Co-funded by the European Union

WP 2 Literature review and collection of good practices

Literature Review: Digital Citizenship and Democracy in the Digital Age- theories and concepts

Key words: Civic Engagement, Conspiracy Theories, Online Violence, Ethical Governance, Personal Data Protection, Inclusion and Diversity, Ethical Frameworks, Algorithms, Content Moderation,

Partner questions (organisation, Country Name): All the partners:

UCA/ France, MDS/ France, IPS/ Portugal, Calabria University/ Italy, Malta University/ Malta, SEHA/ Turkey, Insight/ Luxembourg.

Co-funded by
the European Union

Table of Contents

I. Conceptual Foundations of Digital Citizenship 4

a) Theoretical approaches of Digital Citizenship 4

b) Digital Public Spaces and Democratic Transformation 18

II. Analytical Approach – Theories, Methods, and Tools 34

a) Disinformation and Media Literacy 34

b) Hate Speech and Online Violence 52

c) Fake News 69

III. Ethics and Digital Responsibility 82

a) Personal Data Protection 82

b) Inclusion and Diversity in Digital Spaces 92

c) Ethical Governance of Technologies 102

Synthesis and Discussion of the Results 109

Bibliographical References 112

Introduction

The digital era has fundamentally transformed the ways individuals interact, participate in civic life, and engage with democratic processes. In this context, understanding and fostering digital citizenship has become a priority for educators, policymakers, and researchers alike. The DIGIDEM project aims to address these challenges by developing innovative educational tools and strategies to enhance digital literacy, promote democratic participation, and counteract harmful online behaviours.

This document serves as both a literature review and a collection of good practices to provide a foundation for the development of the interactive educational platform, NetizenSphere. It is structured to explore key aspects of digital citizenship and democracy, emphasizing both theoretical frameworks and practical applications. The goal is to synthesize existing knowledge, identify gaps, and leverage best practices to inform the platform's design and implementation.

The review is organized around three core thematic areas:

- Conceptual Foundations of Digital Citizenship: Including definitions, evolving vocabularies, and critical competencies, in conjunction with digital public spaces and their role in democratic transformation, particularly through social media, algorithms, and communication dynamics shaping public discourse.

- Epistemological (analytical) Analytical Approach – Theories, Methods, and Tools to digital challenges: Disinformation, media literacy, and educational methodologies. In this section, we will explore strategies for preventing hate speech and online violence, including addressing radicalization, conspiracy theories, fake news, disinformation, cyberbullying, propaganda, and online harm.

- Ethics and Digital Responsibility Ethics and Digital: Ensuring privacy, inclusion, and ethical governance in the digital sphere.

Additionally, the document incorporates partner questions to facilitate collaboration and knowledge-sharing across institutions. These questions cover experiences, methodologies, and practical insights into digital citizenship education, civic engagement, and combating disinformation. The exploration of best practices will highlight innovative pedagogical approaches, international comparisons, and successful case studies, offering actionable insights for the project's stakeholders.

By addressing these themes and questions, this document aims to create a comprehensive resource for the DIGIDEM project, guiding the design of the NetizenSphere platform and ensuring its alignment with local and European contexts while maximizing its long-term impact.

Please ensure to refer to bibliographic references and relevant resources in your responses to the questions and cite the bibliographical references using the APA format.

Please note that you are not required to answer all of the questions and concentrating on the innovative aspects regarding the mentioned themes.

If you believe that any aspect or element of the question has been studied in your country or another, please feel free to provide a summary of the relevant previous studies.

I. Conceptual Foundations of Digital Citizenship

a) Theoretical approaches of Digital Citizenship

1. Provide a definition of the following terms: disinformation, fake news, hate speech, and civic engagement, conspiracy theory, pro-democratic trolls supported by relevant bibliographic references and pertinent authors.

France:

Disinformation refers to the deliberate creation and dissemination of false information with the intent to deceive and mislead audiences. This practice is often employed to manipulate public opinion, sow confusion, or undermine trust in institutions. (Howard et al.,2018, p. 11-13) describe disinformation as "the use of automation, algorithms, and big-data analytics to manipulate public life"

Fake news is a subset of disinformation, characterized by fabricated information that mimics the form of mainstream news media but lacks the corresponding editorial standards and intent. (Gelfert, 2018 p. 84–117) defines fake news as "fabricated information that mimics news media content in form but not in organizational process or intent"

Hate speech encompasses any form of communication that belittles or discriminates against individuals or groups based on attributes such as race, religion, ethnic origin, sexual orientation, disability, or gender. (Heinze, 2016, p. 4-5) explores the relationship between hate speech and democratic citizenship, emphasizing the challenges in balancing freedom of expression with the need to protect individuals from harmful rhetoric.

Civic engagement involves individual and collective actions designed to identify and address issues of public concern. It encompasses a range of activities, from voting and participating in public debates to community service and activism, all aimed at contributing to the well-being of the community and the democratic process.

A conspiracy theory is an explanatory belief that events or situations are the result of a secret, often sinister, plot by multiple actors. These theories typically lack credible evidence and are resistant to falsification. Muirhead and (Rosenblum, 2020, p. 9-14) discuss the rise of "conspiracism without theory," where assertions are made without grounding in factual evidence, aiming to delegitimize political systems

Pro-democratic trolls refer to individuals or groups who use deceptive online identities and tactics to promote democratic values and counter authoritarian narratives. While traditional trolling is associated with disruptive behaviour, pro-democratic trolls engage in coordinated efforts to influence public opinion in favour of democratic principles. (Howard et al.,2018, p. 11-13) examines how such actors utilize digital platforms to disseminate their messages and the ethical implications of their strategies

Italy:

There has been a growing scholarly interest in digital citizenship (DC) in recent years. Traditionally, citizenship has been defined as the right to participate in society and live according to its standards fully. It encompasses civil, political, and social rights. Cogan (2000) identifies five attributes of citizenship: identity, rights, obligations, public involvement, and social values. From a general perspective, DC represents the ability to navigate our digital environments in a way that's safe and responsible and to actively and respectfully engage in these spaces. However, defining digital citizenship is challenging because scholars and organizations from various disciplines have each described it according to their norms, discourses, and practices. Additionally, another aspect that should be considered is the teacher's point of view. Understanding teachers' views on digital citizenship is crucial because their perspectives significantly influence classroom practices.

Define the following terms:

a. Disinformation: It is defined as all forms of false, inaccurate or misleading information created for harmful and for-profit reasons (European Commission (EC), 2018, p. 3) and manifests through numerous information disorders (Botha & Strömbäck; Pieterse, 2020; Marcos-Vílchez et al., 2024; Molina et al., 2021).

b. Fake news. According to Shu et al (2017) and Allcott and Gentzkow (2017) there are two accepted definitions of "Fake News" as follows: "Fake news is false news" (as a broad definition) and "Fake news is a news article that is intentionally and verifiably false" (Narrow definition). However, recent literature reviews have been published on this topic (Alghamdi et al., 2024; Botha & Strömbäck, 2022; Farhangian et al., 2024).

c. Hate speech: Numerous interpretations of hate speech are available, but a prevailing thread among them is the utilisation of language that either targets or encourages violence against specific groups characterised by distinct attributes (Rawat et al., 2024). Moreover, the rapid evolution of information and communication technology (ICT) has significantly increased internet use, enabling people to share their opinions more freely (Anjum, Katarya, 2024). This has unfortunately led to the proliferation of hate speech across various platforms. The anonymity and mobility provided by social media platforms allow individuals to hide behind screens and spread hate effortlessly. Recognizing online hate speech (OHS) is crucial in curbing these activities and can help restore public platforms as open marketplaces of ideas.

d. Civic engagement: According to Zaff et al. (2011), civic engagement refers to the involvement and active participation of individuals in their communities and society, particularly in activities that contribute to the betterment of the public good. It encompasses a broader conceptualization of engagement beyond behavioural measures, considering the influence of cultural and political contexts on the development of civic engagement among under-served populations. There have been abundant research studies conducted to examine the

positive impact of civic engagement on youth, especially on their well-being (Fong, To, 2022). A recent systematic literature review examined the relationship between civic engagement and mental health in young adults (Fenn et al., 2024). Another study has explored whether people's environmental exposures affect climate stress/anxiety or the relationship between climate stress/anxiety and civic engagement (Anneser et al., 2024).

e. Conspiracy theory: According to the most recent research in Social Psychology, the branch of psychology that has dealt the most with conspiracy theory along with political psychology, it can be defined as (De Fortuna et al., 2024): "A belief that two or more actors have coordinated in secret to achieve an outcome, and that their conspiracy is of public interest, but not public knowledge. From these defining features of conspiracy theory, other

characteristic features follow logically. Thus, conspiracy theories: (a) are oppositional, which means they oppose publicly accepted understandings of events; (b) describe malevolent or forbidden acts; (c) ascribe agency to individuals and groups rather than impersonal or systemic forces; (d) are epistemically risky, meaning that though individual conspiracy theories are not necessarily false or implausible, conspiracy theories taken collectively are more prone to falsity than other types of belief; and (e) are social constructs that are not merely adopted by individuals but shared with social objectives in mind, and that have the potential not only to represent and interpret reality but also to fashion new social realities" (Douglas & Sutton, 2023, p. 287, in De Fortuna et al., 2024). People's tendency to believe in conspiracy theories, known as conspiracy mentality, varies. Empirical studies aim to explore this social phenomenon. Nera et al. (2024), by conducting three experiments, showed that this mentality predicts support for conspiratorial explanations and rejection of non-conspiratorial ones. However, these tendencies are influenced by whether authorities endorse or reject the explanations. Those with a low conspiracy mentality are more swayed by authorities; endorsements, while conspiracy believers remain unaffected. These findings suggest the need for clearer definitions in research on conspiracy theories.

f. Pro-democratic trolls: Pro-democratic trolls refer to individuals or groups who use trolling tactics online to support democratic values or promote pro- democracy agendas. These trolls often engage in inflammatory or provocative behaviour to influence public opinion, counter anti-democratic narratives, or disrupt the activities of those opposing democratic principles

(https://slate.com/technology/2019/02/trolls-internet-democrats-racist-trump-harris-warren bernie-sanders-trump.html; https://forarpeople.org/political-trolling-101/).

Turkey:

Civic Engagement refers to individual and collective actions aimed at addressing issues of public concern. It involves active participation in public life through actions such as voting, volunteering, activism, and community involvement

Conspiracy Theories are beliefs suggesting that secret and powerful groups manipulate events or conceal significant information from the public. Conspiracy theories typically reject standard explanations and favour simplified, alternative narratives, often spreading rapidly online.

Online Violence involves aggressive, harmful, or threatening behaviours that occur through digital platforms, including harassment, cyberbullying, and hate speech. These behaviours are

characterized by repeated attacks, threats, or harassment through digital communication channels.

Ethical Governance involves frameworks and processes guiding responsible decision-making and actions within digital environments. Ethical governance emphasizes transparency, accountability, fairness, and respect for human rights when managing technological tools and data.

Personal Data Protection refers to practices, policies, and frameworks designed to secure individuals' private information from misuse, unauthorized access, and exploitation. It ensures individuals' privacy rights through regulations such as the European General Data Protection Regulation.

Inclusion and Diversity in digital spaces address equal access and representation for various social, ethnic, gender, and disability groups. It emphasizes creating environments that are accessible, safe, and equitable for all users.

Algorithms are computational rules and processes that platforms use to manage, filter, and distribute digital content. Algorithms significantly shape online discourse, potentially amplifying fake news, hate speech, or echo chambers.

