generated content. Although the EU regulator is more stringent, the Chinese counterpart restricts the need for labelling to cases where the content could mislead the public. See Latham&Watkins, *China's New AI Regulations*, p. 13 August 13, 2023, available at https://www.lw.com/, and S. Migliorini, supra note 60, at 6.

supporters to implement technical measures for training data security and to comply with personal information protection regulations if the training data contains personal information.[74]

Note that the regulation does not explicitly set out any penalties for non-compliance. Article 22 only provides that penalties may be imposed pursuant to other applicable laws and regulations should a deep synthesis service provider or technical supporter violate the regulation's provisions. It also does not specify the consequence of non-compliance by users or online platforms.[75]

5.3. The 2023 Generative AI Regulation

In July 2023, the CAC released the Generative AI Regulation, which was a response to the growth of LLMs and AI technology, such as the chatbot ChatGPT; although it covers the same ground as the Deep Synthesis Regulation, it places more emphasis on text generation and training data by requiring providers to ensure that both are 'true and accurate'.[76]

Covering a broader spectrum of generative AI technologies, this regulation[77] sets stringent requirements for data sourcing, training model accuracy, and user consent. Additionally, it mandates that content generated by AI reflects 'socialist core values,'[78] a unique inclusion that complies with China's ideological priorities.

The term 'generative AI technologies' is defined under the Generative AI Regulation as 'models and related technologies that have the ability to generate texts, pictures, sounds, videos, and other content'.[79] The regulation does not specify use cases or examples, but the term on its face is wider than 'synthesis technologies',

[74] Latham&Watkings, id., p. 10.

[75] Id., p. 23

[76] For a holistic and critical analysis not only of the Generative AI regulation, but of the most significant, so far, Chinese legislation on AI, see M. Sheehan, *China's AI Regulations and How They Get Made,* Carnegy Endowment for International Peace, July 2023, available at https://carnegie-production-assets.s3.amazonaws.com/static/files/202307-Sheehan_Chinese%20AI%20gov-1.pdf.

[77] For the English text of the final updated version of the regulation, see China Law Translate, *Interim Measures for the Management of Generative Artificial Intelligence Services,* July 13, 2023, available at https://www.chinalawtranslate.com/en/generative-ai-interim/. The original Chinese text is available at https://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm.

[78] Latham&Watkins, *China's New AI Regulations*, p. 13, 14, August 13, 2023, available at https://www.lw.com/

[79] Id., p.4

since it captures any content-generating technology; the content can be text, pictures, sounds, videos, and other information that is not on the internet.[80]

Article 4, a key provision, asks for core socialist values to be upheld. That means the prohibition of content such as:

> that inciting subversion of national sovereignty or the overturn of the socialist system, endangering national security and interests or harming the nation's image, inciting separatism or undermining national unity and social stability, [...] promoting ethnic hatred and ethnic discrimination, violence and obscenity, as well as fake and harmful information.[81]

Moreover, the same provides:

> During processes such as algorithm design, the selection of training data [...] and the provision of services, effective measures are to be employed to prevent the creation of discrimination such as by race, ethnicity, faith, nationality, region, sex, age, profession, or health.[82]

Finally, in the same article again, the provision and use of generative AI services must: 'Respect intellectual property rights and commercial ethics, and protect commercial secrets, advantages in algorithms, data, platforms, and so forth must not be used for monopolies or to carry out unfair competition'.[83]

Contrary to the Algorithm Recommendation Regulation, and similarly to the Deep Synthesis Regulation, this regulation does not impose specific penalties. The penalties are authorised by the relevant regulatory departments in accordance with the provisions of the People Republic of China (PRC) Cybersecurity Law, the PRC Data Security Law, the PRC Law on the Protection of Personal Information, the PRC Law on Scientific and Technological Progress, and other relevant legislation. Depending on the nature and scale of violation, besides penalties, the relevant Statal authorities are expected to issue warnings, circulate criticism, or order corrections in a set period of time; if these measures are not obeyed, an order may be given to suspend the provision of the related services.[84]

---

[80] Id.

[81] Article 4 (1), available at China Law Translate, *Interim Measures for the Management of Generative Artificial Intelligence Services,* available at https://www.chinalawtranslate.com/en/generative-ai-interim/.

[82] Id., Article 4(2)

[83] Id., Article 4 (3)

[84] Id., Article 21

The Generative AI Regulation seeks to find a balance between the innovation and development of generative AI technologies and their security.[85] It adopts an inclusive and prudent approach to supervising the application of generative AI technologies. It also lays down a range of legal requirements for providing generative AI services, with a focus on the privacy of pre-training and optimised data, content screening, and other related areas.[86]

All three Chinese regulations discussed share three common structural features: the choice of algorithms as a point of entry; the building of regulatory tools and bureaucratic know-how; and the vertical and iterative approach that lays the groundwork for a capstone AI law.[87] Regarding the final feature, contrary to the European AIA, which chooses a horizontal model of regulation, the Chinese legislators choose a vertical approach, i.e. they focus on a specific application of a technology.

China's iterative approach of the above regulations means that when the government finds a regulation to be ineffective or incorrect, it can simply draft a new one that fills the gap(s). Instead of choosing a 'future-proof' regulation, like the EU is trying to do, Chinese legislators have opted for the iterative approach as the most effective way to confront fast-changing AI technology.

At this point, in keeping with the regulatory and normative paradigms of the EU and the US, the Chinese government, in addition to its role as a regulator, also undertakes a multifaceted role in the AI ecosystem as a policymaker, investor, supplier, and customer.[88] The Chinese government has been a major financier and investor in the AI sector. Government procurement has been of vital importance in the growth of AI technology, specifically in facial recognition; in fact, the Chinese government has been a major consumer of facial recognition AI applications, engaging them mainly in the public services of security, transportation, and finance.

---

[85] For a very concise and accurate description of the regulation see B. Li/A. Zhou, *Navigating the Complexities of AI Regulation in China,* Reed Smith in-depth, August 4, 2024, available at https://www.reedsmith.com/en/perspectives/2024/08/navigating-the-complexities-of-ai-regulation-in-china.

[86] Id.

[87] M. Sheehan, *China's AI Regulations and How They Get Made,* Carnegy Endowment for International Peace, July 2023, p. 16, available athttps://carnegie-production-assets.s3.amazonaws.com/static/files/202307-Sheehan_Chinese%20AI%20gov-1.pdf.

[88] See A.H. Zhang, supra note 58, at 106, 107, and the relative references on the same pages. Politics play a pivotal role in the creation, development, and enforcement of AI regulation in China. Similarly to the EU and the US, China has a popular saying among the legal community: 'Small cases are adjudicated by law, while major cases are decided by politics'.

6. Comparative Analysis of Regulatory AI Initiatives from the EU, US, and China

The explosion of LLMs and generative AI models have resulted in an accelerated performance for AI across a broad spectrum of tasks. The unforeseen nature of AI's outputs raises concerns about the legal compliance and accuracy of the generated content.[89] Overall, the EU, US, and Chinese legal orders have made some headway in establishing a comprehensive, holistic, efficient, and balanced legal, normative, and ethical framework to handle AI's work of predicting, decision-making, and generating content. With the possible exception of the GDPR, however, their efforts do not seem adequate to cope with these novelties and challenges.

Comparing the normative and regulatory choices of the three jurisdictions on the AI ecosystem, a distinction appears with the EU and the US regulatory model falling on one side, and China's on the other. This is not because the EU and the US models are exactly the same, but because they are more alike than not in terms of structure and guiding doctrine. In comparison, there is a greater normative juxtaposition between China and the EU and the US.

Perhaps the most evident difference between the EU and the US has been the decentralised nature of the US regulatory approach to AI. Thus far, the US has a lack of comprehensive, federal AI law and scant enforcement actions against AI firms. States such as Colorado[90] and California,[91] following the risk-based, and not the holistic approach of the EU AIA, have been the pioneers in producing narrowly tailored AI regulations.

At the same time, a very visible similarity between the US and Chinese regulatory paradigms is the lenient and cautious approach shown by the market and regulatory bodies towards AI regulation, and the very strong prioritisation of technological de-

[89] For a greater elaboration on the balance between generative AI applications and the relevant regulatory efforts from the EU law perspective, see the analysis of C. Novelli/F. Casolari/P. Hacker/G. Spedicato/L. Floridi, *Generative AI in EU law: Liability, privacy, intellectual property, and cybersecurity*, 55 Computer Law & Security Review: The International Journal of Technology Law and Practice, pp.1-16, at 14 (2024), available at https://www.sciencedirect.com/science/article/abs/pii/S0267364924001328

[90] The Colorado Artificial Intelligence Act (CAIA), Act of May 17, 2024, ch. 198, 2024 Colo. Sess. Laws 1199 (codified at Colo. Rev. Stat. § 6-1-1701 et seq. (2024), available at https://leg.colorado.gov/bills/sb24-205. The CAIA, compared to the EU AIA is less stringent and more narrowly drawn in several key ways.

[91] The California Generative Artificial Intelligence Accountability Act, chaptered in September 29, 2024, available at https://calmatters.digitaldemocracy.org/bills/ca_202320240sb896.

velopment and innovation over the protection of fundamental rights and the rule of law. Both countries have adopted a more industry-favourable approach that gives the US and Chinese enterprises a greater competitive advantage against their European counterparts.

While all three legal orders adopt the risk-based approach, i.e. categorising AI systems by their potential harm (the differences are in the extent of the risk-based approach), and emphasise transparency, data protection, and accountability in algorithmic processes, China's regulations focus more heavily on content moderation, national security, and ideological alignment. This has resulted in leading Chinese AI enterprises being engaged in private litigation on various fronts:[92] copyright infringement, data privacy violations, defamation, and discrimination.

The breadth of China's regulation suggests significant compliance challenges for entities operating in AI-related sectors. Companies must navigate a complex web of obligations, including algorithm filing, content moderation, and adherence to ethical guidelines. For multinational corporations, the extraterritorial scope of these laws amplifies the need for vigilance, especially in training data and cross-border AI services.

A first strength of the Chinese normative and regulatory approach is its comprehensiveness. The country's layered approach in regulating different AI facets—algorithm use, generative content, and ethical research and development—ensures robust oversight. A second strength is its forward-looking perspective. By addressing emerging risks like deepfakes and generative models, China has positioned itself as a global leader in AI governance. Finally, a third strength is the integration of Chinese legislative acts within existing frameworks, which means that the regulations complement China's broader cybersecurity and data protection laws, creating an interconnected governance system.

On the other hand, there are some significant weaknesses to the Chinese AI regulatory approach. First, the ambiguity of the existing regulations: Provisions like the criteria for 'public opinion attributes' and 'social mobilization capabilities'[93] are vague, potentially leading to uneven enforcement. Second, there are significant en-

[92] For cases regarding AI law, and specifically cases on copyright infringement, see Master List of Lawsuits v. AI, ChatGPT, OpenAI, Microsoft, Meta, Midjourney & Other AI Cos., CHAT GPT IS EATING THE WORLD (Aug. 30, 2024), https://chatgptiseatingtheworld.com/2023/12/27/master-list-oflawsuits-v-ai-chatgpt-openai-microsoft-meta-midjourney-other-ai-cos/, also cited in A.H. Zhang, supra note 57, at 149.

[93] Latham&Watkins, *China's New AI Regulations*, p. 5, August 13, 2023, available at https://www.lw.com/

forcement gaps in the regulatory framework. While penalties are outlined, the lack of clarity on enforcement mechanisms could hinder compliance effectiveness.

And the third—most critical—point is the strong ideological connotation across Chinese regulatory initiatives. The need to protect and adhere to the 'socialist core values' may limit the innovation and global adaptability of Chinese firms operating cross-border. If the Chinese AI regulatory paradigm aims for extraterritoriality and a global role, the adherence to a specific political ideology is a high barrier. This challenge is non-existent in EU and US legislation.

The EU AIA provides a very clear regulatory framework, which categorises AI applications according to a structured, four-level risk-based pyramid. AI applications that threaten people's fundamental rights, safety, or livelihoods are deemed to involve an unacceptable level of risk and are prohibited. Registration, risk assessment, human oversight, and other obligations are included for high-risk AI practices, which constitute the most critical and controversial group of the risk-based approach. The limited risk AI applications are subject to transparency obligations.