Content Moderation is the practice of monitoring and regulating user-generated content on digital platforms to minimize harmful or inappropriate content. Effective content moderation aims to reduce misinformation, hate speech, and harmful behaviours.

Pro-democratic Trolls are individuals or organized groups that deliberately intervene in online discussions to promote democratic values, counteract authoritarian narratives, and challenge disinformation through strategic messaging, often utilizing irony or satire to reduce polarization.

IPS:

- Disinformation: According to Centro da Internet Segura (n.d.), we can define disinformation as any type of content and/or practice that contributes to the increase of falsified, unvalidated or unclear/transparent information to distance citizens from factual knowledge of reality (Centro Internet Segura, n.d.).

- Fake news: Fake news can be defined as news that are distributed with the intention of spreading false information and rumours, independently of the means of communication utilised as well as the motivation associated with them (Centro Internet Segura, n.d.).

- Hate speech: According to APAV (2018), hate speech includes all forms of expression that propagate, incite, promote, or justify racial hatred, xenophobia, homophobia, antisemitism and other forms of hatred based on intolerance (APAV, 2018).

- Conspiracy theory: Generically, conspiracy theories can be defined as ‘attempts to explain the ultimate causes of significant social and political events and circumstances with claims of secret plots by two or more powerful actors’’ (Douglas, et al., 2019). Offering another definition, and within the Portuguese national setting, Candeias (2023) defines conspiracy theories as explanations or circumstances based on malicious acts of secret groups and powerful

people (Candeias, 2023). For the author, they all share some common elements, such as a secret agenda, a conspiracy group, evidence that supports a theory and the attribution of blame to a specific group (scapegoating) (Candeias, 2023).

- Civic engagement: Civic engagement can be defined as an action that tends to involve the participation of a citizen or a group of citizens either in communitarian, civic or political processes (Barreto, 2002).

- [extra relevant definition] Media Literacy: According to the Portuguese National Plan of Media Literacy, we can define media literacy as the ability of people to relate to the media and be able to access, use, produce, and critically understand their message and the context media operates (Governo da República Portuguesa, 2025).

Luxembourg:

European Commission. (2022). Guidelines for teachers and educators on tackling disinformation and promoting digital literacy through education and training. Luxembourg: Publications Office of the European Union. ISBN 978-92-76-55108-9 doi: 10.2766/28248

https://education.ec.europa.eu/document/guidelines-for-teachers-and-educators-on-tackling-disinformation-and-promoting-digital-literacy-through-education-and-training

INCLUDES A VERY INTERESTING AND SIMPLE GLOSSARY

https://www.researchgate.net/publication/337812656_Digital_Citizenship_Education_overview_and_new_perspectives. Reviews existing academic and policy literature on digital citizenship education, highlighting definitions, actors and stakeholders, competence frameworks, practices, emerging trends and challenges.

https://rm.coe.int/prems-003222-gbr-2511-handbook-for-schools-16x24-2022-web-bat-1-/1680a67cab). Intended for teachers and parents, education decision makers and platform providers, it provides a comprehensive approach to digital citizenship education. It offers information, tools and good practice to support the development of competences to empower and protect children, enabling them to live together as equals in today's culturally diverse democratic societies, both on- and offline.

UNESCO (2024, lead author: Richardson. J). Teacher Guidelines. ISBN 978-92-3-100658-6 https://www.gcedclearinghouse.org/resources/global-citizenship-education-digital-age-teacher-guidelines?language=en

Aimed at building the capacities of teachers to prepare learners to act ethically and responsibly in physical and digital environments, these guidelines stress the importance of considering how Global Citizenship Education (GCED) and related knowledge, values, skills and attitudes can be fostered for teaching, learning and engaging for a more sustainable, inclusive, just and peaceful world.

https://doi.org/10.1177/14614448211068530

https://journals.sagepub.com/doi/full/10.1177/14614448211068530

Keywords: Digital media, mass media, media literacy, perceived reality, social media

Abstract: Concerns over the harmful effects of social media have directed public attention to media literacy as a potential remedy. Current conceptions of media literacy are frequently based on mass media, focusing on the analysis of common content and evaluation of the content using common values. This article initiates a new conceptual framework of social media literacy (SoMeLit). Moving away from the mass media-based assumptions of extant approaches, SoMeLit centers on the user's self in social media that is in dynamic causation with their choices of messages and networks. The foci of analysis in SoMeLit, therefore, are one's selections and values that influence and are influenced by the construction of one's reality on social media; and the evolving characteristics of social media platforms that set the boundaries of one's social media reality construction. Implications of the new components and dimensions of SoMeLit for future research, education, and action are discussed. https://doi.org/10.1007/978-3-031-69362-5

Keywords: Cognitive, physical, and psychosocial impacts of digital technology on infants, children, and adolescents; influences of media on individuals as well as relationships, family, culture, and society; Impacts of specific digital domains pertinent to youth (e.g., education technology, video gaming).

Wuyckens, G., Landry, N., & Fastrez, P. 168-182 (2022). https://doi.org/10.23860/JMLE-2022-14-1-12

Keywords: media literacy, information literacy, digital literacy

Abstract: This article presents a systematic meta-review of the scientific literature discussing the concepts of information literacy, media literacy, and digital literacy. While carrying out a cross analysis of the way in which literature reviews specifically address these three concepts, this article identifies, and articulates a critical analysis of, the main findings from the reviewed texts regarding the conceptual landscape that they cover. This work highlights confusion between the constitutive dimensions of literacies, recurrent difficulties in establishing theoretical articulations between contributions, and operationalization problems in observing and assessing these literacies. These issues are the subject of a discussion grounded in the specific field of media education.

Polanco-Levicán, K., & Salvo-Garrido, S. P. 8807 (2022). https://doi.org/10.3390/ijerph19148807

Keywords: Social media literacy, digital literacy, media literacy

Abstract: Nowadays, people spend long periods on social media, ignoring the implications this carries in daily life. In this context, the concept of social media literacy, an emerging concept scarcely developed in the literature, is relevant. This study sought to analyze, descriptively, the main definitions and competences of the concept of social media literacy. The methodology included a systematic search of literature in the databases Web of Science, PubMed, and Scopus between 2010 and 2021, applying filters for English and Spanish, including only scientific articles. A total of 1093 articles were obtained. An article selection process took place, applying the inclusion and exclusion criteria, resulting in a total of 15 articles being selected. The findings indicate that the concept of social media literacy is based on media literacy to then integrate the characteristics and the implications of digital platforms. This is linked to the development of cognitive competences, where critical thinking, socio-emotional competences, and technical competences are fundamental, considering the social context. The development

of socio-emotional competences stands out since social media are a frequent place of interaction between people.

Bal, E., & Akcil, U. P. 445 (2024). https://doi.org/10.3390/su16010445

Abstract: This study aims to develop and evaluate the effectiveness of a digital citizenship curriculum developed using online teaching for undergraduate students. For this purpose, answers were sought to the following general research questions: "How are students performing their digital citizenship skills?", "What is the efficiency level of the developed digital citizenship curriculum?", "What are the students' self-evaluations during the implementation process of the developed digital citizenship curriculum?" and "What are student opinions about the implemented digital citizenship course?". A "parallel-simultaneous design", one of the mixed research methods, was used in this research. The content of the curriculum was determined within the framework of a needs analysis consisting of university students and field experts, and a document review. A quasi-experimental design method was used to implement the programme. A total of 39 teacher candidates, registered in the spring semester of the 2022 academic year, participated in the experimental process. A Moodle learning management system was used as the teaching environment. Participants' digital citizenship behaviours and attitude levels were measured before and after the experimental application. It was observed that there was a significant (positive) difference between the pre-test and post-test average scores in favour of the post-test scores, and the effect value of this difference was high. It turned out that the teacher candidates' opinions on the effectiveness of the online digital citizenship programme were positive. It is understood that the digital citizenship teaching curriculum developed for higher education can be used sustainably for similar university student groups as in this study. However, it is anticipated that the developed programme can be updated along with technological evolutions.

Keywords: curriculum; experimental practices; global citizenship; online learning; undergraduate level

Kim, M., & Choi, D. P. 155-171 (2018).

Abstract: Digital citizens need comprehensive knowledge and technological accessibility to the internet and digital world and teachers have a responsibility to lead them to become digital citizens. However, existing Digital Citizenship Scales contain too broad ranges and do not precisely focus on the target students, so teachers do not have clear criteria for facilitating young people to have digital citizenship, which leads to problems in the direction and goals of digital citizenship education. This study aims to clearly identify the Digital Citizenship Scale for adolescents perceived by teachers who are responsible for their students' digital citizenship education and to present the needs and direction of school-based education to satisfy the identified Digital Citizenship Scales. A five-factor Digital Citizenship Scale called S.A.F.E model, meaning leading character of Self-identity in digital environment, Activity in online(Reasonable Activity and Social/cultural engagement), Fluency for the Digital tools, and Ethics for digital environment, was derived through Exploratory Factor Analysis (EFA) and further cross-validated through Confirmatory Factor Analysis (CFA) with 200 pre-service teachers and in-service teachers. The S.A.F.E model, derived by teachers, shows high reliability and construct validity to be used as a digital citizenship scale for students through concept analysis, EFA, and CFA verification. Based on the theoretically rigorously derived SAFE

model, the educational requirements and direction to become active and critical citizens in the online community were discussed.

Keywords: Adolescent digital citizenship education; Adolescent digital citizenship scale; Factor analysis; Teachers' perspective on digital citizenship scale

Bengio, Y. et al. (2025

https://assets.publishing.service.gov.uk/media/679a0c48a77d250007d313ee/International_AI_Safety_Report_2025_accessible_f.pdf

The purpose of this report:

This report synthesises the state of scientific understanding of general-purpose AI – AI that

can perform a wide variety of tasks – with a focus on understanding and managing its risks.

This report summarises the scientific evidence on the safety of general-purpose AI. The purpose of

this report is to help create a shared international understanding of risks from advanced AI and how

they can be mitigated. To achieve this, this report focuses on general-purpose AI - or AI that can

perform a wide variety of tasks - since this type of AI has advanced particularly rapidly in recent

years and has been deployed widely by technology companies for a range of consumer and

business purposes. The report synthesises the state of scientific understanding of general-purpose

AI, with a focus on understanding and managing its risks.

Boudana, S., & Segev, E. (2024) https://doi.org/10.1080/17512786.2024.2379898

Abstract: This paper explores how news media define and frame fake news, thus orienting public perceptions of fake news in terms of problem, causality, responsibilities, and solutions. It first maps the news discourse on fake news through semantic network analysis of 1,646 articles published in 2017–2018 (when the term reached its peak) and 2021–2022. It then analyses the definitions and frames of fake news in a sample of 122 articles representative of the clusters identified in the previous step of our analysis.

The semantic networks in both periods reveal similar dominant actors - political actors, social media platforms and professional journalists – that shape the “fake news” narrative. Fake news is either defined as disinformation and political propaganda or, more often, as false accusations that Donald Trump and his supporters make against mainstream media for their negative reporting. Political and social media actors are mostly held responsible for undermining public trust in mainstream media through fake news dissemination and/or false accusations. In this polarized conception of the public sphere, the public is positioned as a victim rather than an active contributor to the fake news problem and solution.

Douglas, K. M., & Sutton, R. M. P. 271–298 (2023). https://doi.org/10.1146/annurev-psych-032420-031329

Abstract: Conspiracy theories are abundant in social and political discourse, with serious consequences for individuals, groups, and societies. However, psychological scientists have started paying close attention to them only in the past 20 years. We review the spectacular progress that has since been made and some of the limitations of research so far, and we consider the prospects for further progress. To this end, we take a step back to analyze the defining features that make conspiracy theories different in kind from other beliefs and different in degree from each other. We consider how these features determine the adoption, consequences, and transmission of belief in conspiracy theories, even though their role as causal or moderating variables has seldom been examined. We therefore advocate for a research agenda in the study of conspiracy theories that starts—as is routine in fields such as virology and toxicology—with a robust descriptive analysis of the ontology of the entity at its center.

Keywords: conspiracy theories, conspiracy beliefs, predictors, consequences, communication,

Theory

Sponholz, L. P. 143–163 (2023) https://doi.org/10.48541/dcr.v12.9

Abstract: Hate speech - communication that attacks a person or a group on the basis of identity factors, such as gender, race, or religion - is one of the main digital threats to democracy. Hate speech has manifold, empirically evidenced consequences for targeted individuals and groups experiencing systematic discrimination and for social cohesion as a whole. Yet, while the upheaval of social media has put the concept in the spotlight, such attention has also structurally transformed its meaning, turning hate speech from a concept with clear defining properties into a family resemblance comprising all kinds of online abuse. This process is far from causing only academic issues. It also sidesteps historical oppression as a defining property and as the reason for which one is targeted by hate speech. Thus, the process has been belittling public animosity against historically oppressed groups, reducing hate speech merely to a matter of offensive language on social media. This chapter shows how and why this conceptual change has taken place and the consequences it unleashes. It specifically addresses the problems of concept stretching, concept shrinking, and the inflation of concepts. Finally, it concludes that such conceptual issues jeopardize the potential that digital media research on hate speech has to provide guidance to a broad range of social actors.

Wardle (2024) https://peacekeeping.un.org/sites/default/files/report_-_a_conceptual_analysis_of_the_overlaps_and_differences_between_hate_speech_misinformation_and_disinformation_june_2024_qrupdate.pdf

This report, commissioned by the UN Department of Peace Operations and the UN Office of the Special Adviser on the Prevention of Genocide, provides a comprehensive conceptual analysis to distinguish between hate speech, misinformation and disinformation. It aims to identify their overlaps and differences, particularly in conflict-affected and high-risk areas. The report reviews relevant international human rights law and international humanitarian law, emphasising the need for responses that respect freedom of expression while addressing harmful speech. It also draws on interviews with a small sample of 15 staff working on either or all of these three areas with the UN and humanitarian organizations.

Peissl, Helmut; Sedlaczek, Andrea P. 11 (2022) https://doi.org/10.25656/01:24474

Abstract: Competent and critical media practice is a societal challenge that is also facing adult (political) education. The article discusses two media literacy concepts that stress the political dimension of media use as social action and offer useful starting points for adult education: the UNESCO media and information literacy (MIL) concept and the critical media literacy (CML) concept developed by Douglas Kellner and Jeff Share. Both concepts rest upon the common assumption that media literacy cannot be reduced to a purely technical, functional perspective but that a holistic, intersectional and interdisciplinary orientation is required to meet multiple societal challenges in times of digital transformation. One of the authors' conclusions is that it is necessary to provide adult educators with critical media literacy and to establish it as an interdisciplinary topic in adult education offerings. Since adult educators function as multipliers, they can contribute to the promotion of assessment, reflection and action competencies in all of society in the age of digital transformation and mediatization.

German Federal State of Hesse

Definition of "fake news"

https://www.medienanstalt-hessen.de/hass-hetze-desinformation/

2. What are the relevant experiences and skills in your country in the field of digital democracy and citizenship?

France:

In France, the advancement of digital democracy and citizenship has been marked by several key initiatives and developments. The French government has prioritized open government policies, focusing on transparency, participation, and collaboration. A study analysing 979 open government platforms in France, Italy, and the United Kingdom highlighted that France has made significant strides in e-government and open data. However, it also noted that participatory platforms in France exhibit a higher degree of fragmentation at both transnational and international levels, indicating a need for better integration and coordination by central authorities.

Participatory platforms are digital tools designed to facilitate citizen engagement in decision-making processes. These platforms include online petitions, deliberative forums, crowdsourcing initiatives, and e-consultation portals. Examples in France include "Decider ensemble," which promotes collective decision-making, and "France Connect," a digital identity verification system enabling secure access to public services. These platforms enhance democratic participation by giving citizens a voice in shaping policies and governmental actions.

To address these challenges, France has implemented various programs aimed at empowering citizens in the digital realm. The "France Très Haut Débit" initiative, launched in 2013, aims to provide nationwide high-speed broadband coverage. Its objectives include:

- Ensuring universal access to very high-speed internet (minimum 30 Mbps) by 2022.
- Expanding fiber-optic networks to replace outdated copper infrastructures.

- Reducing the digital divide between urban and rural areas, fostering equal opportunities for digital participation.

Another critical initiative is the "New Deal Mobile," introduced in 2018 to improve mobile network coverage in underserved areas. The program involves collaboration between the French government and telecom operators to:

- Deploy 4G coverage in rural and remote regions.
- Install new cell towers to enhance connectivity.
- Enforce obligations on telecom providers to ensure high-quality network access for all citizens.

These efforts signify a shift towards a more inclusive approach, granting local actors and citizens greater agency in the digital landscape. The government's support for citizen-led digital projects during the COVID-19 pandemic further underscores this commitment to fostering a democratic digital environment.

In the educational sector, the Council of Europe has developed the "Digital Citizenship Education Handbook," which serves as a resource for educators, policymakers, and stakeholders. This handbook outlines the competencies required for responsible digital engagement and provides tools and best practices to cultivate these skills among learners. It emphasizes the importance of empowering individuals to navigate the digital world effectively, ensuring they can participate fully in democratic processes.

Collectively, these initiatives reflect France's dedication to enhancing digital democracy and citizenship by promoting transparency, citizen participation, and digital literacy.

Italy:

In Italy, digital democracy and citizenship have been evolving significantly, especially with the introduction of electronic voting and digital identity systems. For instance, the “Spid vote” initiative aims to streamline democratic processes by enabling electronic voting through digital identities (Borriello, 2023). This approach is part of a broader transformation towards more direct and speedy citizen participation in democratic activities. The COVID-19 pandemic also accelerated the adoption of digital technologies for political participation.

2.What tools or methodologies are currently being used in your country to promote digital citizenship, fake news, disinformation, and hate speech?

a. Parole O_Stili: A project promoting respectful online communication through an

ethical manifesto and educational resources (https://paroleostili.it/)

b. OSCAD (Osservatorio per la Sicurezza Contro gli Atti Discriminatori): Collects and responds to reports of hate crimes, including online hate speech (https://www.interno.gov.it/it/ministero/osservatori-commissioni-e-centri coordinamento/osservatorio-sicurezza-contro-atti-discriminatori-oscad).

c. EU-Funded Projects: Italy participates in European initiatives like EDMO (European Digital Media Observatory), focusing on research, fact-checking, and media literacy (https://digital-strategy.ec.europa.eu/en/policies/european-digital- media-observatory).

Turkey:

In Türkiye, relevant experiences in digital democracy and citizenship are prominently reflected through governmental and civil initiatives aimed at enhancing digital literacy, civic engagement, and combatting online threats. Türkiye's "EBA" (Education Information Network) platform, launched by the Ministry of National Education, supports digital literacy and responsible online behavior among students, integrating digital citizenship concepts into education (MEB, 2023). Several NGOs, such as Dijital Dönüşüm Derneği and Habitat Derneği, actively promote digital skills, data protection awareness, and online ethics through community workshops, training programs, and advocacy campaigns (Habitat Derneği, 2022). Additionally, initiatives like the EU-funded "SafeWeb" project specifically target online hate speech and misinformation, reflecting Türkiye's proactive approach to these challenges (UNDP Türkiye, 2022).

Portugal:

• In Portugal, there have been a few initiatives that have been important to enhance digital democracy and citizenship. We highlight the following:

- Digital Skills Passport: This initiative was launched by the Lisbon City Council and aims to address citizens' digital gaps. This programme targets individuals with no prior knowledge of Internet usage as well as intermediate digital skill users (European Association for the Education of Adults, 2021).
- InCODe.2030: This government programme aims to promote digital democracy skills among Portuguese citizens. The main goal of this programme is to foster digital education and awareness of technology from an early age, as well as to support the digital competencies of educators.
- EUSOUDIGITAL: This initiative aims to promote digital literacy, envisioning to reach at least 1 million adults by the end of 2025, by developing a national network of thousands of volunteers in more than 1.500 spaces across the country. The programme will help adults who never used the Internet before, by providing training sessions developed by volunteers in a familiar context, like schools, assisted living homes and others (DECO, n.d.).
- MUDA (Movement for Active Digital Use): This national movement promoted by the Portuguese Government aims to encourage the participation of Portuguese citizens in the digital space and help them take advantage of the benefits associated with digital services made available by companies and the state (MUDA, n.d.).

Luxembourg: No Answer

3. What tools or methodologies are currently being used in your country to promote digital citizenship, fake news, disinformation, and hate speech?

France:

In France, a multifaceted approach has been adopted to promote digital citizenship and combat fake news, disinformation, and hate speech. This strategy encompasses legislative measures, educational initiatives, and collaborative efforts with digital platforms.

Legislative Measures

Several French governmental institutions are responsible for regulating online content, including the Ministry of Culture, the Ministry of the Interior, the Conseil Supérieur de l'Audiovisuel (CSA, now part of ARCOM), and the judiciary. Key legislative measures include:

The Law Against the Manipulation of Information (2018): This law was designed to counter the spread of false information, particularly during election periods. It enables judicial authorities to act swiftly by ordering the removal of deliberately false content and imposes transparency obligations on digital platforms regarding sponsored content.

The Avia Law (2020): Aimed at combating online hate speech, this law initially required platforms to remove manifestly illegal content within 24 hours. However, key provisions were struck down by the Constitutional Council due to concerns about freedom of expression. The remaining aspects of the law focus on strengthening legal mechanisms, such as establishing a specialized public prosecutor's office and an Online Hate Observatory under the supervision of the Audiovisual and Digital Communication Regulatory Authority (ARCOM).

The Role of ARCOM

ARCOM (Autorité de Régulation de la Communication Audiovisuelle et Numérique) is the French regulatory authority responsible for overseeing digital and audiovisual communications. It plays a crucial role in:

- Ensuring compliance with laws on disinformation and hate speech.
- Monitoring digital platforms to verify their efforts in tackling illegal content.
- Issuing recommendations and sanctions to enforce regulations.

Educational Initiatives

To strengthen media and information literacy, France promotes digital citizenship through various educational programs. These initiatives, supported by the Ministry of National Education and the CLEMI (Centre for Media and Information Literacy Education), aim to:

- Encourage critical thinking and responsible online behaviour.
- Help users differentiate between credible and misleading information.
- Integrate digital literacy into school curricula to foster awareness among young people.

Collaborative Efforts with Digital Platforms

French authorities work closely with major digital platforms such as Google, Facebook (Meta), Twitter (X), and YouTube to improve content moderation and transparency. These partnerships typically involve:

- Co-regulation agreements, where platforms commit to removing illegal content while adhering to legal obligations.
- Reporting mechanisms, allowing users to flag illicit content easily.
- Algorithmic transparency requirements, ensuring platforms disclose how their recommendation systems influence the spread of information.

Regulatory Framework

France's digital content regulation aligns with both national and European frameworks, including:

- The French Law on Digital Services (Loi sur les Services Numériques), which integrates aspects of the EU's Digital Services Act (DSA) to hold platforms accountable for harmful content.
- The General Data Protection Regulation (GDPR), which governs online data protection and privacy.