Under the AIA, prohibited AI practices (Article 5) are subject to administrative fines of up to EUR 35,000,000 or, if the offender is an undertaking, up to 7% of its total worldwide annual turnover for the preceding financial year, whichever is higher. In contrast, the relevant penalties in China are based on AI rules, which are a set of departmental regulations focusing on different aspects of AI systems. China has no AI-specific law, but drafting such a law has been included in the State Council's 2023 legislative work plan.

Finally, the EU, on 16 June 2022, launched the Strengthened Code of Practice on Disinformation (the EU Disinformation Code) to reduce the impact of online disinformation.[94] Preceding the final release of the AIA, the EU Disinformation Code brings together a more diverse range of stakeholders than ever, empowering them to contribute to wide-ranging improvements by signing up to precise commitments relevant to their field. Such commitments include demonetising the dissemination of disinformation; guaranteeing transparency of political advertising; enhancing cooperation with fact-checkers; and facilitating researchers access to data.[95]

[94] European Commission, *The Strengthened Code of Practice on Disinformation 2022,* available at file:///C:/Users/User/Downloads/2022_Strengthened_Code_of_Practice_Disinformation_TeAETn-7bUP XR57PU2FsTqU8rMA_87585.pdf

[95] European Commission, Shaping Europe's digital future-Policy and legislation, 2022 Strengthened Code of Practice on Disinformation, available at https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation.

The EU Disinformation Code requires that large technology companies implement their commitment to counter deepfakes and false information. In addition, it focuses on regulating misinformation and disinformation, with other areas left to the GDPR (personal information protection), the amended Digital Services Act (transparency obligations and certain user protections applicable to online services and platforms), and the AIA. In contrast, the Chinese Deep Synthesis Regulation imposes broader obligations on a wider range of entities than the EU Disinformation Code. It applies to all participants involved in deep synthesis services, including service providers, technical supporters, users, and online platforms, providing a comprehensive framework of obligations from various perspectives, including cybersecurity, data management, personal information protection, algorithm audit and filing, real-name verification, and content moderation.[96]

All the above differences notwithstanding, and despite the fact that the EU, the US, and China have chosen a normative and regulatory path towards AI, in accordance with their different historical, cultural, ideological, jurisprudential background, there are five areas of alignment that bring forward a common normative basis.[97]

These five areas include: a) the formulation of the various AI regulations in line with the core principles of AI as published by the Organisation for Economic Co-operation and Development (OECD)[98] and endorsed by the G20.[99] These principles include respect for fundamental rights, sustainability, transparency, and strong risk management; b) a risk-based approach to regulating AI, which all three jurisdictions have adopted to varying extents. This means the EU, the US, and China shape and formulate their AI regulations to reflect the potential risks of AI applications to core values and fundamental rights such as transparency, non-discrimination, and security. The more likely an AI application is to cause harm to the above, the higher the risk

[96] Latham&Watkins, *China's New AI Regulations*, p. 24, August 13, 2023, available at https://www.lw.com/

[97] This remark embeds not only the three jurisdictions covered in the present paper, but almost all major AI regulatory attempts that have appeared globally in the last three years. Specifically, besides the EU, the US, and the China initiatives, other significant AI regulatory attempts have been in Canada, Japan, Brazil, Korea, Singapore, and the UK. For an overview and a synthesis of all the above jurisdictions, see FTI Consulting, *Bridging the Gap Between Artificial Intelligence Implementation, Governance and Democracy*, 2024, available at www.fticonsulting.com.

[98] For the OECD's AI strategy and AI principles see OECD AI, *OECD's live repository of AI strategies and policies*, January 2023-2024, available at https://oecd.ai/en/dashboards/overview, and OECD AI, *OECD Artificial Intelligence (AI) Principles*, May 2019, available at https://oecd.ai/en/ai-principles.

[99] For the G20 position on AI governance, see Group of 20 – forum for international economic cooperation, January 2023, available at https://www.g20.org/en/about-g20/#overview.

category and the stricter the obligations; c) both a sector-agnostic and sector-specific approach. Because of the varying use cases of AI, some jurisdictions focus on the need for sector-specific rules, in addition to sector-agnostic regulation.[100] A significant normative advantage lies in sector-agnostic regulations—due the fact that they are dependent in the short term on a specific technology, they qualify as 'future proof', and, thus, more rarely in need of updating; d) AI regulations drafted and enforced in line with other digital policy regulatory priorities, such as boosting the digital economy, personal data/privacy protection, cybersecurity, and intellectual property rights protection;[101] and e) multi-stakeholder collaboration, which applies more in the EU and US, rather than in China: AI rule-making is based not only on a central, governmental authority, such as the EU Commission, the US federal or State legislators, or the CAC, but also on the private sector, either in the form of the big tech companies, or through citizen initiatives. This model speaks for the transnational way of building a normative order, transcends the traditional multi-level constitutionalism of national, international, and supranational constitutionalism, and aspires to create a normative order that envisions collaboration between all affected stakeholders.

7. In Lieu of an Epilogue

What has become evident is that a comprehensive, sound, and either a holistic or sectorial AI regulatory framework needs to be in place in order to enable digital economy to reap the benefits of an innovative AI environment, and thrive within the AI digital ecosystem while comprehensively addressing the issues of fundamental rights, accountability, fairness, explainability, sustainability, transparency, safety and security, as well as all the governance, regulatory, and organisational risks related to AI.

A proactive stance in AI regulation is commendable for its foresight and scope, focusing on the ethical, normative, legal, and societal dimensions of AI. However, the probability for over-regulation, ideological constraints, and operational ambiguities presents challenges that participating stakeholders must navigate carefully. As global AI ecosystems evolve, the regulations discussed in this article may serve as both a benchmark and a cautionary tale for balancing innovation and accountability.

[100] EY, *The Artificial Intelligence (AI) global regulatory landscape. Policy trends and considerations to build confidence in AI*, 2023, p. 5, available at https://www.ey.com/content/dam/ey-unified-site/ey-com/en-gl/insights/ai/documents/ey-the-artificial-intelligence-ai-global-regulatory-landscape-final.pdf

[101] id.

References

Ada Lovelace Institute, *Regulate to innovate: A route to regulation that reflects the ambition of the UK AI Strategy,* 2021, available at https://www.adalovelaceinstitute.org/report/regulate-innovate/

Beck U., *Risk Society: Towards a Newest Modernity,* 1992, available at http://www.river-simulator.org/Resources/Anthropology/RiskSociety/RiskSocietyTowardsAnewModernity1992Beck.pdf

Bogdan M., *Concise Introduction to Comparative Law,* p. 18, Europa Law Publishing (2013)

Boyle J., Foucault in Cyberspace: Surveillance, Sovereignty, and Hardwired Censors, in 66 University of Cincinnati Law Review, pp. 177-205 (1998)

Bradford A., *Europe's Digital Constitution,* VerfBlog, 06/09/2023, available at https://verfassungsblog.de/europes-digital-constitution

California Generative Artificial Intelligence Accountability Act, chaptered in September 29, 2024, available at https://calmatters.digitaldemocracy.org/bills/ca_202320240sb896

Celeste E., *Digital Constitutionalism The Role of Internet Bills of Rights*, pp. 84,85, Routledge (2023)

Chemerinski E., *States' Misguided Attempts to Regulate Social Media*, March 1, 2024, available at https://www.governing.com/policy/states-misguided-attempts-to-regulate-social-media

China Law Translate, *Interim Measures for the Management of Generative Artificial Intelligence Services,* July 13, 2023, available at https://www.chinalawtranslate.com/en/generative-ai-interim/

China: Regulations on the Management of Algorithm Recommendation for Internet Information Services (Unofficial translation by Digital Policy Alert), available at https://clairk.digitalpolicyalert.org/documents/china-regulations-on-the-management-of-algorithm-recommendation-for-internet-information-services/raw

Christensen Institute, *Disruptive Innovation Theory: The theory that influenced Steve Jobs, Jeff Bezos, and Andy Grove*, 2024, available at https://www.christenseninstitute.org/theory/disruptive-innovation/

Collingridge D., *The Social Control of Technology,* Frances Pinter (1980)

Colorado Artificial Intelligence Act (CAIA), Act of May 17, 2024, , available at https://leg.colorado.gov/bills/sb24-205.

Comply Advantage, *AI Regulation in the US — A Guide for Compliance Officers*, 2023, available at www.complyadvantage.com.

Council of Europe, *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, September 5, 2024, available at https://rm.coe.int/1680afae3c

Deguchi A./C.Hirai/H.Matsuoka/T.Nakano/K.Oshima/M.Tai/S.Tani, What is Society 5.0?, available at https://www.researchgate.net/publication/341746797_What_Is_Society_50.

Eliot L., *AI Ethics Wary About Worsening Of AI Asymmetry Amid Humans Getting The Short End Of The Stick,* 19 August 2022, available at https://www.forbes.com/sites/lanceeliot/2022/08/19/ai-ethics-wary-about-worsening-of-ai-asymmetry-amid-humans-getting-the-short-end-of-the-stick/

European Commission, *Shaping Europe's digital future-Policy and Legislation*, 2 June 2024, available at https://digital-strategy.ec.europa.eu

European Commission, Shaping Europe's digital future-Policy and legislation, 2022 Strengthened Code of Practice on Disinformation, available at https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation

European Commission, *The Strengthened Code of Practice on Disinformation 2022,* available at file:///C:/Users/User/Downloads/2022_Strengthened_Code_of_Practice_Disinformation_TeAETn7bUP XR57PU2FsTqU8rMA_87585.pdf

Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, October 30, 2023, available at https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30 /

EY, *The Artificial Intelligence (AI) global regulatory landscape. Policy trends and considerations to build confidence in AI*, 2023, available at https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/ai/ey-the-artificial-intelligence-ai-global-regulatory-landscape.pdf

Friedrich Ebert Stiftung, China's Regulations on Algorithms, January 2023, available at https://library.fes.de/pdf-files/bueros/bruessel/19904.pdf

FTI Consulting, *Bridging the Gap Between Artificial Intelligence Implementation, Governance and Democracy*, 2024, available at www.fticonsulting.com

Genus A. /A.Stirtling, *Collingridge and the dilemma of control: Towards responsible and accountable innovation,* 47(1) Research Policy, 2018, pp. 61-69, available at https://doi.org/10.1016/j.respol.2017.09.012

Georgieva K., *AI Will Transform the Global Economy,* January 14, 2024, available at https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity

Hacker R., *The Asymmetry of Knowledge: How AI Unveils the Invisible Realms of Science,* 7 June 2024, available at https://roberthhacker.medium.com/the-asymmetry-of-knowledge-how-ai-unveils-the-invisible-realms-of-science-ca0499847514

Interesse G., *China to Regulate Deep Synthesis (Deepfake) Technology Starting 2023,* December 20, 2022, available at https://www.china-briefing.com/news/china-to-regulate-deep-synthesis-deep-fake-technology-starting-january-2023/

Kempf A.L. /N. Rauer, *A guide to high-risk AI systems under the EU AI Act,* 13 February 2024, available at https://www.pinsentmasons.com/out-law/guides/guide-to-high-risk-ai-systems-under-the-eu-ai-act

Kiosse S. /K.Fath/K.Harbaugh, *300 Days Since Biden's AI Executive Order: What have Federal Agencies Accomplished and What is on the Horizon?,* September 11, 2024, available at https://www.privacyworld.blog/2024/09/300-days-since-bidens-ai-executive-order-what-have-federal-agencies-accomplished-and-what-is-on-the-horizon/

Koops B.J. /R. Leenes/P. De Hert (eds.), *Constitutional Rights and New Technologies: A Comparative Study Covering Belgium, Canada, France, Germany, Sweden, and the United States*, Tilburg University- Institute for Law, Technology, and Society (2007), available at www.uvt.nl/tilt.

Latham&Watkins, *China's New AI Regulations*, p. 1, August 13, 2023, available at https://www.lw.com/

Lei Y.-W., *Delivering Solidarity: Platform Architecture and Collective Contention in China's Platform Economy*', 86 (2) American Sociological Review, pp. 279–309 (2021)

Li B. /A. Zhou, *Navigating the Complexities of AI Regulation in China*, Reed Smith in-depth, August 4, 2024, available at https://www.reedsmith.com/en/perspectives/2024/08/navigating-the-complexities-of-ai-regulation-in-china.