This comprehensive strategy demonstrates France's commitment to fostering a safer and more informed digital environment through regulation, education, and cooperation with digital stakeholders.

Italy: No answer

Turkey:

In Türkiye, various tools and methodologies are currently used to promote digital citizenship and address disinformation and hate speech. The Ministry of National Education integrates digital literacy curricula emphasizing critical thinking, recognizing fake news, and responsible online behaviour through projects such as EBA (Education Information Network) and FATİH (Movement to Enhance Opportunities and Improve Technology) (Gökçe & Dinçer, 2021). NGOs like Teyit.org provide fact-checking services and training programs aimed at debunking disinformation and misinformation (Teyit, 2022). Social media platforms, especially Twitter and Facebook, collaborate with local organizations and fact-checkers to reduce hate speech and disinformation.

Portugal:

• In Portugal, there have been a few tools that are utilised to promote digital democracy and citizenship. We highlight the following:

- National Plan of Media Literacy (Plano Nacional de Literacia Mediática 2025-2029): The Portuguese government has created the National Plan of Media Literacy, inspired by the best practices implemented in Europe (such as the EDMO Guidelines for Effective Media Literacy Initiatives 2024), to tackle disinformation and promote media literacy. This national plan aims to promote media literacy among all audiences; combat misinformation and fake news; promote responsible and informed content consumption and promote more informed and participatory citizenship (Governo da República Portuguesa,

2025). With these goals at the forefront, this plan will respond to the socioeconomic, cultural and technological challenges that the country faces in the present and future (e.g., socioeconomic inequalities, aging population, rise of migrant communities) (Governo da República Portuguesa, 2025).

- [generic yet including Portugal] The Empowering Europeans Against Disinformation project (EEAD): This programme started in 2018 and since then has been offering complex support to civil society organisations in countering disinformation. This initiative, which focuses on Czech Republic, Hungary, Poland, Portugal and Romania, aims to increase the understanding around disinformation threats, strengthen the capacity of civil actors in dealing with disinformation and boost their knowledge on the field (Hive Mind, 2023).
- [European] The Digital Services Act: This law was created to protect the fundamental rights of the online users and has become a groundbreaking piece of legislation for the online space, holding platforms accountable for illegal and harmful content, including disinformation and hate speech (Executive, 2024).

Luxembourg: No Answer

b) Digital Public Spaces and Democratic Transformation

1. How do digital public spaces contribute to the transformation of democratic processes?

France:

Digital public spaces, including social media platforms (Facebook, Twitter/X, YouTube), discussion forums (Reddit, Quora), and civic engagement platforms (Change.org, Decidim, Parlement & Citoyens), have significantly transformed democratic processes by enhancing citizen engagement, facilitating information dissemination, and promoting participatory governance. These platforms enable individuals to access political information in real-time, engage in debates, and hold public officials accountable, thereby fostering a more inclusive and transparent democratic environment.

How Do These Platforms Enhance Democratic Transparency?

- Open Government Initiatives: Platforms such as France's Open Data Portal (data.gouv.fr) provide citizens with access to governmental data, improving transparency and enabling data-driven public discourse.
- Live Political Discussions & Fact-Checking: social media enables real-time interaction between politicians and the public. Platforms like Twitter/X and YouTube host live debates, while fact-checking initiatives such as AFP Factual and Check News (Libération) help combat misinformation.
- Petitions & Grassroots Movements: Websites like Change.org and MesOpinions.com empower citizens to influence policy decisions through petitions and advocacy campaigns.

The Role of Politicians on social media

Many politicians actively use social media to connect with citizens, share policy updates, and shape public opinion. Examples include:

- Emmanuel Macron (President of France) – Uses Twitter/X and Instagram to communicate government actions, particularly during crises such as COVID-19 and pension reform debates.
- Jean-Luc Mélenchon (France Insoumise) – Engages with followers through YouTube live streams and Twitter, promoting his party's policies and responding to current events.
- Marine Le Pen (Rassemblement National) – Uses Facebook and TikTok to appeal to younger audiences and discuss nationalist policies.

Challenges & Considerations

While digital public spaces strengthen democratic engagement, they also present challenges:

- Misinformation & Echo Chambers: False information can spread rapidly, influencing public perception and election outcomes.
- Digital Divide: Unequal internet access can exclude marginalized communities from political discourse.
- Regulatory Needs: Striking a balance between content moderation, free speech, and platform accountability remains a key challenge for policymakers.

Italy:

The potential offered by web 4.0 allows for immersive experiences by offering spaces for co-design, collaboration in learning, entertainment, but also democratic participation, as they offer citizens new ways and tools to express their opinions and suggestions on issues and policies, therefore the new platforms play a central role in the processes of active involvement of the community (Bentivegna & Boccia Artieri 2021, Lovari 2022). Even if Digital public spaces have reshaped democratic processes, making political, social and democratic debate more accessible and immediate (Miconi, 2020) some studies emphasize that online interaction, often characterized by impulsive can negatively impact the quality of democratic debate reducing awareness of traditional social norms, and increasing aggression (Boccia Artieri, 2012).

Turkey:

Digital public spaces significantly transform democratic processes by enhancing citizen participation, enabling real-time dialogue, and reshaping public discourse. They provide platforms for political engagement beyond traditional forms, allowing users to engage directly with policymakers, express opinions, and mobilize collective actions (Dahlgren, 2021). Social media platforms, forums, and online communities facilitate broader civic involvement, particularly among youth and marginalized groups, amplifying previously underrepresented voices (Bennett & Pfetsch, 2018). However, digital public spaces also pose risks, such as echo chambers, misinformation, and polarization, which challenge constructive dialogue and democracy itself (Sunstein, 2017).

Portugal:

General:

- Technology can have an important role in revitalising democratic governance by empowering citizens through new kinds of democratic institutions and practices that promote different data imaginaries and offer practical counter interventions (Fischli & Muldoon, 2024).

- A positive step in democratic participation was achieved by the adoption of digital voting systems that have enabled the participation of individuals living in remote areas, people with disabilities and people who live outside of their homeland (Bi, 2024). This mechanism is also important to stimulate the participation of the younger generation, as they might be more prone to participating through digital platforms (Bi, 2024).

Portugal:

- In Portugal, Participa.pt, a governmental platform was created by the Portuguese government to allow citizens to actively participate digitally in public conversations, as well as to stay up to date with news, legislative proposals, this way countering and/or eliminating some barriers felt.

Luxembourg: No Answer

2. How do social media, online forums, and digital platforms influence public discussions and political participation, and in what ways do they shape public opinion, information sharing, and engagement with political issues?

France:

Social media, online forums, and digital platforms have profoundly reshaped public discussions and political participation by providing unprecedented opportunities for information sharing, engagement, and opinion formation. These platforms enable the rapid dissemination of information, allowing users to access diverse viewpoints and participate in political discourse. However, they also introduce significant challenges, such as the spread of misinformation, the reinforcement of ideological echo chambers, and the suppression of offline conversations.

One of the key transformations brought by these platforms is the creation of new opportunities for engagement. Social media allows individuals to connect with political movements, participate in discussions, and mobilize around causes in ways that were previously impossible. A study by (Keith et al., 2016, p. 178-190) highlights that while social media fosters political engagement, it can also suppress offline conversations. The study suggests that the widespread visibility of political opinions online contributes to a "spiral of silence," where individuals may refrain from discussing controversial topics in face-to-face settings if they fear social backlash or perceive their views to be in the minority.

Furthermore, the attention economy plays a crucial role in shaping public opinion and information consumption. Social media platforms are designed to prioritize content that maximizes user engagement, often amplifying emotionally charged and sensational material. According to the European Commission's Joint Research Centre (2020), this dynamic can contribute to the spread of polarizing messages, reinforcing ideological divisions and reducing exposure to diverse perspectives. As a result, political discourse can become more fragmented, and misinformation can gain traction more easily.

Despite these challenges, digital platforms also offer enhanced opportunities for political engagement. A report by the Pew Research Center (2013, p. 307-332) found that a growing number of individuals use social networking sites to participate in political activities, such as sharing political content, following political figures, and engaging in activism. These digital spaces provide new ways for marginalized voices to be heard and facilitate broader political mobilization.

Understanding these dynamics is crucial for fostering informed and constructive political discourse in the digital age. Policymakers, researchers, and platform developers must navigate the balance between promoting free expression and mitigating the negative effects of misinformation, polarization, and reduced offline interactions. By recognizing both the opportunities and challenges posed by digital platforms, societies can work towards more inclusive and informed public discussions.

Italy:

The use of platforms such as Facebook, Twitter and Instagram as a communication and information system has led to a radical change in the way information is conveyed and how participation in public discussions occurs. Communication is more open and direct and it is possible to reach a large and diverse audience for the sharing of messages, ideas and programs in real time. This level of accessibility and interaction has strongly influenced public and political communication, encouraging greater user participation, but also the development of problems and controversies.

Turkey:

Digital public spaces in Türkiye influence democratic processes significantly by increasing citizen participation and shaping public opinion. Platforms such as Twitter, Facebook, and YouTube enable rapid dissemination of information, fostering civic engagement and political activism, especially among youth. Social media facilitates mobilization and organization during key political events, such as elections or social movements, making democratic participation more accessible. However, these platforms can also amplify polarization and misinformation, influencing perceptions and trust in democratic institutions negatively (Aydın & Ayhan, 2022). Effective management of these digital spaces requires enhancing media literacy and transparency to strengthen democracy.

Portugal:

General:

- Existing research has shown that social media and digital platforms have a significant impact on shaping public opinion and also influencing economic decision-making. One way that social media and digital platforms have impacted politics is through the amplification of political messages. Before the media upsurge, politicians depended strongly on traditional media to reach their voters, but now millions of people can be reached by a simple tweet or Facebook post. Free access to information allows for the construction of politically conscious citizens, making the population's participation more effective in the moment of choice, as previous knowledge is an important step to democracy (Paraense, 2022). For this, governments need to take social media's influence into account when creating economic and social policies and work on improving digital literacy (Ausat, 2023).

Portugal:

A study conducted with Portuguese youth showed that young people who share frequently current affairs on social media platforms are also more active in other forms of political participation (such as voting, signing petitions, and participation in movements and protests) (Feio, Oliveira, & Martins, 2023).

This has been a growing concern in Portugal, since the General Directorate for Education has been working with GILM (Grupo Informal sobre Literacia Mediática), to promote annual encounters which educate youth towards media. In 2024, the encounter was focused on hate and peace discourses in time of war (Público na Escola, 2024). Nevertheless, and interestingly, an Iberifier study (Disinformation consumption patterns in Portugal and Spain), showcased how Portuguese youth, ages 18-24, is the one with the least trust in media and journalism as the solution for disinformation (Lusa, 2024).

Luxembourg: No Answer

3. How do algorithms and content moderation impact democratic transformation in digital public spaces?

France:

Algorithms and content moderation play a crucial role in shaping democratic transformation in digital public spaces by influencing the way information is disseminated, the accessibility of diverse perspectives, and the overall quality of political discourse. While these mechanisms can enhance the integrity of online discussions by filtering out harmful content and reducing misinformation, they also raise concerns about bias, censorship, and their broader impact on democratic institutions.

One of the primary ways algorithms influence democratic transformation is by shaping information flow. Social media and digital platforms use algorithmic systems to prioritize content based on user engagement, relevance, and other factors. These algorithms determine what users see in their feeds, often amplifying content that generates strong emotional reactions. As a result, political discussions can become skewed toward sensationalist or polarizing narratives, limiting exposure to balanced and nuanced perspectives.

Additionally, these algorithms can contribute to limiting access to a wide range of perspectives and information. By tailoring content based on users' past interactions, recommendation systems create filter bubbles and echo chambers, where individuals are primarily exposed to viewpoints that align with their existing beliefs. This selective exposure reduces opportunities for critical debate and weakens the democratic principle of an informed and diverse public sphere.

Beyond content curation, content moderation policies also impact democracy by influencing what is deemed acceptable discourse in digital public spaces. While moderation is essential to combating hate speech, disinformation, and harmful content, it can also have unintended consequences, such as destabilizing democratic institutions. When moderation decisions lack

transparency or appear biased, they can fuel distrust in platforms and institutions. For example, perceived political bias in content removal or the inconsistent enforcement of moderation policies can lead to accusations of censorship, further polarizing public opinion and undermining trust in democratic processes (Gorwa et al.,2020, p.12).

Given these challenges, scholars advocate for greater transparency in how algorithms function and how content moderation decisions are made. Researchers and policymakers emphasize the need for clearer guidelines, accountability mechanisms, and independent oversight to ensure that digital platforms support rather than hinder democratic participation. Increased transparency would allow users to understand why certain content is promoted or removed, reducing concerns about manipulation and bias.

This approach which emphasizes algorithmic accountability, diverse content exposure, and fair moderation practices aims to strike a balance between protecting democratic discourse and mitigating the risks of misinformation and polarization. By implementing policies that promote openness and user empowerment, digital platforms can better serve as spaces for constructive political engagement and democratic deliberation.

Italy:

The algorithms used by social platforms can create information bubbles or "filter bubbles". Users are encouraged to view mainly content that reflects their pre-existing beliefs and preferences, thus creating information bubbles that isolate them from different opinions or perspectives. This personalized selection of content can lead to the polarization of public opinion, as people may end up being exposed only to content that reflects their opinions, preventing them from encountering different and contrasting perspectives, and facilitating the spread of fake news.

Turkey:

Algorithms and content moderation significantly shape democratic transformation in Türkiye's digital spaces. Algorithms determine visibility, often reinforcing user biases through personalized content, which can create echo chambers and polarization (Akser, 2020). Effective moderation combats hate speech, misinformation, and harmful content, yet overly restrictive measures risk suppressing free speech (Özkula & Yıldırım, 2021). In Türkiye, balancing free expression with regulation is challenging, as platforms like Twitter implement strict content controls responding to local regulations, sometimes criticized as limiting political discourse (Yesil & Sözeri, 2018). Transparent moderation practices and algorithmic accountability are essential to maintaining a healthy digital democracy.

Portugal: No Answer

Luxembourg: No Answer

4. What challenges exist in balancing free speech with the need to curb hate speech and disinformation in digital public spaces?

France:

Balancing free speech with the imperative to curb hate speech and disinformation in digital public spaces is a complex challenge that involves legal, technological, and ethical

considerations. Ensuring that online environments remain open for diverse expression while mitigating harmful content requires nuanced policies and coordinated efforts.

Defining Hate Speech and Disinformation

(Vasist et al., 2024, p. 14). A fundamental difficulty in addressing hate speech and disinformation is establishing clear and universally accepted definitions. These interpretations vary across cultures and legal systems, making it challenging to develop standardized regulations. What constitutes hate speech in one country may be considered protected speech in another, leading to inconsistencies in enforcement. Similarly, defining disinformation is problematic, as distinguishing between harmful falsehoods and subjective political opinions is not always straightforward. This ambiguity complicates the development of effective regulatory frameworks.

Regulatory Frameworks: A Delicate Task

Crafting legal frameworks that effectively address harmful content while upholding freedom of expression is a delicate task. Overly restrictive laws may lead to censorship and the suppression of legitimate discourse, while insufficient measures might fail to protect individuals from harm, allowing harmful rhetoric and false information to proliferate. Governments must strike a balance between preventing abuse and preserving democratic values, which requires ongoing legal and ethical debates.

Platform Responsibility and Inconsistent Enforcement

Social media platforms play a central role in moderating content, yet their enforcement practices can be inconsistent. Policies often differ between platforms, and enforcement mechanisms may be influenced by corporate interests, public pressure, or political considerations. This inconsistency can lead to accusations of bias or arbitrary censorship, further eroding trust in digital platforms. Determining the appropriate level of responsibility for tech companies in content moderation remains a contentious issue, especially as private corporations become de facto arbiters of online speech.

Technological Challenges and Algorithmic Influence

The rapid evolution of digital platforms and the sophistication of disinformation campaigns make it difficult to keep pace with emerging threats. Many platforms rely on automated moderation systems that use algorithms to detect and remove harmful content. However, these algorithms can inadvertently promote sensational or polarizing content, including hate speech, as they are designed to prioritize engagement. This dynamic exacerbates ideological divides and contributes to the spread of harmful narratives, making effective moderation even more challenging.

The Global Nature of the Internet and Jurisdictional Issues

The borderless nature of the internet complicates enforcement efforts, as content can spread across multiple jurisdictions in an instant. National governments may struggle to regulate digital platforms that operate globally, leading to conflicts between local laws and international policies. Ensuring effective enforcement requires cooperation between governments, tech companies, and international regulatory bodies, but achieving consensus on best practices remains a significant challenge.

User Anonymity and Accountability

User anonymity is a double-edged sword in digital spaces. While it protects privacy and enables free expression, it also can embolden individuals to engage in harmful speech without fear of real-world consequences, making it difficult to hold perpetrators accountable. The challenge lies in developing solutions that protect users' rights while preventing anonymity from being exploited for abusive purposes. Some platforms have experimented with identity verification measures, but such approaches raise concerns about privacy and surveillance.

Impact on Public Discourse and Democratic Processes

The widespread presence of hate speech and disinformation has broader implications for public discourse. Individuals may feel discouraged from engaging in online discussions due to hostility, harassment, or the overwhelming presence of misleading information. This deterioration of online dialogue undermines democratic processes, as misinformation can shape public opinion and influence elections. A lack of trust in online discourse can also lead to disengagement from civic participation, weakening democratic engagement.

A Collaborative Approach for Effective Strategies

Addressing these challenges requires a collaborative approach that involves governments, technology companies, civil society organizations, and users. Effective strategies should include:

- Transparent content moderation policies to ensure fairness and consistency in enforcement.
- Enhanced digital literacy programs to help users critically assess online information.
- Stronger international cooperation to tackle cross-border challenges.
- Algorithmic transparency to mitigate the promotion of harmful content.

Italy:

The European Commission launched its Code of Conduct in May 2016 together with four major IT companies (Facebook, Microsoft, Twitter and YouTube), in an effort to respond to the proliferation of racist and xenophobic hate speech online. During 2018, Instagram, Snapchat and Dailymotion joined the Code of Conduct, Jeuxvideo.com joined in January 2019 and TikTok in September 2020. On 25 June 2021, LinkedIn also announced its participation. The aim of the Code is to ensure that requests to remove content are dealt with swiftly. When companies receive a request to remove content from their online platform that they believe is illegal, they assess the request against their rules and EU guidelines and, where appropriate, national laws, which implement EU law on combating racism and xenophobia. Companies commit to reviewing most of these requests in less than 24 hours, and to remove the content where necessary, respecting the fundamental principle of freedom of speech. To date, eight companies have signed up to the Code, namely Facebook, YouTube, Twitter, Microsoft, Instagram, Dailymotion, Snapchat and Jeuxvideo.com.

Turkey:

Balancing free speech with efforts to curb hate speech and disinformation poses several challenges in Türkiye’s digital spaces. Firstly, there's a constant tension between protecting

freedom of expression and implementing effective regulations to prevent online harm (Yesil & Sözeri, 2018). Over-regulation risks silencing legitimate dissent and can lead to accusations of censorship, while insufficient moderation allows harmful content to spread. Additionally, the subjective nature of defining "hate speech" complicates consistent enforcement, as interpretations vary among authorities, platforms, and civil society (Saka, 2018). Another challenge is the rapid spread and algorithmic amplification of misinformation, making timely intervention difficult without affecting free discourse.

Portugal:

General:

- In a democratic society, freedom of expression (or free speech) is a generally held principle that allows people to talk freely without fearing retaliation, censorship or government interference (Tanvir, 2022). Social media platforms and the online world have drastically transformed the dynamics of free speech and hate speech, enabling billions of people to express their opinions - either positive or negative- often without fully understanding the consequences of their words or actions. These platforms usually come under scrutiny for either allowing too much hate speech or for censoring too much content, creating a challenge to navigate this fine line between free expression and protecting their users from harm.

- Another possible challenge faced in balancing these two lines of conduct, is the challenge faced by legislators and regulators to create proper instruments that guarantee that content moderation processes and interventions are ‘’accountable both at an internal and external level, particularly when it comes to errors and harm’’ (Barata, 2024).

- Challenges can also be felt in the protection of basic human rights, such as the protection of personal data, the right to respect private life and the right to respect family life (Waal, et al., 2020).

Portugal:

- Recently, in Portugal, some cases of hate speech in the parliament have motivated political parties and citizens to call for a revision of the code of conduct for political representatives, advocating for the implementation of sanctions (Lusa, 2025). Centro da Vida Independente, an organisation that fights for the rights of people with functional impairment, has criticised the position of the president of the Portuguese parliament, as he allows for discrimination under the premise of freedom of speech (RTP, 2025). The normalisation of hate speech on the private and public spheres contributes to legitimisation of discrimination, violence and hate crimes (Casa do Brasil em Portugal, 2025).

- Comprehension of grey areas: The project kNOw Hate has developed a tool to detect and flag hate speech online, namely in YouTube and Twitter comments. Nevertheless, this tool only detects previously indicated words, without being able to signal out, comments or contents that might spread disinformation or hate covertly.

Luxembourg: No Answer

5. How do digital public spaces affect the inclusivity and accessibility of democratic participation?

France:

Digital public spaces play a crucial role in shaping the inclusivity and accessibility of democratic participation by providing platforms for diverse groups to engage in civic discourse and decision-making processes. These digital environments can lower traditional barriers to participation such as geographic constraints, physical disabilities, and institutional gatekeeping thereby fostering broader political engagement and more equitable representation. However, their effectiveness in promoting inclusivity depends on multiple structural, technological, and social factors.

Expanding Access to Marginalized Communities

One of the most significant contributions of digital public spaces is their ability to amplify the voices of marginalized communities that have historically been excluded from political discourse. A study by (Trevisan, 2022, p. 496-513) highlights how digital inclusivity in U.S. progressive politics has empowered underrepresented groups, including individuals with disabilities, to engage more actively in political processes. Online platforms enable these communities to organize, mobilize, and advocate for policies that address their specific needs, fostering greater civic participation.

Additionally, digital platforms help bridge geographical and logistical barriers that often limit political engagement. Individuals in rural or remote areas, for example, can participate in discussions, attend virtual town halls, and engage with policymakers without the need for physical presence. Similarly, people with disabilities benefit from assistive technologies that improve accessibility, such as screen readers, voice-to-text applications, and customizable user interfaces.

Digital Platforms as Modern Public Spaces

The digital sphere serves as a contemporary equivalent of traditional public forums, where people come together to exchange ideas and debate political issues. (Mazzoleni, 2015, p. 174-183) explores the concept of an inclusive digital public sphere, emphasizing how digital media fosters open discussions and facilitates the formation of public opinion. Online spaces allow individuals from different backgrounds to interact, share perspectives, and contribute to collective decision-making, thereby creating a more inclusive democratic environment.

Moreover, digital platforms can enhance direct engagement between citizens and political institutions. Governments increasingly use digital tools such as participatory budgeting websites, e-petitions, and virtual consultations to solicit input from the public. These mechanisms make governance more transparent and responsive to diverse societal needs, strengthening democratic accountability.