Mantelero A., *The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template*, 54 Computer Law & Security Review: The International Journal of Technology Law and Practice, pp. 1-31 (2024), available at https://sciencedirect.com/sciene/article/pii/S0267364924000864.

Marwala T. /E. Hurwitz, *Artificial Intelligence and Asymmetric Information Theory*, October 2015, available at https://www.researchgate.net/publication/282906709_Artificial_Intelligence_and_Asymmetric_Information_Theory

Migliorini S., *China's Interim Measures on generative AI: Origin, content and significance*, 53 Computer Law & Security Review, pp. 1-7 (2024), available at https://www.sciencedirect.com/science/article/abs/pii/S0267364924000529

Millar C. /M.Lockett/T.Ladd, *Disruption: Technology, innovation and society*, Special Issue of Technological Forecasting and Social Change an international journal, pp. 254-260, 2017 available at https://www.researchgate.net/publication/321015457_Disruption_Technology_innovation_and_society#

Nosta J., *The Precarious Asymmetries of Human-AI Relationships. Bridging the communication chasm between humans and AI,* 1 April 2024, available at https://www.psychologytoday.com/us/blog/the-digital-self/202404/the-precarious-asymmetries-of-human-ai-relationships

Novelli C. /F. Casolari/P. Hacker/G. Spedicato/L. Floridi, *Generative AI in EU law: Liability, privacy, intellectual property, and cybersecurity*, 55 Computer Law & Security Review:

The International Journal of Technology Law and Practice, pp.1-16 (2024), available at https://www.sciencedirect.com/science/article/abs/pii/S0267364924001328

OECD AI, *OECD's live repository of AI strategies and policies*, January 2023-2024, available at https://oecd.ai/en/dashboards/overview

OECD AI, *OECD Artificial Intelligence (AI) Principles*, May 2019, available at https://oecd.ai/en/ai-principles

OECD, *Artificial Intelligence & Responsible Business conduct*, 2018, available at https://mneguidelines.oecd.org/RBC-and-artificial-intelligence.pdf.

Regulatory Horizons Council, *The Future of Technological Innovations and the role of Regulation,* pp. 6-7, 2021

Rodotà S., *A Constitution for the Internet* (2006), available at http://www.federalist-debate.org

Salokannel P., *The Impact of AI: How Artificial Intelligence is Transforming Society*, March 7, 2024, available at https://www.3dbear.io/blog/the-impact-of-ai-how-artificial-intelligence-is-transforming-society

Sheehan M., *China's AI Regulations and How They Get Made,* Carnegie Endowment for International Peace, July 2023, available at https://carnegie-production-assets.s3.amazonaws.com/static/files/202307-Sheehan_Chinese%20AI%20gov-1.pdf

Siems M., *Comparative Law*, p. 4, Cambridge University Press (2014)

Sobel M., *Demystifying the White House Executive Order on AI. What enterprise leaders need to know,* available at https://writer.com/blog/executive-order-ai/

Stacks S, *Testimony: Hearing on Promoting Competition, Growth, and Privacy Protection in the Technology Sector Senate Finance Subcommittee on Fiscal Responsibility and Economic Growth,* 2022, available at https://www.finance.senate.gov/imo/media/doc/Samm%20Sacks%20Testimony%20-%20Senate%20Finance%20-%20December%207%202021.pdf

Tarver E., *The Biggest Companies in Silicon Valley*, November 23, 2024, available at https://www.investopedia.com/articles/markets/103015/biggest-companies-silicon-valley.asp.

Teubner G., Verfassungsfragmente: Gesellschaftlicher Konstitutionalisierung in der Globalisierung, Suhrkamp, Berlin (2012)

Teubner,G. Globale Zivilverfassungen: Alternativen zur staatszentrierten Verfassungstheorie, in 63 Zeitschrift für ausländisches öffentliches Recht und Völlkerrecht, pp. 1-28 (2003)

Tony Blair Institute for Global Change, *The Impact of AI on the Labour Market*, November 2924, available at https://assets.ctfassets.net/75ila1cntaeh/5yiS3TBdF96J4Kt0tFmj3k/c4b260667ed70fc791020268cfed443b/1mABBNuF6Anp7zTv1ZeIxe--155801112024

Trump AI Executive Order, available at https://www.federalregister.gov/documents/2025/01/31/2025-02172/removing-barriers-to-american-leadership-in-artificial-intelligence

Tsagourias N., The legal status of cyberspace, in N. Tsagourias/R.Buchan (eds), *Research Handbook on International Law and Cyberspace*, pp 13-29, Edward Elgan Publishing (2015)

United Nations, *The Global Digital Compact*, September 2024, available at ttps://www.un.org/techenvoy/global-digital-compact.

US Department of Commerce/National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*, January 2023, available at https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

Utterback J.M. /H.J. Acee, Disrupted Technologies: An Expanded View, 9(1) International Journal of Innovation Management, pp. 1–17 (2005)

Veale M. /F.Z. Borgesius, *'Demystifying the Draft EU Artificial Intelligence Act. Analysing the good, the bad, and the unclear elements of the proposed approach',* 4 Computer Law Review International, pp. 97-112 (2021)

Vergerio C., *Beyond the Nation-State,* 2021, available at https://www.bostonreview.net/articles/beyond-the-nation-state/

White House, *Applying the Blueprint for an AI Bill of Rights*, available at https://www.whitehouse.gov/ostp/ai-bill-of-rights/applying-the-blueprint-for-an-ai-bill-of-rights/

White House, *Blueprint for an AI Bill of Rights. Making automated systems work for the american people,* available at https://www.whitehouse.gov/ostp/ai-bill-of-rights/

White House, *National artificial intelligence Research and development Strategic plan 2023 update,* available at https://www.whitehouse.gov/wp-content/uploads/2023/05/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf

Witzel L., *5 Things You Must Know Now About the Coming EU AI Regulation*, 2021, available at https://medium.com/@loriaustex/5-things-you-must-know-now-about-the-coming-eu-ai-regulation-d2f8b4b2a4a9

Wright D. /P. De Hert (eds), *Privacy Impact Assessment*, Springer (2012)

WTO, *Introduction: 'Disruptive…' or (just) 'Emerging' technologies*?, 2022, available at https://www.wto.org/english/res_e/booksp_e/wco_wto_introduction.pdf

WTO, *The Policies of AI and Trade*, November 2024, available at https://www.wto.org/english/res_e/booksp_e/trading_intelligence_ch3_e.pdf

Yang S. /C. Fung/B. Zhou, *AI Ethics: Overview (China),* January 2025, available at https://www.chinalawvision.com/2025/01/digital-economy-ai/ai-ethics-overview-china/

Zhang A.H., *The Promise and Perils of China's Regulation of Artificial Intelligence*, 63 (1) Columbia Journal of Transnational Law, pp. 101-157, at 103 (2024), available at https://ssrn.com/abstract=4708976

Zhang L./Library of Congress, *China: Provisions on Deep Synthesis Technology Enter into Effect*, April 26, 2023, available at https://www.loc.gov/item/global-legal-monitor/2023-04-25/china-provisions-on-deep-synthesis-technology-enter-into-effect/

Ziccardi G., *Resistance, Liberation Technology and Human Rights in the Digital Age*, pp. 152-159, Springer (2013).

OPINION PAPER

Freedom of Expression and the Right to Respect Private Life

Loukis Loucaides[1]

Freedom of expression, as safeguarded in various constitutional and international law instruments, has been developed and expanded by the case-law and the doctrine of western countries and by the struggles of journalists in a dynamic and progressive manner, in line with the liberal trends and ideas of democracy that have emerged during the years following the Second World War. However the concept of freedom of expression was enlarged with an idealistic approach, to such an extent that other legitimate interests or rights, such as the right to respect the private life and the reputation of individuals, had on occasions to be subjected to inroads and lost, to some substantial degree, the required protection which those values deserve. An attempt will be made to deal with the topic of freedom of speech and the right of privacy on the basis of the provisions of the European Convention on Human Rights and the relevant case-law of the European Court of Human Rights. Freedom of speech is safeguarded by Article 10 of the Convention, the relevant parts on which we need to concentrate being as follows:

> '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 exercise of these freedoms, since it carries with it duties and responsibilities, may be subject to such conditions, restrictions or penalties as are prescribed by law, and are necessary in a democratic society... for the protection of the reputation or rights of others...'

In *Handyside* v. *UK,* the Court, in a much-quoted passage, said:

> 'Freedom of expression constitutes one of the essential foundations of a [democratic] society, one of the basic conditions for its progress and for the development of every man. Subject to paragraph 2 of Article 10, it is applicable not only to "information"' or "ideas" favourably received or regarded as inoffensive, but also to those that offend, shock or disturb the state or any sector of the popula-

[1] Former Judge of the European Court of Human Rights.

> tion. Such are the demands of that pluralism, tolerance and broadmindedness without which there is no "democratic society"'.[2]

In another case the Court stated the following:

> 'One factor of particular importance for the Court's determination in the present case is the essential function the press fulfils in a democratic society. Although the press must not overstep certain bounds, in particular, in respect of the reputation and rights of others and the need to prevent the disclosure of confidential information, its duty is nevertheless to impart -in a manner consistent with its obligations and responsibilities- information and ideas on all matters of public interest. In addition, the Court is mindful of the fact that journalistic freedom also covers possible recourse to a degree of exaggeration, or even provocation. In cases such as the present one the national margin of appreciation is circumscribed by the interest of democratic society in enabling the press to exercise its vital role of "public watchdog" in imparting information of serious public concern.'[3]

In the *Lingens* case,[4] the Court pointed out that it is incumbent on the press 'to impart information and ideas on political issues just as on those of other areas of public interest' and added that '...freedom of the press affords the public one of the best means of discovering and forming an opinion of the ideas and attitudes of political leaders'. This is why, in this context, '...the limits of acceptable criticism are accordingly wider as regards a politician as such than as regards a private individual'. Unlike the latter, the former inevitably and knowingly lays himself open to close scrutiny of his every word and deed by both journalists and the public at large, and he must consequently display a greater degree of tolerance'.

However, in my view, the principle established by the jurisprudence, that there is more latitude in the exercise of freedom of expression in the area of political speech or debate, and in cases of criticism of politicians, should not be interpreted as allowing the publication of any unverified defamatory statements. To my mind this principle means simply that in those cases that I have just mentioned, some excessive offending effect should be tolerated and should not be sanctioned. But the principle does not mean that the reputation of politicians is at the mercy of the mass media or other persons dealing with politics, or that such reputation is not entitled to the same

2 Series A 24 §49 (1996).

3 *Bladet Tromso and Stensaas* v. *Norway*, no. 21908/93 ECHR 1999-111.

4 *Lingens* judgment of 8 July 1986, Series A no. 103, 41.

legal protection as that of any other individual. Reputation is a sacred value for every person, including politicians, and is safeguarded as a human right under the Convention for the benefit of every individual without exception.

It is important to draw attention to the fact that, for many years, the jurisprudence of the Court has developed on the premise that, while freedom of speech is a right expressly guaranteed by the Convention, the protection of reputation was simply a ground of permissible restrictions on the right in question, which may be regarded as justified interference with the freedom of expression only if it was 'necessary in a democratic society', in other words if it corresponds to 'a pressing social need' and is 'proportionate to the aim pursued' and if 'the reasons given were relevant and sufficient'. As a consequence of this approach, the case-law on the subject of freedom of speech has on occasion shown excessive sensitivity and granted over-protection in respect of interference with freedom of expression, as compared with interference with the right to reputation. Freedom of speech has been upheld as a value of primary importance, which in many cases could deprive deserving plaintiffs of an appropriate remedy for the protection of their dignity.

The case of *Tromso and Stensaas* v. *Norway*[5] illustrates, in my opinion, the over-protective approach of the Court as regards freedom of speech in contrast to the right to reputation. The relevant facts of the case may be summed up as follows: The applicant's newspaper had published an interview with a certain Mr Lindberg who, following the instructions of the Ministry of Fisheries, had carried out an inquiry and prepared a report in relation to allegations that certain seal hunters of a seal hunting vessel had used illegal hunting methods. The report was, by order of the Ministry in question, exempted from public disclosure, the ground being that it contained allegations of statutory offences. The applicant's newspaper reproduced some of the statements of Mr Lindberg included in his report, concerning the alleged breaches of the seal-hunting regulations by members of the crew of the vessel. Defamation proceedings were instituted by the persons concerned against the applicant seeking compensation and requesting that certain of the statements in question be declared null and void. Judgment was entered in favour of the plaintiffs for certain statements which were found to be false and defamatory.

The Court, by a majority, found that the judgment against the applicant declaring null and void the above statements and awarding compensation to the plaintiff amounted to a violation of Article 10 of the Convention which safeguards freedom of

[5] No. 21908/93 ECHR 1999-III.

expression. The majority found that the impugned judgment served the legitimate aim of protecting the reputation of the crew members, but came to the conclusion that such judgment was not necessary 'in a democratic society'. In this respect, the majority stressed that the publications in question concerned a matter of public interest,[6] and pointed out the duty of the Court to conduct a careful scrutiny of measures interfering with the freedom of speech when such measures are capable of 'discouraging the participation of the press in debates over matters of legitimate public concern'. Reference was also made to the duty of journalists 'to act in good faith in order to provide accurate and reliable information in accordance with the ethics of journalism'.[7] It was also accepted that the impugned expressions consisted of factual statements. Nevertheless, the Court found that the applicant's newspaper was dispensed from the 'obligation to verify factual statements that were defamatory of private individuals' because such statements 'did not emanate from the newspaper itself',[8], but originated in a report drawn up by Mr Lindberg in an official capacity and could in the circumstances of the case be considered as reliable without being necessary for the newspaper 'to carry out its own research into the accuracy of the facts reported'. As regards the nature of the defamation, the Court found that, although the four impugned statements 'implied reprehensible conduct they were not particularly serious'. The other two allegations -that seals had been skinned alive and that furious hunters had beaten up Mr Lindberg and threatened to hit him with a gaff- 'were more serious but were expressed in rather broad terms and could be understood by readers as having been presented with a degree of exaggeration'.

One could also refer to other cases manifesting the same trend. The case of *Nilsen and Johnsen* v. *Norway*[9] contains basic elements of such a trend. In that case the applicants were police officers representing the Norwegian and the Bergen Police

[6] Eleni Micha, in her article 'Defamation: Dignity Lost?' 16(3) *Netherland Quarterly of Human Rights* 272 points out '... the subjectivity of the distinction between matters of public and private concern. Who could really decide on the issue when the mass media -having a really mighty role today- are able to manipulate people's opinions and promote as matters of public concern, issues really insignificant?' c.f. Justice Brennan's criticism of the 'public-private dichotomy created by the majority' with 'almost no guidance' concerning its extent with the consequence of containing an impoverished definition of 'public concern': *Dun & Bradstreet* v. *Greenmoss Builders* 472 U S 749 (1985).

[7] In the case of *Fressoz and Roire* v. *France*, 21 January 1999, the Court confirmed the principle that Article 10 'protects journalists' rights to divulge information on issues of general interest provided that they are acting in good faith and on an accurate factual basis and provide "reliable and precise" information in accordance with the ethics of journalism'.

[8] No. 21908/93 ECHR 1999-III, § 66.

[9] ECHR 1999 - VIII - (25/11/99).

Associations respectively. They were found responsible for having published false defamatory statements concerning Professor Bratholm who, in a book published by him, attributed to the police incidents of unlawful violence, brutality and other forms of misconduct. The impugned statements amounted to allegations that Professor Bratholm 'with the intent of harming the police is deliberately imparting false information on police brutality, that he has dishonest motives and that he is taking part in a private investigation for the purpose of fabricating allegations of police brutality'. The domestic court found that these allegations were assertions of fact, the truth of which could be tested by evidence. In finding in favour of the professor, the Norwegian Supreme Court observed *inter alia* that:

> 'However, freedom of expression does not go as far as [allowing] every statement in a debate, even if the debate relates to matters of public concern. Freedom of expression must be weighed against the rights of the injured party. ... Nor do accusations of lies, improper motives, dishonesty ... serve to promote the freedom of expression but, perhaps, rather to suppress or prevent debate which should have been allowed to take place.'

The European Court found, by majority, that there has been a violation of the freedom of expression as a result of the judgment against the applicants. As regards the statements which imputed improper motives or intentions to Professor Bratholm, the Court found that they should not be regarded as allegations of facts requiring the applicants to prove their truth, because, according to the Court, those statements 'were intended to convey the applicants' own opinions and were thus rather akin to value judgements'. The majority took into account the fact that the applicants were speaking on behalf of the members of the police associations criticised by Professor Bratholm and therefore they 'were not entirely unjustified in claiming that they were entitled to "hit back in the same way"'.

Nobody can disagree with the importance of the freedom of speech, especially that of the media, as an essential factor of democratic society. However, the question is whether the protection accorded to such a freedom may, under any circumstances, be so excessive as to deprive the victims of false defamatory statements or of unnecessary intrusion in their private life of the necessary effective remedy. I reiterate here the view that I expressed in the case of *Tromso*,[10] which concerns the balancing of freedom of speech on the one hand and the right to reputation on the other as follows:

[10] No. 21908/93 ECHR 1999-III.

> 'The press is, in our days, an important and powerful means of influencing public opinion. The impressions that may be created through a publication in the press are usually more decisive than the reality because until the reality is found out the impressions prevail. And the reality may never be discovered or when it is disclosed it may be too late to remedy the damage done by the original impressions[11]. The press is in effect exercising a significant power and should be subject to the same restraints applicable to any exercise of power, namely it should avoid abuse of its power, it should act in a fair way and respect the rights of others. Therefore, whenever it engages in imparting information of a defamatory nature it should respect basic rules of natural justice such as the rule of *audi alteram partem*'.

I take this opportunity to point out some adverse consequences resulting from over-protection of freedom of expression at the expense of the right to reputation. The main argument in favour of protecting freedom of expression, even in cases of inaccurate defamatory statements, is the encouragement of uninhibited debate on public issues. But the opposite argument is equally strong: the suppression of untrue defamatory statements, apart from protecting the dignity of individuals, discourages false speech and improves the overall quality of public debate through a chilling effect on irresponsible journalism. Moreover, such debates may be suppressed if the potential participants know that they will have no remedy in the event that false defamatory accusations are made against them. The prohibition of defamatory speech also eliminates misinformation in the mass media and effectively protects the right of the public to truthful information. Furthermore, false accusations concerning public officials, including candidates for public office, may drive capable persons away from government service, thus frustrating rather than enhancing the political process.

In recent years the Court has expressly recognised that protection of reputation is a right which is covered by the scope of the right to respect for one's private life under Article 8 § 1 of the Convention,[12] even though the relevant jurisprudence has

[11] As rightly observed by the US Supreme Court in the Gertz case op. cit. (n. 9) 'Of course, an opportunity for rebuttal seldom suffices to undo harm of defamatory falsehood. Indeed, the law of defamation is rooted in our experience that the truth rarely catches up with a lie'. See also Thomas David Jones 'Human Rights: Group Defamation, Freedom of Expression and the Law of Nations', Chap. Ill, 'The Myth of Absolutism: The First Amendment Right to Freedom of Expression' at p. 68: 'Once tarnished, reputation is virtually impossible to restore'.

[12] *See*, *Chauvy and Others* v. *France,* no. 64915/01, ECHR 2004-VI; *Abeberry v. France* (dec.), no. 58729/00, 21 September 2004; and *White* v. *Sweden*, no. 42435/02, 19 September 2006.

not expanded on this novel approach, nor has it been invoked in other cases involving freedom of speech and defamation. In the light of this case-law, any defamatory statement amounts to interference with the right guaranteed by the Convention which has the same status as that of freedom of expression and such interference can only be justified if it satisfies the requirements of permissible restrictions on the exercise of such right, i.e., it must be necessary in a democratic society etc. It will be more difficult to defend a defamatory statement for the purposes of the protection under the Convention when it is examined as interference with a right recognised under the Convention, rather than as a necessary restriction on freedom of expression.

The right to respect for private life is expressly safeguarded under Article 8 of the Convention. It is subject to restrictions similar to those of Article 10, which safeguards freedom of expression. As to the concept of private life, the jurisprudence of the Court has established the following: Private life is a broad term not susceptible to exhaustive definition. Elements such as gender identification, name, sexual orientation and sexual life are important elements of the personal sphere protected by Article 8. That Article also protects a right to identity and personal development, and the right to establish and develop relationships with other human beings and the outside world and it may include activities of a professional or business nature.[13] There is, therefore, a zone of interaction of a person with others, even in a public context, which may fall within the scope of 'private life'[14] Under the influence of prevailing moral and social values and conditions, developing jurisprudence determines the scope of the concept of private life. It is also inevitably affected by the passage of time. In any event it should be borne in mind that freedom of expression and the right to privacy are in principle of equal value. If the Court has to decide in any particular case which of the two rights should prevail over the other, the matter is decided through the application of a balancing test of the competing rights on the basis of the facts and circumstances of the case. As rightly pointed out by Judge Dean Spielman when he was asked in an interview how this balance is found: 'It all boils down to the issue of proportionality. If sacrifice of one value has been disproportionate then we rule in favour of the sacrificed value. It really depends on the circumstances of the case and also on the issue of whether the protected value is at the heart of privacy or on the boundaries of it'.

[13] *See* inter alia Peck v. UK, application no. 44647/98

[14] See *P.G. andJ.H.* v. *the United Kingdom,* no. 44787/98, § 56, ECHR 2001-IX.

This is illustrated by the approach of the Court in relevant cases to some of which I am now going to refer. An important case on this subject is that of *Von Hannover* v. *Germany*, the facts of which are as follows. Princess Caroline von Hannover was campaigning since the beginning of the 1990s to prevent the publication of photographs revealing her private life. On this basis she unsuccessfully applied, on several occasions, to the German courts to obtain an injunction preventing any further publication of a series of photographs which had appeared in the 1990s in certain German magazines. She claimed that they infringed her right to protection of her private life and her right to control the use of her image. In the landmark judgment of 15 December 1999, the German Federal Constitutional Court granted the applicant's injunction regarding the photographs in which she appeared with her children, on the ground that their need for protection of their intimacy was greater than that of adults. However, the Constitutional Court considered that the applicant, who was undeniably a contemporary 'public figure', had to tolerate the publication of photographs of herself in a public place, even if they showed her in scenes from her daily life rather than engaged in her official duties. The Constitutional Court referred in that connection to the freedom of the press and to the public's legitimate interest in knowing how such a person generally behaved in public.[11]

The applicant maintained that the decisions of the German courts infringed her right to respect for her private life, as guaranteed by Article 8 of the Convention, since they failed to afford her adequate protection from the publication of photographs taken without her knowledge by paparazzi, on the ground that, in view of her origins, she was undeniably a contemporary 'public figure'. In its judgment of 24 June 2004,[15] the Court observed at the outset that there was no doubt that the publication by various German magazines of photographs of the applicant in her daily life, either on her own or with other people, fell within the scope of her private life. It was therefore necessary to balance protection of the applicant's private life against freedom of expression, as guaranteed by Article 10 of the Convention.[16] Although freedom of expression applied to the publication of photographs, this was an area in which the protection of the rights and reputation of others took on a particular importance, as it did not concern the dissemination of 'ideas', but of images containing very personal or even intimate 'information' about an individual. Furthermore, photos appearing in the tabloid press were often taken in a climate of continual harassment which in-

[15] *von Hannover v. Germany,* no. 59320/00, ECHR 2004-VI.

[16] § 53 and § 58.

duced in the person concerned a very strong sense of intrusion into their private life or even of persecution.[17]

The Court considered that the decisive factor in balancing the protection of private life against freedom of expression should lie in the contribution that the published photographs and articles made to a debate of general interest. In the case before it, the photographs showed scenes from the applicant's daily life, engaging in activities of a purely private nature. The Court noted the circumstances in which the photographs had been taken: without the applicant's knowledge or consent and, in some instances, in secret. It was found that they made no contribution to a debate of public interest, since the applicant exercised no official function and the photographs and articles related exclusively to details of her private life.[18] Furthermore, while the general public might have a right to information, including, in special circumstances, on the private life of public figures, they did not have such a right in this instance. The Court considered that the general public did not have a legitimate interest in knowing Caroline von Hannover's whereabouts or how she behaved generally in her private life, even if she appeared in places that could not always be described as secluded and was well known to the public. Even if such a public interest existed, just as there was a commercial interest for the magazines to publish the photographs and articles, those interests had to yield, in the Court's view, to the applicant's right to the effective protection of her private life.[19]

The Court reiterated the fundamental importance of protecting private life from the point of view of the development of every human being's personality and said that everyone, including people known to the public, had to have a 'legitimate expectation' that his or her private life would be protected. The criteria that had been established by the domestic courts for distinguishing a figure of contemporary society *'par excellence*' from a relatively public figure, were not sufficient to ensure the effective protection of the applicant's private life and she should, in the circumstances of the case, have had a 'legitimate expectation' that her private life would be protected.[20] Having regard to all the foregoing factors, and despite the margin of appreciation afforded to the State in this area, the Court considered that the German courts had not struck a fair balance between the competing interests. Accordingly, it held that there had been a violation of Article 8 of the Convention.