Challenges to Digital Inclusivity

While digital public spaces offer new opportunities for democratic participation, their inclusivity and accessibility are not guaranteed. Several factors influence their effectiveness, including:

- Platform Design and Accessibility Features: Many online spaces are not built with accessibility in mind, limiting participation for individuals with disabilities. Inadequate captioning, poor screen reader compatibility, and complex navigation structures can exclude certain users from political engagement.
- Digital Literacy and Technological Barriers: Effective participation in digital democracy requires a certain level of digital literacy. Individuals with limited technological skills may struggle to navigate online platforms, verify information, or engage in meaningful discussions. Additionally, disparities in internet access often referred to as the digital divide disproportionately affect low-income and rural populations, further restricting their involvement.
- Online Harassment and Disinformation: The presence of hate speech, disinformation, and toxic discourse can discourage participation, particularly among marginalized groups. Women, ethnic minorities, and LGBTQ+ individuals often face targeted harassment online, which can deter them from engaging in democratic discussions. Addressing these issues is essential to maintaining a safe and inclusive digital environment.

The Path Forward: Addressing Key Factors for Inclusive Digital Democracy

To fully realize the potential of digital spaces in enhancing democratic participation, policymakers, technology companies, and civil society must work together to address these challenges. Key strategies include:

- Improving platform accessibility by incorporating universal design principles and assistive technologies.
- Promoting digital literacy through educational initiatives that empower users to navigate online political spaces effectively.
- Enhancing protections against online harassment and misinformation through stronger moderation policies and community guidelines.
- Expanding internet access to underserved communities to ensure equal opportunities for civic participation.

Italy:

Today gender and age differences in using digital platforms or exclusion and discrimination may discourage democratic online participation….

Turkey:

Digital public spaces significantly enhance inclusivity and accessibility of democratic participation by providing platforms where diverse voices, including marginalized or underrepresented groups, can express opinions, engage in dialogue, and participate in political processes without traditional barriers (Hindman, 2018). In Türkiye, platforms like Twitter and Ekşi Sözlük enable broader public discourse and facilitate youth and women's participation in political conversations (Saka, 2018). However, digital divides due to socioeconomic factors or digital literacy can limit full participation, potentially deepening inequalities (Bulut & Yörük, 2017). Effective digital citizenship education and inclusive technology policies are thus critical in addressing these challenges.

Portugal:

General:

- Digital spaces have the potential to enhance the inclusivity and accessibility of democratic participation by making it easier for people to access information about political and current events around the world. However, the access to this information, doesn't come without its challenges. First and foremost, we need to consider that not all individuals can afford the technology and infrastructure to participate in digital democracy. While some upper- or middle-class families or communities might have better access to digital tools that enabled them to provide an active participation, poorer communities might not possess these advantages (Tewdwr-Jones, 2021).

Portugal:

- In Portugal, this topic is yet to be developed, nevertheless, the government has taken its first steps by developing the Participa.pt platform. By centralising and aggregating all governmental-related information, inclusion and accessibility towards democratic participation can be promoted.

Luxembourg: No Answer

6. What strategies are effective in fostering constructive dialogue and reducing polarization in online discussions?

France:

Reducing polarization and fostering constructive dialogue in online discussions requires a combination of linguistic strategies, psychological interventions, and structured educational programs. The goal is to create an environment where participants engage in meaningful discussions, respect differing perspectives, and work toward common ground rather than deepening ideological divides.

1. Subjective Phrasing: Softening Communication for Openness

The way people phrase their statements can significantly impact how their messages are received in online discussions. Incorporating subjective language, such as "I think" or "In my experience," helps make comments appear less confrontational and more constructive. This approach signals that the statement reflects a personal viewpoint rather than an absolute truth, reducing defensive reactions from others. Research suggests that subjective phrasing can soften disagreements, making participants more willing to consider alternative perspectives and engage in respectful dialogue. (Van Zomeren et al., 2008, p. 504).

2. Perspective-Taking: Enhancing Empathy and Understanding

One of the most effective strategies for reducing affective polarization the emotional division between opposing groups is encouraging individuals to actively take the perspective of others. This technique involves prompting users to reflect on their own experiences with disagreement before engaging in discussions. For example, studies have found that when individuals recall a disagreement with a friend before encountering opposing political views online, they become more empathetic and open to understanding different perspectives.

Additionally, exposure to personal stories from people with differing viewpoints can humanize political opponents, making it harder to dismiss or dehumanize them. Digital platforms can facilitate this by integrating storytelling features where users share experiences rather than just opinions, encouraging deeper emotional engagement and reducing hostility.

3. Educational Programs: Providing Tools for Constructive Engagement

Structured educational interventions can equip individuals with the skills necessary for engaging in civil discourse and resolving conflicts. Programs such as the "Perspectives" initiative have demonstrated success in reducing polarization by teaching participants how to engage in dialogue across ideological divides. These programs typically include:

- Conflict resolution training to help individuals de-escalate online arguments.
- Fact-checking exercises to encourage critical thinking and reduce misinformation-driven polarization (Lewandowskyet al., 2017, p. 353-369).
- Active listening techniques that promote mutual understanding.

4. Facilitated Workshops: Promoting Reciprocal Reflection and Active Listening

Facilitated workshops provide structured spaces for participants to engage in reciprocal group reflection and active listening, both of which are crucial for breaking down ideological barriers. These workshops can be conducted online through video conferencing platforms or community discussion forums, allowing for broad participation. (Tufekci, 2017, p.15).

Key components of effective facilitated workshops include:

- Personal storytelling: Participants share experiences related to divisive topics, fostering emotional connections.
- Guided discussions: Trained moderators help steer conversations away from confrontation and toward mutual understanding.
- Empathetic listening exercises: Participants practice paraphrasing others' points before responding, ensuring that each viewpoint is acknowledged.

Such interventions have been shown to reduce stereotypes, increase trust among participants, and promote more balanced discussions.

By implementing these strategies subjective phrasing, perspective-taking, educational programs, and facilitated workshops digital platforms and online communities can cultivate a healthier, more constructive discourse environment. (Kross et al., 2011, p. 187-191). Encouraging respectful interactions and reducing polarization is essential not only for improving online conversations but also for strengthening democratic engagement and social cohesion in the digital age.

Italy:

Particularly polarizing and divisive topics attract users. Acting effectively when disinformation is already in place and circulating is useless, and from a practical point of view, when a false news item is circulating, stopping it and correcting it is almost impossible. Furthermore, these corrective actions do not seem to work because they do not reach those who really need them and instead end up reinforcing the initial idea.

Instead of "fighting disinformation", we should "improve the quality of information in the system" through automatic identification systems for online discussions, taking into account different levels: violent, offensive or inappropriate language, which does not favor healthy communication. (rif.: https://ilbolive.unipd.it/it/news/disinformazione-fake-news-hate-speech-punto-azioni).

Turkey:

1. Encouraging Civil Discourse and Moderation:

Implementing clear moderation guidelines and community standards encourages respectful interactions. Platforms enforcing consistent moderation rules minimize hostility and promote civil discourse (Gillespie, 2018, p. 5).

2. Structured and Moderated Online Discussions:

Structured dialogue platforms (e.g., Kialo) or moderation frameworks allow balanced argumentation and reduce polarization by focusing on evidence-based arguments (Gurin et al., 2013, p. 13).

3. Media Literacy Education:

Media literacy training helps individuals critically evaluate information, reducing susceptibility to misinformation and polarization by promoting analytical thinking (European Commission, 2021).

4. Counterspeech and Positive Engagement:

Encouraging counter-narratives ("counterspeech") rather than censorship can neutralize hate speech and misinformation, fostering healthier discourse (Benesch, 2014, p.18-25).

Different digital platforms uniquely influence public opinion and democratic engagement based on their structures, user demographics, and content dissemination methods.

Social Media (Twitter, Facebook, Instagram):

Social media platforms strongly shape public opinion by rapidly spreading information, influencing political mobilization, and facilitating direct communication between citizens and politicians. Their algorithms can reinforce polarization through selective content exposure, amplifying extreme views (Cinelli et al., 2021, p. 8).

Online Forums (e.g., Reddit):

Forums support detailed and specialized discussions, often promoting niche or marginalized perspectives. However, due to anonymity and community moderation, they may also foster echo chambers or misinformation (Massanari, 2017, p. 21-22).

Messaging Apps (WhatsApp, Telegram):

Messaging platforms facilitate private group communication, fostering rapid mobilization and grassroots activism. However, the lack of effective moderation can quickly spread misinformation and propaganda (Resende et al., 2019,p 4-6).

Video Platforms (YouTube):

Video platforms affect public opinion significantly through recommendation algorithms that often lead users toward polarized or extremist content, impacting democratic discourse negatively (Ribeiro et al., 2020, p. 9).

Portugal:

General:

- Promote positive engagement: Some suggestions to foster constructive dialogue could be to promote positive engagement. Platforms could achieve this goal by enforcing rules that encourage positive interactions, ensuring that conversations are respectful and productive. They could also foster active listening, asking insightful questions and making users express their views in a respectful manner (CoExist, 2025).

- It is also important for social media platforms to take an active role in managing and curating content to ensure that accurate and reliable information is shared (CoExist, 2025).

Portugal:

- Human rights framework and education: Following the Manual para o 'Combate do discurso de ódio online através da educação para os direitos humanos' (Manual for the combat of online hate speech via a human rights education), there are different games and activities that could be used with youth to mitigate polarising discourses online, including fact checking games (Keen, 2016).

Luxembourg: No Answer

7. How do different digital platforms vary in their influence on public opinion and democratic engagement?

France:

Digital platforms play a crucial role in shaping public opinion and democratic engagement, with their impact varying based on platform design, algorithms, and user interactions. Each type of platform whether social media, search engines, or public service media has distinct mechanisms that influence how people access, interpret, and engage with political information.

1. social media: Rapid Information Dissemination and Political Mobilization

Social media platforms like Facebook, Twitter, and Instagram facilitate the fast spread of information, enabling users to participate in political discussions, share news, and organize collective action for various causes. These platforms enhance engagement by providing spaces for activism, citizen journalism, and direct interaction with political figures. However, their influence on democratic participation is not always positive.

- Misinformation and Echo Chambers: social media can also amplify misinformation and partisan content, creating echo chambers where users are primarily exposed to viewpoints that align with their existing beliefs. This limits exposure to diverse perspectives and reinforces political polarization.

- Algorithmic Prioritization: The design of social media algorithms often favors sensational and emotionally charged content, increasing the likelihood that false or misleading information will spread widely. This can distort public perceptions of political events and undermine informed decision-making in democratic processes. (Joint Research Centre, European Commission)

2. Search Engines: Expanding Access to Information and Political Knowledge

Unlike social media, search engines such as Google provide users with access to a broader range of political information, allowing them to research different perspectives and make informed political choices. Studies suggest that search engines contribute to political literacy and influence voter behaviour, particularly in direct democracy settings.

Enhancing Political Knowledge: By providing users with fact-based sources, search engines can help counter misinformation and promote informed electoral decisions.

Algorithmic Bias and Selective Exposure: However, search engine results are also influenced by ranking algorithms, which can shape users' political understanding by prioritizing certain sources over others. Personalization features may further reinforce selective exposure, subtly shaping public opinion. (Routledge.com)

3. Public Service Media: Reliable Information and Balanced Discourse

Public service media platforms aim to provide fact-checked, balanced news coverage, supporting an informed and engaged citizenry. Unlike commercial media, their objective is to serve the public interest rather than maximizing engagement.