17 § 59.
18 §§ 60 et seq.
19 § 77.
20 Esp. § 78.

In the more recent case of *Craxi (n° 2j)*[21] concerning the publication of a private conversation, the Court, in balancing the different interests at stake -namely that of freedom of the press and the right to privacy- applied the same test it usually applies whilst examining the justification of an interference by public authorities. In the judgment, the following are stated:

> 'However, public figures are entitled to the enjoyment of the guarantees set out in Article 8 of the Convention on the same basis as every other person. Inparticular, the public interest in receiving information only covers facts which are connected with the criminal charges brought against the accused. This must be borne in mind by journalists when reporting on pending criminal proceedings and the press should abstain from publishing information which is likely to prejudice, whether intentionally or not, the right to respect for the private life and correspondence of the accused persons. The Court observes that, in the present case, some of the conversations published in the press were of a strictly private nature. They concerned the relationships of the applicant and his wife with a lawyer, a former colleague, a political supporter and the wife of Mr Berlusconi. Their content had little or no connection at all with the criminal charges brought against the applicant. This is not disputed by the Govemment. In the opinion of the Court, their publication by the press did not correspond to a pressing social need. Therefore, the interference with the applicant's rights under Article 8 § 1 of the Convention was not proportionate to the legitimate aims which could have been pursued and was consequently not "necessary in a democratic society" within the meaning of the second paragraph of this provision.'

In *Peck v The U.K.,* the applicant was suffering from depression as a result of personal and family circumstances. On 20 August 1995 at 11.30 p.m. he walked alone down the high street towards a central junction in the centre of Brentwood with a kitchen knife in his hand and he attempted to commit suicide by cutting his wrists. He stopped at the junction and leaned over a railing facing the traffic with the knife in his hand. He was unaware that a close circuit television camera, mounted on the traffic island in front of the junction, filmed his movements. The relevant footage later disclosed did not show the applicant cutting his wrists, the operator being solely alerted to an individual in possession of a knife at the junction. The police were notified by the operator of the TV camera and arrived. They took the knife from the appli-

[21] *Craxi* v. *Italy (no. 2),* no. 25337/94, 17 July 2003.

cant, gave him medical assistance and brought him to the police station. Thereafter he was detained in the mental hospital. This incident in a public street was recorded by the closed-circuit television and that footage was shown on television on several occasions. The Court referred to the concept of private life as not being susceptible to exhaustive definition and reiterated its jurisprudence according to which:

> 'There are a number of elements relevant to a consideration of whether a person's private life is concerned in measures effected outside a person's home or private premises. Since there are occasions when people knowingly or intentionally involve themselves in activities which are or may be recorded or reported in a public manner, a person's reasonable expectations as to privacy may be a significant, although not necessarily conclusive, factor. A person who walks down the street will, inevitably, be visible to any member of the public who is also present. Monitoring by technological means of the same public scene (for example, a security guard viewing through closed-circuit television) is of a similar character. Private life considerations may arise, however, once any systematic or permanent record comes into existence of such material from the public domain.'[22]

The Court found that there was a violation of the right to respect for private life of the applicant because of the disclosures of the incident through the television which were considered by the Court as disproportionate and therefore unjustified interference with his private life and a violation of Article 8 of the Convention.

In the case of *Hachette Filipacchi Associés v. France*[23] the family of the deceased Prefect of Corsica who was murdered and his body was briefly lying on a street complained that the well-known magazine Paris Match published a photograph of this scene without their consent. The owners of the periodical in question were ordered by the domestic courts to publish, in their next edition of their magazine, that the publication of the photograph in question was made without the consent of the family of the deceased Prefect and that the members of his family consider that such publication infringes on their private life. The publishers of the magazine in question complained before the European Court that the order in question was an interference with their freedom of expression. The Court, by five votes to two, found that there has not been a breach of the right to freedom of speech and that the domestic courts were right in finding that there has been an unjustified interference with the right to

[22] In *P.G. and J.H.* (§ 57)

[23] Appl. No 71111/01, 14 June, 2007

private life of the relatives of the Prefect. I was one of the judges who disagreed with this judgment, having myself found that the effect of the photograph was to convey the full tragic impact of the assassination, correctly described in the accompanying comment as "a tragic page in [French] history" and that it is in the general interest for such events to be publicised, in full detail, so that the public is informed about matters that affect society and the country as a whole. The public also has a right to receive that information, a right that cannot be overridden by the subjective feelings of the victim's family.

I have given you examples of the case-law of the ECHR, where, in balancing the competing rights of freedom of expression and the right of privacy, the latter prevailed. Apart for one case with somewhat exceptional aspects, I could not trace any other cases where, after balancing the two rights in question, freedom of expression was found to succeed. The case is the one[24] in which the widow and sons of President Mitterrand were able to restrain dissemination of a book *Le Grand Secret* by his doctor, which revealed that the late President knew he was suffering from cancer in 1981, a secret he kept from his family as well as the public. The Paris Court of Appeal rejected a freedom of expression argument holding that it was necessary to stop the distribution of the book which was infringing the right to privacy of the deceased and his heirs. The ECHR held that the permanent injunction could not be justified. According to the judgment: 'The Court considers that the book was published in the context of a wide-ranging debate in France on a matter of public interest, in particular the public's right to be informed about any serious illnesses suffered by the head of State, and the question whether a person who knew that he was seriously ill was fit to hold the highest national office. Furthermore, the secrecy which President Mitterrand imposed, according to the book, with regard to his condition and its development, from the moment he became ill and at least until the point when the public was informed (more than ten years later), raised the public-interest issue of the transparency of political life.[25] However the Court found that the interim injunction, in contrast with a permanent one, was justified because the distribution of the book was very close to the death of President Mitterrand, a fact which could have deepened the family's grief.

In any event, freedom of expression may understandably be upheld, for example in respect of disclosures about public officials or candidates for political posts regard-

[24] *Plon* v. *France* 18 May 2004.

[25] Par.44.

ing their disreputable behaviour in their private life; even the marital infidelities of people holding sensitive posts in the State may be disclosed if they create commitments or situations incompatible with the functions of their office, while details of a politician's health may also be disclosed because there is a reasonable assumption that it may affect the execution of his duties; secret sources of income may as well be published. There is no doubt that freedom of expression, especially by the mass media, is the cornerstone of democracy and the most effective way to control those in a position of power. However, we should not lose sight of the fact that the mass media are nowadays to a great extent commercial enterprises with uncontrolled and virtually unlimited strength, interested more in profitable, flashy news than in disseminating proper information to the public, in controlling government abuse or in fulfilling other idealistic objectives. And although they may be achieving such objectives incidentally, accidentally or even deliberately, they should be subject to certain restraint out of respect for the truth and for the dignity and private life of individuals. Like any power, the mass media cannot be accountable only to themselves. A contrary position would lead to arbitrariness and impunity, which undermine democracy itself.

BOOK REVIEWS

Free in Captivity. The Ayios Kassianos Women's March in 1989
[Ελεύθερες στην Αιχμαλωσία. Η Πορεία των Γυναικών στον Άγιο Κασσιανό το 1989]

Tasos Hadjianastasiou
Koukkida Publications
Athens, 2024 (pp. 350)
ISBN: 978-618-208-133-4

This is an important book that captures the essence of a specific historical event, the 1989 women's march in Ayios Kassianos. The book was published, following an initiative by the Pancyprian Women Walk Home Movement, which provided significant support to the author to carry out his research, including access to the Movement's archive, as well as to several of the primary actors, who participated in the women's march, and who provided their testimony, either orally or in writing. Access to such a wide array of unpublished primary sources is of particular significance, and complements the exhaustive research undertaken by the author in contemporary news reports, as well as secondary bibliography.

This is the first comprehensive study for this significant moment in modern Cypriot history, and therefore is essential reading for anyone interested in the Cypriot Women Movement, and, more generally, in the historical efforts of Cypriot civil society to mobilise against the Turkish occupation of Cyprus. Indeed, the author places the Cypriot Women Movement within a wider research framework covering the Cypriot anti-occupation movement, and within the broader context of resistance movements throughout the globe. The women's march, in July 1989, was undeniably one of the significant moments of this overall anti-occupation movement; their attempt to enter the area of Ayios Kassianos was met with expected resistance from UN peacekeepers, but the march continued, ignoring warnings, and the women were, eventually, left alone to face the armed Turkish occupation forces. In total, 111 people, 98 women, ten men, and three foreign correspondents, were apprehended. Among the men was the current Archbishop of the Orthodox Church of Cyprus, who was then an Archimandrite.

The book consists of ten main chapters. The first chapter explains how the wid-

er anti-occupation movement in Cyprus is part of the typology of mass movements, since it consists of collective actions taken in order to promote common interests or defend a common objective outside formal institutions. As already indicated, the author places the Cypriot women's movement as an indispensable part of the Cypriot anti-occupation movement, and in the second chapter provides a detailed account of the evolution of the anti-occupation marches of women since the Turkish invasion of 1974. The first, highly successful, march took place in April 1975, just nine months after the Turkish invasion. 12 years later, in 1987, the women movement intensified its efforts, with a march in Aronas and subsequently, a major march in Ayios Pavlos, and then, in June 1988, with the symbolic occupation of the Acropolis in Athens, and the marches in Lympia and Achna, in 1989.

The subsequent chapters chronicle the march to Ayios Kassianos. Chapter 3 analyses the preparation of the march, whereas chapters 4 and 5 assess the first two days of the march. Chapter 6 elaborates on the captivity of the arrested persons in prison dungeons. The most important contribution of Chapters 3-6 is the extensive use of written and oral testimonies, which undeniably enriches our understanding of the conditions of detention of the captives, the efforts to intimidate and torture them, and the resistance they showed, until the last captives were finally freed, nearly ten days later. Chapters 7-9 assess the outside impact of the march and the subsequent captivity; the massive demonstrations and mobilisations for support in Cyprus, Greece and abroad, the stance of Turkey and of the Turkish Cypriot leadership, international reactions, as well as the position taken by the Vasiliou Government. Chapter 10 and the epilogue of the book, briefly refer to what happened after the Ayios Kassianos incident and how this shaped the history of Cyprus.

The author praises the decisiveness and fighting spirit of the women, and it is difficult to dispute that Ayios Kassianos was a milestone in the efforts for resistance of the Cypriot civil society against the ongoing occupation of their lands, and the deprivation of their right to return to their homes. The women's mobilisations were instrumental in establishing the importance of Greek Cypriot women as active political actors, demanding their rights, irrespective of responses by the State. Rather than passive agents, the grieving mothers, sisters, wives and daughters of the missing persons, the widows and the refugees, were understood as actors with primary agency, seeking liberty and justice.

The author is correct that the Cypriot Government, and part of the Cypriot establishment, felt that the Ayios Kassianos march would prove detrimental to efforts to solve the Cyprus Problem, since they would lead to further securitisation. This is a re-

curring theme in Cypriot history: the fear that if civil society were to demonstrate too loudly against the occupation, then it would be more difficult to negotiate a solution. In practice, however, efforts to solve the Cyprus Problem have ended up being unsuccessful, irrespective of the intensity of civil society responses. The author is critical of the Government's stance, and implies that the lack of support of resistance from civil society has effectively contributed to the deterioration of the Greek Cypriot position in the negotiations. This is partly accurate, but it should be noted that the Cyprus Problem has been a frozen conflict for decades and no rational government would actively encourage turbulence in the island, especially considering the military power that Turkey already employs in the occupied part of the island.