- Contributing to Informed Public Discourse: By prioritizing journalistic integrity, these platforms counteract the spread of misinformation and promote media pluralism.
- Challenges and Limitations: The effectiveness of public service media can be hindered by funding constraints, political pressures, and competition from commercial digital platforms. In some cases, government influence may compromise editorial independence, affecting the credibility and neutrality of the information provided.

The interplay between social media, search engines, and public service media shapes public opinion and democratic participation in complex and multifaceted ways. While digital platforms provide unprecedented access to political information and engagement opportunities, they also introduce challenges related to misinformation, algorithmic influence, and political bias. Developing media literacy skills and encouraging critical engagement with digital content are essential to ensuring that these platforms contribute to a well-informed and active democratic society.

Italy:

The platforms provide opportunities to express opinions, connect with other engaged citizens and organize protest movements. Digital platforms playing a crucial role in the dissemination of news and allowing people to organize and share their experiences in real time. These movements often develop and spread rapidly through online social networks, involving large numbers of people in a short time.

Turkey: **No Answer**

Portugal:

- Digital platforms play a significant role in shaping public opinion on political issues. A study dedicated to analysing people's views on platforms conducted with data from 8 countries has shown that search engines are the most trusted source of news and information about democratic life, with video network platforms usually in second place (Ejaz, Fletcher, Nielsen, & McGregor, 2024).

- In Portugal, most youngsters (16-30 years old) still based their public and political opinions on the news they get from TV, radio and online newspapers, rather than through social media. For those below 16, social media prevail as the main source of information (Monteiro, 2025).

Luxembourg: No Answer

II. Analytical Approach – Theories, Methods, and Tools

a) Disinformation and Media Literacy

1. What theories and frameworks have been used in your country to analyze disinformation in the digital age?

France:

In France, several theories and analytical frameworks have been employed to understand and address disinformation in the digital era. These frameworks focus on categorizing different types of information disorder, examining their emotional and ritualistic aspects, and utilizing advanced technological methods to track and counteract disinformation.

1. The Council of Europe's Information Disorder Framework

One of the most widely used interdisciplinary frameworks in France is the Council of Europe's (CoE) Information Disorder Framework, which classifies information disorder into three distinct types:

- Misinformation: False or misleading information spread without the intent to deceive (e.g., sharing incorrect information unknowingly).
- Disinformation: False information deliberately created and spread with the intent to deceive (e.g., propaganda or fake news campaigns).

- Misinformation: Genuine information shared with the intent to cause harm (e.g., leaking private data to damage reputations). (coe.int)

This framework examines disinformation through three key phases:

- Creation: Who is producing the content and why?
- Production: How is the message shaped and framed?
- Distribution: How does the content spread and who engages with it?

Additionally, the framework emphasizes the emotional and ritualistic elements of disinformation, arguing that content designed to evoke strong emotions (fear, outrage, or shock) spreads more rapidly due to human psychological tendencies. Understanding these dynamics is crucial for designing strategies to combat the spread of false information. (coe.int)

2. The Commission Bronner's Definitions of Information Disorders

The Commission Bronner, a specialized French governmental body, has further refined these definitions by developing clear distinctions between different types of problematic information. These definitions aim to enhance precision in policy-making and media literacy initiatives. (fr.wikipedia.org)

The commission defines:

- Misinformation: Incorrect or misleading information spread without malicious intent.
- Disinformation: False information deliberately spread to manipulate public opinion.
- Malinformation: Authentic but damaging information spread to harm individuals or institutions.

This nuanced approach allows policymakers, media professionals, and researchers to better identify and address different types of information disorders, leading to more effective counter-disinformation measures in digital spaces. (fr.wikipedia.org)

3. Technological and Behavioural Approaches: Social Bots and Cognitive Modelling

Beyond theoretical frameworks, technological studies have also played a significant role in analysing disinformation in France. A study by (Emilio Ferrara, 2017, p.5) examined the role of social bots in amplifying disinformation during the French presidential election. (arxiv.org)

This study applied machine learning algorithms and cognitive behavioural modelling to:

- Differentiate between human users and automated bots spreading false information.
- Analyse patterns of coordinated disinformation campaigns on social media.
- Understand how disinformation manipulates public perception and political discourse.

By providing insights into the mechanics of digital disinformation, this research has informed efforts to detect and mitigate automated influence campaigns in the French media ecosystem.

A Multi-Layered Approach to Combating Disinformation

France employs a diverse set of theories and frameworks to analyze and address disinformation, including:

- The Council of Europe's categorization of information disorder.
- The Commission Bronner's definitions, which provide a refined understanding of misinformation, disinformation, and malinformation.
- Technological studies using AI and behavioral modeling to track disinformation dynamics.

This comprehensive and interdisciplinary approach offers valuable insights into the spread, impact, and mitigation of digital disinformation, enabling France to develop more effective policies, educational programs, and technological solutions to safeguard democratic discourse.

Italy:

Disinformation is complex and difficult to define because we are dealing with an all-encompassing phenomenon. Recent literature addresses the need for other terms that better reflect the problem's many different facets, such as "misleading information" (Fallis 2015) or "information disorder" (Wardle & Derakhshan 2017) and can circumscribe its perimeter: the former, for instance, distinguishes between misinformation and disinformation; the latter between misinformation, disinformation and mal information. According to Fallis (2015), the concept of misleading information can be broken down into two categories: disinformation, defined "as a bit of false information deliberately aimed at deceiv[ing]" (Giglietto et al. 2016: 5), and misinformation, "intended as a false or inaccurate information circulating as a result of honest mistakes, negligence or unconscious biases" (Ibid.). However, while Giglietto and colleagues' study (2016) recognizes the distinction between misinformation and disinformation, it includes both in a "meta category" that takes a more markedly process-oriented perspective. In turn, the term "information disorder" comprises and distinguishes between misinformation and disinformation, but adds mal-information, which is "when genuine information is shared to cause harm, often by moving information designed to stay private into the public sphere" (Wardle & Derakhshan 2017). The literature has also introduced the "Infodemic" (Cinelli et al. 2020), comparing the diffusion of fake news to the dynamics of viral infection (Tambuscio et al. 2015).

In the recent Italian literature, "Disinformation" is generally used to refer to the deliberate creation and/or spread of false or misleading information. The conceptual framework of disinformation operations argues that disinformation is best understood as a top-down, centrally coordinated intervention that relies on tactics of systemic deception to mislead the public for a strategic, political purpose. Disinformation campaigns can also occur within different degrees of coordination between multiple, "networked factions" of disinformation agents. In planning a campaign, disinformation agents tap into existing trauma within the public, and extract, weaponize, and manipulate facts with the addition of half-truths, exaggerations, or lies. Hence, essentially, disinformation operations are attacks against a liberal epistemic order that places its trust in essential custodians of factual authority (e.g., political and scientific institutions). Pasquetto et al. (2022) define disinformation infrastructure as a multi-layered, relational and distributed sociotechnical entity that supports the maintenance and distribution of information that is false or misleading. Their research is rooted in an alternative line of work within

disinformation studies that adopts a sociotechnical approach to the study of disinformation. This approach sees disinformation as the product of the participatory work of disinformation agents who create, share, and amplify disinformation content via ad hoc tactics of manipulation. Hence, disinformation is neither simply nor solely passively received by audiences, but it is through the proactive and communal work of specific and highly committed disinformation agents that disinformation can "trade up the chain" of the digital media ecosystem, reach audiences across multiple platforms, and find a loyal, sympathetic public.

Recent Italian literature analyses different categories of disinformation and distinguishes their boundaries. In particular: 'fabricated content', when content is created ad hoc and therefore completely false, with no relation to reality; 'manipulated content', when true information and images are distorted in order to deceive; 'impostor content', when false content is presented as an authentic source; 'false context', when true content is disseminated out of context, making the overall information false. These types of content carry a high risk of disinformation. A risk that is reduced when the user comes into contact with: 'misleading content', i.e. when true information is misused to frame an issue or a person; 'false context', when the title, text or image used is out of context (Acampa, 2022). The LUISS University research on "Identifying and Countering Coordinated Disinformation Operations" (Ministero degli Affari Esteri, 2023) examined the complex dynamics of communities involved in spreading false information in Italy. According to the research, the groups that organise and disseminate disinformation campaigns in Italy are well organised from a technical and infrastructural point of view. They create real infrastructures of digital disinformation, mainly based on social pages and accounts, and then include a whole series of websites, news aggregators, databases, alternative disinformation channels, blogs, forums, etc. It is also noted that within the disinformation infrastructure, blogs seem to occupy an increasingly prominent position, perhaps in response to the anti-disinformation measures implemented by platforms and required by the new international legal landscape. Indeed, blogs remain one of the few online spaces that are not moderated or censored. Disinformation infrastructures are run by disinformation agents (or disinfluencers), such as influencers and public figures, supported by an army of loyal followers interested in the cause. The collaboration and active participation of followers is essential for the dissemination and maintenance of the disinformation infrastructure over time. The Italian context began to prioritise the issue of disinformation after the notorious «Facebook-Cambridge Analytica» scandal. However, despite empirical evidence demonstrating the dangers of disinformation, the Italian political system has not adopted regulatory measures to address the issue (Barbieri, 2024).

Turkey:

In Türkiye, several theories and frameworks are employed to analyze disinformation in the digital age. A prominent approach is the Information Disorder Framework, which classifies problematic information into misinformation (unintentional false information), disinformation (intentionally misleading content), and malinformation (true information intended to harm) (Wardle & Derakhshan, 2017). Additionally, analytical studies often use content analysis methods to understand narratives, ideological biases, and patterns in disinformation (Kırdemir, 2022). Türkiye also utilizes AI-driven analysis to track disinformation campaigns, assessing their global implications for security and diplomacy (Özdemir & Şeker, 2023);

Portugal:

- Portuguese Regulatory Authority for the Media (ERC): This entity was established by the Law No. 53/2005 from the 8th of November, and it is responsible for regulating and monitoring all entities carrying out media activities in Portugal.

- Portuguese Charter on Human Rights in the Digital Age (Carta Portuguesa de Direitos Humanos na Era Digital): This charter aims to ensure the respect for the rights, freedoms and guarantees of citizens in the media space. It was also implemented to ensure the effective implementation of the European Action Plan against Disinformation, to foster a collective response and to promote cooperation among member states to tackle this issue (Oliveira et al., 2024).

Luxembourg:

Switzerland Klopfenstein Frei, N., Wyss, V., Gnach, A., & Weber, W. P. 100-121 (2024) https://doi.org/10.1177/14648849221123385

Abstract: News media in Switzerland are confronted with the challenge of not reaching young people, as youth-specific news platforms and formats are not used by the target group. Our study aims to determine how and where young people can be reached with news. It uses a mixed methods approach to analyze the expectations of young people towards news content and formats and to determine their news consumption patterns. The results show that young people's news consumption is characterized by four dimensions: 1. duration and times of consumption, 2. news consumption habits and behavior, 3. restrictions, 4. media literacy. The significance of these dimensions varies between three age groups determined through the study. The news consumption of the 12-to-14-year-olds is strongly restricted by parents and school. This group consumes news mainly through media available at home. In the group of 15- to-17-year-olds, parental influence and restrictions decrease, while peer influence increases. This age group spends a lot of time on social media platforms where young people stumble upon news rather accidentally. Between 18 and 20 years of age, news consumption stabilizes, and individual patterns emerge. This age group accesses news via selected apps and social media channels. Young people in Switzerland prefer visual formats like pictures, videos, and memes. When scrolling through social media platforms, they come across news content which arouses their interest in certain information and leads them to search for it on media websites. Swiss Youth want to be informed about relevant news or topics that are "in vogue".