There is a lingering question that the author has, presumably intentionally, not attempted to address in his, otherwise excellent, narration. Why did the women's movement cease after the Ayios Kassianos incident? As the author painstakingly argues, this was the peak of the civil society anti-occupation movement, that caused wide mobilisations, both in Cyprus and abroad. However, whereas the women's movement had intensified its collective action between 1987-1989, there was no effective follow-up after Ayios Kassianos. One can but wonder whether the division of the women's movement eventually led to its own failure, and whether the forcefulness of the Ayios Kassianos incident was effectively such, that it self-sabotaged any attempt for the movement to continue. The author briefly referred, in Chapter 2, to the division amongst the initial women movement, that led to two distinct women movements; the first movement encouraged reconciliation with Turkish Cypriots and discouraged the use of national symbols and instructed orderly action, whereas the second movement supported more forceful action and considered that reconciliation was only possible with the withdrawal of Turkish troops. The second movement was the one responsible for the Ayios Kassianos march, which led to their arrest. Ayios Kassianos proved to be both the peak, and the end of the road. Its history has been written, but the next step is to assess it within the framework of the wider impact of the internal conflict and division of the movement, and of the prevalence of the more hardline school of thought. There was no room after the Ayios Kassianos incident for the first movement to function, but, effectively, there was also no room for the second movement to continue its operations either.

Achilles C. Emilianides

Accumulation of Treasures, Material Gleaned from Cypriot Newspapers (1879-1962) [Θησαύρισμα - Στοιχεία Σταχυολογηθέντα από Κυπριακές Εφημερίδες (1879-1962)]

Nicos Panayiotou
Private publication
Nicosia, 2025, v. I-II, (pp. 1437)
ISBN: 978-9925-7371-5-4

The new book by Nicos Panayiotou is an important contribution to our knowledge of the period of the British Rule, in general, and the function of the press during this period, in particular. Journalistic writing has, generally speaking, an ephemeral character. However, whereas nowadays a journalist or columnist might have a justified suspicion that their work might be relevant to a researcher in the future, this was not the case or the expectation during the period of British rule, as a columnist could not even imagine that someone might one day, nearly a century later, take an interest in his contribution. At the end of the 19th and the beginning of the 20th century, in a time when newspapers were not preserved in electronic online archives and were often destroyed either because of paper deterioration or simply through neglect, columnists wrote without ever suspecting that, a century later, people would still be reading their texts, and would use their essays in order to recapture the reality of a previous era.

This is a remarkable book, as it focuses on the recording of moments, in finding fellow travelers who share the author's love for unknown or forgotten historical snapshots or even linguistic expressions. The author does not attempt to write the history of the British Rule, nor to emphasise the famous and celebrated events, or the names that might still attract attention today for various reasons. He is more interested in the overlooked details and sidenotes, in those elements that depict for the reader the atmosphere of an era, both in terms of events and language usage. The author makes extensive use of excerpts from newspapers of the British colonial period because he wants the reader to experience not only the news itself, but also the 'flavour' of the publication.

Work of this kind unavoidably depends on delicate balances, and it would have

been easy for the effort to prove unsuccessful, especially given that it covers the study of newspapers over a span of more than 80 years. Panayiotou's writing experience, and particularly his experience as a bibliographer, has helped him carry out this titanic task. The primary aim of a bibliographer is to assist the researcher in locating material related to the subject under study through the organised classification of material, allowing the researcher to use himself the necessary sources, if needed. This work, of course, is not a bibliography, since it includes not merely a listing of material but also commentary and synthesis aimed at portraying the period. Yet it shares a similar objective: to guide the reader toward deeper research and to inspire further inquiry. The author's intention is to provide stimuli for a new journey. Indeed, a young researcher might well discover, through reading this book, the topic of a dissertation, while a more experienced scholar might find inspiration for a specialised academic paper.

The author has been comprehensive in the range of subjects he has recorded, and as a result the book contains something for everyone. It need not be read sequentially; its nature as a treasury allows readers to choose the subjects that interest them most and begin there. If one is interested in music or the theatre, education or local communities, one can start there and then move on to the rest. The author includes extensive analyses of newspaper articles (with national aspirations naturally occupying center stage), chronicles, news reports, feuilletons, literary material, and even topics often overlooked despite being central to understanding the era, such as correspondence and advertisements. And, of course, he provides thorough accounts of the individuals mentioned, along with detailed indexes.

Through seemingly insignificant news items and letters, we are prompted to reflect on the social customs of the time. I must confess that I was deeply affected by the repulsive report in *Salpinx* from 1885, which urged people not to burn children born deformed, 'as you always do,' among other reasons because 'when they die, they are embalmed and purchased at very high prices by travelers and various other naturalists.' The author rightly comments that this reflected the mentality of an era and that humanity traveled a thorny path before reaching even a minimally humane treatment of fellow human beings. Similarly, I found particularly interesting the 1931 advertisement in *Neos Kypriakos Fylax* for Philips radios, which were presented as the perfect expression of civilization, reproducing sound with complete naturalness and creating the impression that the artist performed inside your living room. I lived through the era when we listened to sound from old radios, as well as the arrival of

private radio broadcasting. A century after this advertisement was originally written, technological progress led to similar advertising expressions for virtual reality glasses.

Regarding court cases, the author gives extensive attention to some of the most turbulent trials of the British colonial era, such as the appeal by the governor of Famagusta, Arthur Young, who had been a parliamentary candidate in the elections of 1891 and succeeded in annulling the election of Achilleas Liasidis and Giorgos Siakallis, on the grounds that they had exercised widespread voter intimidation with the support of the Orthodox Church. Or the trial of the 'White Politician' Ioannis Kyriakidis. Equally important is the reference to the 1923 trial of Leonidas Pavlidis, editor of *Imerisios Kiryx*, and the citation of an excerpt from the speech of Antonis Triantafyllidis, who, as the author rightly observes, masterfully declared that:

'It is not the duty of the press to make a Government popular. Journalists are not salaried preachers of the Government. A journalist, as a member of a people, has the duty to represent and interpret public opinion. In this case, the Minister's reply offended the dignity of a people, and it was the duty of the press to express the people's indignation. The Minister has an interest in learning the impression made by his words. And those who seek to silence the press render a poor service to the Government, which, remaining ignorant of the people's thoughts and sentiments, will wander astray and fall into error'.

In sum, this is a valuable book, which contributes to the wider historical understanding of the British Era.

Achilles C. Emilianides

The British Sovereign Base Areas in the Island of Cyprus

Kyriacos Demetriades
Cambridge Scholars Publishing
Newcastle upon Tyne, 2026 (pp. 328)
ISBN: 978-1-0364-6124-9

According to Article 1 of the Treaty of Establishment, the areas referred to as the Akrotiri SBA (British Sovereign Base Area) and the Dhekelia SBA, which cover 250 sq. km of the land area of the island of Cyprus, remained 'under the sovereignty of the UK', whereas a further 13 sites were retained for use by the UK. According to Article 2 section 2 of the Treaty of Establishment, the Republic of Cyprus shall cooperate fully with the UK in order to ensure the security and effective operation of such military SBAs, as well as the full enjoyment, by the UK, of the rights conferred by the Treaty. The legal status of the SBAs was confirmed by Protocol 3 of the Treaty of Accession of the Republic of Cyprus in the EU, where it was provided that the Treaty shall not apply to the SBAs of Akrotiri and Dhekelia, except to the extent necessary to ensure the implementation of the agreements set out in the Protocol. Following Brexit, a new Protocol was attached to the Withdrawal Agreement of the UK, in order to safeguard that the arrangements applicable to the relations between the EU and the SBAs after Brexit would continue to be defined within the framework of the Republic of Cyprus' EU membership; the SBAs thus remained part of the customs territory of the EU after Brexit.

The continuation and extent of the UK's sovereignty over the SBAs has been a topic of controversy, and, in recent years, the issue has received significant attention, due to the SBAs allegedly having been used to host US forces, or to provide support to Israel in its conflict with Hamas in Gaza. The UK Government has insisted that the SBAs were strictly used to support the delivery of humanitarian aid to Gaza, but not to provide military equipment. However, the SBAs have been used as a base for UK air strikes against the Houthis in Yemen. Following the launch of military strikes by the USA and Israel against Iran, and the launch of counterstrikes by Iran in the region, a drone hit the Akrotiri SBA. While confirmed that Iran was not responsible for the strike, the involvement of Cyprus, as a geographical area, in the dispute, and

the failure of the UK to inform the Cypriot residents about the strike, prompted political upheaval. The President of the Republic of Cyprus announced his intention to discuss the SBAs future with the UK, and this was recorded in the conclusions of the European Council during the Cypriot Presidency in April 2026. Whereas the UK was adamant that it was not willing to negotiate the status of the SBAs, other issues were left open for negotiation between the parties.

Within this framework, the recent book by Kyriacos Demetriades on the SBAs is both timely and an important contribution to the literature. Demetriades has extensive practical experience on the topic, having served, for many years, as the Area Officer in the Dhekelia SBA, with responsibility for civil administration. His book is comprehensive and covers many aspects of the SBAs which have not been discussed elsewhere, in detail. The book includes a historical introduction and a chapter on the negotiations leading to the establishment of the SBAs, a topic which I covered in my 2021 book.[1] The background is very helpful for the reader, but the most important contribution of the book are Chapters V-VIII.

Chapter V details the development of Dhekelia during the 1950s. It shows how the Dhekelia area was ready, in 1960, to evolve into the Dhekelia SBA, due, in a significant part, to good planning on behalf of the British Administration. The historical evolution of Dhekelia, with the development of the quarantine station, the prison, the camps established for Jewish immigrants near Dhekelia, the Maltese Camp, the rifle rangers, the desalination plant, and the military hospital, are only parts of this analysis, which is of particular importance.

Chapter VI elaborates on the SBAs administration, its features and fundamental principles. Most of the studies regarding the SBAs focus on their legal status, the political controversies, and their interrelation with the Republic of Cyprus, but few focus on the internal workings of the SBAs, and therefore this contribution is a welcome addition. With the background of the 1960 Declaration on the Administration of the SBAs, Demetriades examines the British views on the legal status of the Declaration itself, the enactment of Orders in Council and Ordinances, and the development of the domestic law of the SBAs. As the author acknowledges, for the SBAs to function effectively, they must enjoy an intertwined relationship with the Republic of Cyprus, and this is one of their peculiarities that should not be underestimated. There are Cypriot citizens who reside and work permanently in the SBAs, and despite their mil-

[1] A.C. Emilianides, *A Long-Drawn-Out Game of Chess: The Secret Negotiations about the British Bases (1959-1960)*, Nicosia: Hippasus, 2021 (in Greek).

itary character, the SBAs are dependent upon the Republic of Cyprus for their proper function. They are not, and cannot realistically be, isolated within the island, and this is why it was not practical to differentiate between a drone strike against the SBAs and a strike against the Republic of Cyprus; whereas technically and legally there might be a differentiation, in practical terms differentiation is almost impossible.

The focus on the SBAs Administration is complemented by the analysis included in Chapter VII, which refers specifically to the personnel of the SBAs, its functions and peculiarities. The SBAs may be military bases, but they have civil administration and personnel, an Attorney-General, judges and administrative officers. The implementation of Non-Military Development in the SBAs, as agreed with the Republic of Cyprus, is also discussed in this chapter. Furthermore, Chapter VIII offers a brief assessment of the historical-cultural heritage of areas located in the SBAs, with emphasis on selected places, such as Ormidhia, Xylotymbou, Xylophagou, and the Akrotiri Peninsula.

Overall, the book offers helpful insights on the internal workings of the SBAs, specifically aspects of their civil administration and historical evolution. The author has utilised a wide range of materials, both primary and secondary sources, as well as his own experience, to provide a comprehensive overview of critical elements of this under-researched topic. Anyone interested in the evolution and actual reality of the SBAs should find a wealth of information which, hopefully, can lead to additional research on specialised issues related to the SBAs in the future.

Achilles C. Emilianides

Horizontal Effect of Fundamental Rights and its Reception in the Cypriot Legal Order [Η Οριζόντια Επίδραση των Θεμελιωδών Δικαιωμάτων και η Υποδοχή της στην Κυπριακή Έννομη Τάξη]

Petros Constantinides
Sakkoulas Publications
Athens-Thessaloniki, 2025 (pp. 491)
ISBN: 978-618-247-088-6

Drittwirkung of fundamental rights is one of the most interesting aspects of public law, as it effectively interrelates with private law. Petros Constantinides had originally published on this topic a decade ago,[1] but his new monograph, which hails from his PhD, significantly expands our horizons on this complex issue. Drittwirkung, in its more traditional sense, enables an individual plaintiff to rely directly on the constitutional provisions safeguarding fundamental rights in order to sue another individual or the government for the violation of those rights. Whereas, the traditional legal view is that constitutional rights protect the individual from State interference, rather than from interference by other private individuals, the expansion of the notion of 'positive obligations' of the State has indirectly allowed actions against the State also from interference by private individuals. Drittwirkung, however, allows the plaintiff to directly sue the private individuals, rather than the State.

In the Cypriot legal order, the foundation for the reception of drittwirkung is the judgment of the Supreme Court of Cyprus in *Yiallouros v. Nicolaou*, which impliedly adopted drittwirking.[2] Yiallouros, the Director of the Sewerage Board of Nicosia, had, during a whole year, taped and monitored the telephone conversations of Nicolaou, an employee of the Sewerage Board; in addition, Nicolaou was put on probation in light of the content of his telephone conversations. Nicolaou sued Yiallouros before the District Court, requesting compensation for the violation of his constitutional rights, namely his right to privacy safeguarded by Article 15 of the Constitution and secrecy of communications safeguarded by Article 17. The District Court held that

[1] P. Constantinides, *The Drittwirkung of Fundamental Rights in the Cypriot Legal Order*, Hippasus 2016 (in Greek).

[2] [2001] 1 CLR 558 (in Greek).

Yiallouros was liable to pay compensation to his employee. Yiallouros appealed, arguing that there is no direct right to sue for violation of constitutional rights, when such violations do not also constitute torts in accordance with the Torts Law, Cap. 148. The Supreme Court held that the violation of constitutional rights vests to the victim the right to invoke legal remedies for the vindication of his rights. The effective application of fundamental rights presupposes the right of a person whose constitutional rights have been violated to receive compensation for any damage suffered because of the violation. Consequently, the violation of constitutional rights was held to be directly actionable before District Courts.

It should be noted that, in *Yiallouros*, the Supreme Court did not explicitly use the term drittwirkung (or horizontal application), and did not expand on the theoretical foundations, or practical distinctions, of the use of its principles. It restricted itself in acknowledging that the right derived directly from the Constitution, and that there should be an effective remedy, without setting any comprehensive criteria. Savvas Papasavvas had correctly noted that the application of the principles in the judgment was confusing, specifically since the Court opted to rely on the common law principles that there should be a remedy when there is a wrong.[3] The Supreme Court had, however, referred to the principle of drittwirkung earlier, albeit with no detailed discussion, when it observed that drittwirkung could be characterised as the impact of fundamental rights to the creation of legal principles in the entire field of law, and that fundamental rights should be applied erga omnes.[4] There is little doubt that the Cypriot case-law provides, in theory, for wider protection of the drittwirkung of fundamental rights compared to other European countries. However, this remains mostly theoretical, due to the dearth of concrete cases where the principle has been applied in favour of the individual. Indeed, very few relevant cases even cite drittwirkung, and none attempts to formulate legal principles for its implementation.

Constantinides' book addresses this theoretical vacuum. Part I comprehensively discusses the foundations of drittwirkung, elaborating on the grounds which might justify the horizontal effect of fundamental rights in the field of private law. The author critically approaches the theory that domination should be the principal distinguishing characteristic, and relies on the very idea of liberty as the main ground justifying drittwirkung. This is a fundamentally liberal approach, which is analysed

[3] S. Papasavvas, *There Is Drittwirkung of Human Rights in the Cypriot Legal Order: Comments and Thoughts on the Judgment of the Supreme Court in Civil Appeal 9931 Yiallouros v. Nicolaou*, 15 *Dikaiomata tou Anthropou* 837–861 (2002) (in Greek).

[4] *President of the Republic* v. *House of Representatives* [1997] 3 CLR 36 (in Greek).

within a comparative context. The analysis of horizontal effect within a comparative framework that is not restricted in continental law, but also expands to common law, EU law and the ECHR, is an innovation of the book, and is consistent with the inherently mixed nature of the Cypriot legal system. In Part II, the author specifically addresses horizontal effect in the Cypriot legal order, focusing also on specific examples, in the field of the right to contract, freedom of expression and employment relations. The book is complemented by a detailed list of bibliography and case-law.

The author's attempt to set out specific criteria for the implementation of the principle of horizontal effect and to examine their application in the Cypriot legal order, is particularly useful. This is, undeniably, the most comprehensive work on drittwirkung in the Cypriot legal order and should prove valuable to the further study on this significantly under-researched topic.

Achilles C. Emilianides

Cyprus and the First World War

Hubert Faustmann & Georgios Kazamias (eds)
Bloomsbury
London, 2025 (pp. 223)
ISBN: 978-1-3504-4697-7

Published research on Cyprus and the First World War is extremely limited, in any language, and therefore, the publication of an English-speaking edited volume by a respected publishing house would unavoidably imply that this is essential reading for anyone interested in the period. However, restricting the contribution of this edited volume to the self-evident fact that it covers a vacuum, would be a disservice to its quality and originality. This would have been an important contribution and essential reading, even if the topic had not been so under-researched. All papers are of high academic quality, and, as a whole, the volume presents a comprehensive perspective of the era, and the critical aspects of Cyprus' participation in the Great War.

The Ottoman Empire assigned to Great Britain the rights of possession and administration of Cyprus by signing the Convention of Defence Alliance on 4 June 1878 in Constantinople. With the signing of the Treaty of 1878, the period of British rule in Cyprus officially began. Great Britain was granted the power to enact laws and conventions for the island of Cyprus and to regulate its commercial and consular relations and affairs. An Annex to the Treaty of 1878 provided that such Treaty would be terminated and Cyprus would be restored to the Ottoman Empire, so long as Russia restored Kars, Ardahan and Batoum to the former. Thus, while, in law, Cyprus still theoretically belonged to the Ottoman Empire and could eventually return on the fulfilment of certain conditions, in effect, the real and only sovereign of Cyprus was Great Britain, which exercised the administration of the island. The Anglo-Turkish Mixed Arbitral Tribunal held in Parounak, that the position of Cyprus during that period was analogous to that of a protectorate in the sense that it fell within the designation of a country under the protection of Great Britain, which in fact governed the island. This peculiar situation ceased in 1914, when, immediately after the outbreak of the First World War, Great Britain annexed Cyprus by Order in Council of 5 November 1914.

The editors correctly acknowledge that this shift alone, from Ottoman suzerainty to colonial status, would imply that the First World War was transformative for the

history of the island. Prior to this shift, the British counterargued that enosis, the union of the island with Greece, could not take place because Britain was not the legal owner, but only the possessor. The return of the Ottomans was therefore an implied threat. The unilateral annexation of Cyprus by Britain also meant that Greek Cypriots could demand enosis in a more structured and insistent manner. Soon thereafter, the opportunity for enosis arose in October 1915, when Britain offered to cede the possession of Cyprus on condition that Greece give immediate and full military support to Serbia. This proposal was independent from any other land that Greece could potentially secure at the end of the war, if this turned out to be successful for the Allied Powers, as evidently it did. The Zaimis Government rejected the offer, as Greece decided that it was not yet ready to abandon its neutrality stance.

The British offer is discussed by George T. Mavrogordatos, who describes it as the best ever, if not the only, opportunity for enosis. Mavrogordatos harshly criticises Constantine and his supporters, even calling him as the 'man who had actually buried Enosis'. He argues that whereas support for Constantine had a rational basis for the Old Greece, it was completely irrational in the case of Greek Cypriots, who overwhelmingly continued to support Constantine, rather than Venizelos. Mavrogordatos, being pro-Venizelos, further argues against the widely held view that Venizelos failed to raise the Cyprus claim at the subsequent Peace Conference, by contextualising it within Venizelos' wider political agenda; he argues that Venizelos could not risk offending British sensibilities or jeopardise British support, and that Cyprus simply did not have the urgency or magnitude required for it to be a top priority. This is a fair assessment; however, Mavrogordatos is less willing to similarly contextualise the stance of the anti-Venizelos supporters for neutrality at a time when it was not yet clear who would actually win the Great War.

Contrary to Mavrogordatos, the late Heinz A. Richter is much more critical than Venizelos in his chapter regarding Cyprus and the Peace negotiations in Paris in 1918. His conclusion is that the Cypriot effort to achieve enosis at the Paris Peace Conference failed, because they could not participate and had to rely on Venizelos, who was not interested in Cyprus as he had other priorities at the time, namely the dreams for Megali Idea. Admittedly, even if the Cypriot delegation had had the chance to participate, I doubt this would have achieved anything without Venizelos support; the stakes were simply too high. Richter blames Venizelos for being too idealistic and optimistic and for placing too much trust to the British that they would keep their promises. The two conflicting portraits of Venizelos, as presented by Mavrogordatos on the

one hand, and Richter on the other, are indicative of why he remains such a leading and divisive figure, more than a century after the events described in the book.

I have been discussing Venizelos more than Cyprus so far, but it is, of course, Cyprus that is at the core of the volume. The perception of the Greek Cypriot press to the Great War is analysed in detail by Andreas Karyos. The Greek Cypriot press appreciated, from the outset, that the War would result to enormous territorial changes, and would be catastrophic for everyone concerned. Once Britain had joined the War, the Greek Cypriot press sided with the Entente, and against the Ottoman Empire and the *opportunist and arrogant* Austria-Hungary and their allies. The development of the Great War was perceived as an opportunity for Greek Cypriots to secure their own political aspirations, and Karyos emphasises how the Paris Peace Conference was viewed as the culmination of this effort, the great moment when the victorious powers would consent to enosis.

Robert Holland offers a balanced overall account of Cyprus' participation in the Great War, noting that normalcy continued to prevail in the general life of the island, mainly due to Cyprus' not having any large-scale military contribution to the war. It is perhaps striking that, despite Cyprus' oft-discussed geographical position, it had not been a focus of military operations in many wars. As Holland observes, whereas Cyprus indeed participated in the Great War, the war itself was not great in the island. Cyprus living conditions mostly remained the same, whereas a practical wartime reality was that Cyprus was even more cut off from the outside world, although this probably had more effect on the isolation of British government officials, who could not take leave in the UK, rather than for Cypriots themselves.

Anastasia Yiangou tackles the Orthodox Church during this period. In 1912, following the end of the lengthy rivalry for the election of a new Archbishop, the Archbishop was placed as chair of the newly established Political Organisation, which adopted a more forceful enosis policy. 1914 was the 50th anniversary of the British cession of the Ionian island to Greece, in 1864, and this led to the organisation of the first enosis plebiscite. The annexation of Cyprus by the British, and, as the author correctly acknowledges, the British offer of Cyprus to Greece marked a significant psychological shift in the Church's attitudes towards Britain; enosis had become an imminent objective. Venizelos close associate, Meletios Metaxakis, the Bishop of Kition, lost the 1916 archiepiscopal elections, with the new Archbishop Cyrill III, a moderate, leading the expected forceful demand that the enosis aspirations be realised without further delay. The hardening of attitudes, as the author acknowledges, both within the Orthodox Church and within the colonial authority, was, therefore,

a reality since 1917, and in many ways the experience of the First World War had helped trigger this change in the relationship between the Church and the British.

Altay Nevzat covers an important topic, which is the viewpoint of the island's Turkish Muslim community during the period. The chapter is of particular importance, since it challenges the prevailing, admittedly simplistic, approach, that excludes them from the main narrative which, from a political point of view, unavoidably focuses on the Greek Cypriots, the Church, and the demand for enosis. The change of ruler in the island was particularly challenging for the Turks of Cyprus, since the legal ruler was, until then, the Ottoman motherland. The dominant viewpoint, deriving from the British colonial administrators themselves, refers to the Cypriot Turks as subdued and loyal, or even happy, to disassociate from the Ottoman Empire. However, Nevzat acknowledges that the situation was much more complex, as the Turks of Cyprus were not unreservedly loyal to Britain, with clashes of identity, both within the community, and vis-à-vis the British.

National aspirations aside, the issue that dominated political discussions during the early period of the British Rule, was the high taxation, particularly so that the British could pay financial compensation to the Ottoman Porte -the Tribute- as consideration for the right to administer the island. Diana Markides has, fairly recently, authored a monograph on the Cyprus Tribute, and her chapter in this volume succinctly elaborates on this important topic. The Great War undermined the legitimacy of the Tribute, and following the annexation of Cyprus by Britain, it seemed like the perfect opportunity for Cypriots to achieve a termination of this debt charge. However, this did not happen. Markides explains how the two major communities of the island differed on their perspectives; whereas, the Greeks of the island passionately requested the termination of the Tribute, the Muslims were reluctant to request the abolition of a Tribute to their motherland. The British administration could take advantage of the growing polarisation amongst the two communities, and the lack of concerted action on this issue, so as to avoid terminating the Tribute. The emphasis of Greek Cypriots on their political aspirations for enosis, and their optimism that Britain would fulfil such aspirations, led them to restrict the tenacity of their demands in other issues.

This is a general theme that is also acknowledged in Kyriacos Iacovides' illustration of the social and economic impact of the Great War. Iacovides notes that the hardship the Cypriots faced did not lead to social turmoil during the period. This was partly due to enlistment for service, which decreased unemployment and offered some monetary relief, but the main reason was the wish of Greek Cypriots to avoid criticising the British, at a time when they were expecting them to act as the liberators

and fulfil the aspiration of enosis. In addition to Iacovides' work, which refers to the economic and difficult social realities of the island, Maria Panayiotou provides an overview of the impact of the Great War on imports and exports, and how this affected the gradual transformation of the Cypriot economy into a non-self-supporting economy that imported more foot than it exported.

Georgios Kazamias assesses the most numerous military contribution of Cyprus to the Great War, the Cyprus Mule Corps. Near 10.000 or 11.000 individuals served as muleteers in the Macedonian Front, the vast majority being Greek Cypriots, motivated by the attractive salary, as well as the prospect of being sent to the Greek lands. Kazamias is successful in the difficult task of reconstructing the daily lives of the muleteers, despite the lack of written memoirs, by carefully utilising all available sources. As he notes, enlisting proved to be an optimal choice during the war years, as it required the assumption of only a limited war risk, and replaced emigration as an economically viable option. Kazamias' analysis of the life and death in the Corps, including of issues such as formation and deployment, questions of nationality and identity, health and illness, and even naming convention, makes this one of the volume's more important contributions.

Antigone Heraclidou focuses on the significant non-military contribution of the island during the period. Cyprus became a suitable place for the recuperation of injured soldiers and the detention of prisoners of war. Heraclidou details the establishment of military convalescent homes in Mount Troodos, the construction of a vast camp for prisoners of war in Famagusta, the accommodation of Russian refugees, and the presence of Armenian volunteers. The significance of Heraclidou's contribution is that it convincingly credits Cyprus' non-military contribution during the Great War, a topic which had so far received relatively little emphasis.

The volume's last chapter is by Hubert Faustmann and it includes wider thoughts on the impact and legacy of the First World War on Cyprus. It is fittingly the last chapter, as it effectively argues for a re-interpretation of the importance of the Great War on Cyprus, and implicitly makes the case for the significance of the edited volume itself. Faustmann is correct in arguing that the impact of the Great War and its legacy on the British colonial rule of Cyprus is considerably more significant than commonly acknowledged. The change of status of the island, the beginning of a new, very different phase, for the enosis movement, and the change in the dynamics both between the two communities and their relationship with the British authorities, are lasting consequences that fundamentally altered the political and legal landscape and paved the way for the next chapter in the British Rule era. It is tradition to divide the era in

two parts, the first until 1914, and the second after 1914 and the annexation of the island by Britain. The edited volume comprehensively reassesses the period during which this fundamental shift took place, and will be the starting point for additional research work on the period.

Achilles C. Emilianides

De-ethnicising the Ethnography of Cyprus: Political and Social Conflict between Turkish Cypriots and Settlers from Turkey

Yael Navaro-Yashin
In Y. Papadakis, N. Peristianis, & G. Welz (Eds.)
Divided Cyprus: Modernity, History, and an Island in Conflict
Indiana University Press
Bloomington, 2006 (pp. 256)
ISBN:978-0253218513

Introduction

Yael Navaro-Yashin's chapter, *De-ethnicizing the Ethnography of Cyprus,*published in *Divided Cyprus: Modernity, History, and an Island in Conflict* (Indiana University Press, 2006), offers a significant intervention in the study of the Cyprus conflict. Moving beyond conventional analyses that frame the island's division as a binary ethnic struggle between Turkish Cypriots and Greek Cypriots, Navaro-Yashin challenges the assumption of internal homogeneity on the Turkish side.

This review argues that Navaro-Yashin's work makes a significant contribution to the analysis of the Cyprus issue in terms of internal social tensions and the reshaping of daily life, but that is flawed due to its one-sided ethnographic perspective and its failure to engage sufficiently with class and broader comparative frameworks.

Summary

Rather than reproducing the dominant narrative of interethnic conflict, Navaro-Yashin demonstrates that the primary tensions shaping Turkish Cypriot daily life emerge from within their own community. Specifically, she highlights the strained relationship between native Turkish Cypriots and settlers from mainland Turkey following the island's division in 1974.

Her ethnographic findings reveal that Turkish Cypriots' anxieties are less directed toward Greek Cypriots and more toward demographic change, cultural differences, and the pervasive influence of the Turkish military presence. The official discourse portraying settlers as 'compatriots' or 'saviours' is contradicted by everyday practices that construct clear social and cultural boundaries between the two groups.

Through interviews and observations, Navaro-Yashin shows how Turkish Cypriots perceive settlers as culturally different and, at times, socially disruptive. These perceptions are expressed in complaints about changing social norms, rising crime, and feelings of alienation within their own homeland. A striking example is the comparison made by an interlocutor who describes coexistence with Greek Cypriots as more predictable and secure than living alongside settlers -illustrating how ethnic categories lose coherence in lived experience.

The study also foregrounds the role of militarisation in shaping everyday life. The omnipresence of military symbols, restricted zones, and hierarchical authority structures produces a sense of constraint and subordination among Turkish Cypriots. This environment fosters what the author interprets as a form of learned helplessness, where individuals internalise their limited agency under military-political dominance.

Finally, Navaro-Yashin explores how cultural practices -such as language use, secular lifestyles, and gender norms- become markers of distinction and tools for identity preservation. These practices evolve into a political stance against perceived assimilation, encapsulated in expressions like 'Turks come and go' which reflect resistance to demographic and cultural transformation.

Critical Analysis

One of the key strengths of Navaro-Yashin's work lies in its methodological and theoretical innovation. By grounding the Cyprus issue in ethnographic detail, she shifts the analytical lens from macro-level geopolitics to micro-level social relations. This approach reveals how power operates through everyday practices, perceptions, and spatial arrangements.

The use of Pierre Bourdieu's concepts -particularly *habitus*, *cultural capital*, and *distinction*- is especially effective. Navaro-Yashin demonstrates how Turkish Cypriots construct cultural hierarchies that symbolically differentiate themselves from settlers. Everyday judgments about manners, domestic practices, or speech, become mechanisms through which social boundaries are reproduced. These forms of distinction function not merely as cultural preferences, but as strategies of identity preservation in a context where political autonomy is perceived as limited.

However, the study would benefit from a more explicit engagement with broader theoretical traditions, particularly postcolonial and settler-colonial studies. The dynamics described -settler influx, demographic engineering, cultural hierarchies, and perceived loss of autonomy- closely parallel patterns observed in other contexts.

Situating the Cyprus case within this comparative literature would strengthen the analytical depth of the argument and clarify its broader relevance.

In terms of limitations, the most significant issue is the one-sided nature of the ethnographic perspective. While the study powerfully captures Turkish Cypriot narratives of marginalisation, it does not sufficiently incorporate the voices of settlers themselves. Their motivations for migration, experiences of exclusion, and strategies of adaptation remain largely unexplored. This absence weakens the claim to ethnographic completeness and risks reproducing the very asymmetries the study seeks to critique.

Additionally, the analysis underplays the role of economic structures. Many settlers come from rural and economically disadvantaged backgrounds, working in precarious or seasonal jobs. As a result, the cultural distinctions drawn by Turkish Cypriots may also reflect class-based hierarchies rather than purely cultural differences. The derogatory labeling of settlers can thus be interpreted as a form of class prejudice, where ethnicity, nationality, and socioeconomic status intersect. A more explicit engagement with this intersectionality would significantly enrich the analysis.

Conclusion

Navaro-Yashin's chapter makes an important contribution to the sociology of conflict by demonstrating that ethnic divisions are neither fixed nor exhaustive explanations of social tension. By uncovering intra-community hierarchies and everyday forms of power, the study challenges dominant narratives of the Cyprus issue and opens new avenues for ethnographic and sociological inquiry.

At the same time, its analytical potential is constrained by a limited engagement with comparative theory, economic factors, and the perspectives of settlers themselves.

Ultimately, this study serves as an example of how ethnography can challenge simplified ethnic frameworks; however, its lasting contribution to sociology lies in paving the way for the interpretation of multifaceted narratives and the relationships between settler and indigenous populations in Cyprus within the context of social stratification and broader global patterns.

Parstuğ Coşkun
Ibn Haldun University

The Protection of the National Identity of the Member States and the Principle of Supremacy in EU Law [Η Προστασία της Εθνικής Ταυτότητας των Κρατών Μελών & Η Αρχή της Υπεροχής στο Ευρωπαϊκό Δίκαιο]

Theognosia Kouspi
Hippasus
Athens-Nicosia, 2021 (pp. 162)
ISBN: 978-9963-676-20-0

This monograph by Theognosia Kouspi confronts one of the major issues in the process of European integration and EU law: the principle of the supremacy of EU law and its relationship with the protection of the national identity of the Member States. In other words, it examines the conflict between national constitutions and EU law, and the dialogue between national supreme courts and the Court of Justice of the European Union (CJEU). One might assume that this topic has already been extensively discussed; nevertheless, Kouspi offers a bold and comprehensive approach to the issue. She critically examines the concept of constitutional pluralism and engages in dialogue both with the case law of the CJEU and with the jurisprudence of national courts, particularly in an effort to interpret the significance of the clause protecting the national identity of Member States.

The study's greatest contribution, beyond its thoroughness, lies in the fact that the author avoids the trap of ideological polarisation that often characterises approaches to this complex issue. As a result, she does not hesitate to seek a balance between the faithful application of the principle of supremacy and the cases in which its limitation may be justified. Her emphasis on the methodology of carrying out this dialogue, through the proportional application of balancing methods used in public law, contributes in this way to a more comprehensive treatment of the subject she addresses.

Any interpretation of economic integration, which remains detached from the broader political and institutional integration of the EU, unavoidably remains incomplete and incapable of contributing to the understanding and resolution of problems, crises, or challenges. The process of European integration was built during the Cold War, in a period in which ideal economic and political conditions prevailed for such an undertaking. The 2004 enlargement, which followed the end of the Cold War, led

to a period of political and ideological euphoria. However, the rejection of the Constitutional Treaty became the catalyst for the externalisation of the EU's problems and initiated a period of institutional crisis and skepticism regarding the future of the EU. Although the primary objectives of the movement toward European integration were political, the main means used to achieve them were economic. Yet the undeniable progress achieved in economic integration, especially after the establishment of the common European currency, could not by itself lead to political integration, nor could it resolve the problems arising from the interconnection between politics and economics in the field of international relations. Ignoring the connection between political and economic integration was one of the most misguided approaches, and it revealed its inherent weaknesses during the management of the repeated crises, economic, institutional, and political, in the last fifteen years.

Although the EU possesses decision-making procedures and mechanisms for allocating competences among its institutions and between itself and the Member States, and despite significant progress regarding European citizenship, the protection of human rights, economic integration, the method of electing the European Parliament, the appointment of the Commission, and the importance of the case law of the CJEU, the EU nevertheless lacks any fully developed political or social system of governance. The political culture of the EU is therefore bound by the political culture of its Member States. The latter sets limits on the development and completion of the former. Consequently, discussion concerning the institutional and constitutional future of the EU will always remain incomplete without a parallel discussion concerning the institutional and constitutional future of the national Member States themselves. If the sovereignty of the EU derives from the peoples of Europe (and not from their leaders), then the EU -both at the level of national leaderships and at the level of Community bureaucracy- must strengthen the institutional relationship between the peoples of Europe and the institutions of the EU.

Achilles C. Emilianides