2. How can the effectiveness of media literacy programs in combating disinformation be evaluated?

France:

Evaluating the effectiveness of media literacy programs in combating disinformation requires a multifaceted approach that incorporates various research methodologies to assess both immediate and long-term impacts. The following key strategies are commonly employed:

- Pre- and Post-Intervention Assessments: One of the most direct methods to evaluate the effectiveness of media literacy programs is through pre- and post-

intervention assessments. These assessments typically involve surveys or tests conducted before and after the educational intervention to measure changes in participants' ability to identify and critically analyze misinformation. A significant improvement in scores post-intervention would suggest that the program has enhanced participants' media literacy skills. Studies such as those by (Cetroet al., 2020, p. 5) and (Grosbois,2022, p. 8) highlight the importance of structured assessments in tracking progress and refining educational approaches.

- Behavioral Tracking: Beyond self-reported surveys, behavioral tracking provides deeper insights into how media literacy education influences real-world actions. This method involves monitoring participants' media consumption patterns, fact-checking behaviors, and the frequency with which they share unverified information. A decline in the dissemination of false or misleading content post-intervention would be a strong indicator of the program's success. Research by (Lukasik et al., 2023, p. 20) and (Ostian,2019, p. 8) underscores the role of digital platforms in amplifying disinformation and the necessity of tracking engagement metrics to assess behavioral changes.
- Longitudinal Studies: While short-term assessments provide valuable insights, longitudinal studies are crucial for understanding the lasting effects of media literacy education. These studies follow participants over an extended period to evaluate whether the skills acquired through the program are retained and continue to influence their ability to critically engage with media content. Findings from (Géré, 2018, p. 9) and (Fries, 2021, p. 3) emphasize the importance of sustained interventions to counteract the evolving nature of disinformation campaigns.
- Control Groups and Experimental Designs: To establish a causal relationship between media literacy interventions and changes in participants' ability to combat disinformation, researchers often employ experimental designs that include control groups. By comparing individuals who have undergone media literacy training with those who have not, researchers can isolate the specific effects of the intervention. (Piguet, 2022 p. 16) and (Monneret, 2020, p. 11) discuss the effectiveness of controlled experimental settings in measuring cognitive and behavioral shifts related to misinformation resilience.
- Recent research supports the efficacy of these evaluation methods. A meta-analysis by Vraga and (Tully, 2021, p. 422-448) found that media literacy interventions generally improve resilience to misinformation, with a significant effect size (d = 0.60). Additionally, (Guess et al, 2020, p. 15536-15545) demonstrated that digital media literacy training enhances individuals' ability to differentiate between mainstream and false news in both the United States and India. Furthermore, (De Montbrial et al., 2017, p.14) discusses the broader implications of the information war and emphasize the necessity of structured interventions to counteract disinformation efforts at a societal level.

Italy:

It is necessary to organise interventions against disinformation, such as deplatforming. If deplatforming is not done in a timely manner, it will have a limited effect in eradicating such groups and limit the effectiveness of campaigns. Deplatforming is never a definitive solution, but must be carried out in an iterative manner, as disinformation groups develop precise tactics and strategies that allow them to adapt to new technological conditions from time to time. Traditional media can also play a central role in preventing and reducing disinformation. On the one hand, through the adoption of strategic silence by the media, a tactic of journalistic coordination that involves a conscious recognition of the media's role in amplifying disinformation campaigns and a consequent decision not to give visibility to such campaigns run by disinformation groups. Secondly, the development of a journalistic curriculum aimed at educating new generations of journalists on how disinformation works online is recommended.

In order to better understand the phenomenon in the long term and to develop appropriate technical and regulatory countermeasures, it is also necessary to promote exchanges and cooperation between universities, public institutions and private companies by allocating structural funds for the creation and maintenance of national and international research and discussion groups, both formal and informal (Ministero degli Affari Esteri, 2023).

Turkey:

1. Pre- and Post-Assessment Tests:

Evaluating participants' critical thinking skills and their ability to identify misinformation before and after training (Bulger & Davison, 2018).

2. Behavioral Change Analysis:

Monitoring whether participants apply learned skills in real scenarios, such as checking sources and fact-checking information (Huguet et al., 2019).

3. Surveys and Interviews:

Collecting qualitative feedback on attitudes and confidence levels regarding disinformation recognition after the program (European Commission, 2021).

4. Longitudinal Studies:

Tracking the long-term retention of media literacy skills to understand sustained impact on behaviour (Livingstone & Helsper, 2007).

Portugal:

- Available data was not found to answer this question. However, the effectiveness of media literacy programs could be evaluated by utilising a combination of quantitative and qualitative methods that are employed for other types of programmes, courses or interventions. For instance, pre- and post-questionnaires have proved effective in analysing the effectiveness of other types of programmes. By comparing post-testing outcomes with the baseline data from the pre-testing, researchers and investigators can assess whether a given intervention or programme has produced a desired impact, helping to assess its effectiveness (Majka, 2024).

Luxembourg:

EDMO Guidelines for Effective Media Literacy Initiatives

https://edmo.eu/areas-of-activities/media-literacy/raising-standards/

This set of principles and guidelines for effective media literacy initiatives has been developed by the EDMO Council on Effective Media Literacy Initiatives with input from the EDMO Hubs and other practitioners and experts during a consultation process which included a public workshop and an open call for written feedback. Its goal is to increase the effectiveness of media literacy initiatives across Europe by developing guidelines that new and existing practitioners can consult. This is crucial to EDMO's wider mission, since better media literacy is likely to increase the public's resilience to online mis- and disinformation.

These guidelines have been designed for use by anyone involved in the development of media literacy initiatives – civil society, educators, policy makers, those in the media or tech industries. In an area as complex and diverse as media literacy, there can be no one-size-fits-all approach. Media literacy initiatives – such as training courses, published lesson plans or other resources, online games, or campaigns – vary considerably in scope, size, duration and focus. As a result, not all of these guidelines will be relevant to all projects. It is entirely up to the user to select the most appropriate advice for their initiative.

In an area as complex and diverse as media literacy, there can be no one-size-fits-all approach. It is entirely up to the user to select the most appropriate advice for their initiative.

While EDMO's focus is disinformation, the guidelines can be applied widely across diverse media literacy initiatives. For example, these might include news literacy initiatives, which will have a greater focus on the value of independent journalism in the media ecosystem; wider digital literacy initiatives, which will focus on explaining how digital media operate and how to use them; algo-literacy initiatives, which will focus on understanding the role of algorithms in our media consumption, or many more.

The guidelines checklist is intended to be helpful in raising issues for consideration, based on the experiences of prior media literacy initiatives. It should not become a 'tick box' exercise or be used to limit flexibility in developing future initiatives. The Guidelines will be reviewed on an annual basis, to propose changes to the text, please click here.

We have grouped the guidelines under 12 principles, and grouped these into three sections: development, delivery and review.

Germany

Meyer, F., Becker, J., & Bock, A. P. 118–138 (2023). https://doi.org/10.21240/mpaed/00/2023.06.01.X

Germany

Gerke, A., & Wegner, C. P. 61-68 (2022). file:///C:/Users/quint/Downloads/2022_3_15.pdf

3. What methodologies, strategies or plans are employed to identify and trace the sources of disinformation?

France:

1. Machine Learning and Artificial Intelligence (AI) Algorithms:

AI-powered tools are increasingly used to detect and analyze disinformation patterns. Machine learning algorithms process vast amounts of online data to identify inconsistencies, repeated falsehoods, and synthetically generated content. These models utilize natural language processing (NLP) techniques to recognize misleading narratives, linguistic patterns associated with disinformation, and automated content generation, such as deepfake videos and AI-generated text.

Example: Facebook and Twitter have implemented AI-driven systems to detect and flag misinformation. For instance, Facebook's AI identified and removed fake accounts linked to Russian interference in the 2016 U.S. presidential election.

2. Fact-Checking and Verification Tools:

Dedicated fact-checking organizations and media outlets employ human fact-checkers in conjunction with automated verification tools. Platforms such as PolitiFact, Snopes, and FactCheck.org work to assess the accuracy of viral claims, cross-referencing information with credible sources. Automated tools, including reverse image search engines like Google Lens and TinEye, assist in tracking the origins of manipulated images and videos.

Example: During the COVID-19 pandemic, fact-checkers debunked widespread false claims, such as misinformation about vaccine ingredients and efficacy.

3. Network Analysis and Social Media Mapping:

Disinformation often spreads through coordinated campaigns across social media platforms. Network analysis helps track how false narratives originate and disseminate through digital ecosystems. By examining metadata, retweet patterns, and engagement clusters, researchers can identify bot networks, troll farms, and coordinated inauthentic behaviours. Social network analysis tools like Gephi and NodeXL facilitate the visualization of disinformation networks and the identification of key influencers driving misinformation campaigns.

Example: The "troll factory" operated by the Internet Research Agency (IRA) in Russia used coordinated social media campaigns to spread divisive political content in the U.S. and Europe.

4. Crowdsourcing and Open-Source Intelligence (OSINT):

Crowdsourced efforts, such as citizen journalism and public reporting platforms, contribute to the identification of disinformation. Open-source intelligence (OSINT) methods involve analyzing publicly available data, including social media posts, domain registrations, and leaked documents. Investigative organizations, such as Bellingcat, leverage OSINT techniques to trace disinformation campaigns and expose their sources.

Example: Bellingcat exposed Russian involvement in the downing of Malaysia Airlines Flight MH17 by analyzing social media posts and satellite imagery.

5. Digital Forensics and Metadata Analysis:

Digital forensic techniques are used to trace the authenticity and origin of disinformation content. Metadata embedded in images, videos, and documents can reveal crucial details about their creation, including timestamps, geolocation, and device information. Tools like InVID and FotoForensics help analyze the integrity of multimedia content and detect potential manipulations.

Example: Journalists used metadata analysis to confirm that images from the Syrian Civil War were being misrepresented and repurposed to spread disinformation.

6. Tracing Financial and Political Incentives:

Disinformation campaigns are often motivated by financial or political interests. Investigative journalists and researchers examine financial transactions, advertisement spending, and sponsorship records to trace funding sources behind disinformation networks. Political affiliations and lobbying efforts linked to misinformation campaigns are also scrutinized to uncover vested interests in shaping public discourse.

Example: The Cambridge Analytica scandal revealed how political consulting firms harvested Facebook data to influence voter behavior through targeted misinformation campaigns.

7. Collaboration with Governments, Tech Companies, and NGOs:

Governments, technology companies, and non-governmental organizations (NGOs) play a crucial role in countering disinformation. Policy measures, regulatory frameworks, and public awareness campaigns contribute to mitigating the spread of false information. Social media platforms deploy moderation algorithms, while independent organizations advocate for transparency and digital literacy initiatives.

Example: The European Union established the "EUvsDisinfo" initiative to counter Russian disinformation campaigns targeting European elections.

Case Study and Research Reference:

Research conducted by Friggeri, Adamic, Eckles, and Cheng, p. 14 (2014) explores the phenomenon of "rumor cascades," demonstrating how false information spreads across social networks and the effectiveness of intervention strategies. Their study highlights the importance of monitoring engagement patterns and employing real-time detection techniques to curb disinformation proliferation.

Italy:

Overall, experts warn that the absence of a clear legal framework in Italy let platforms adopt very different and often ambiguous approaches to how they intend to counter disinformation. The Italian case thus appears to lack regulatory instruments in the counter-disinformation policy, due to a fragmented political apparatus unable to tackle the issue in the appropriate institutional fora (Di Mascio et al. 2021). Previous attempts to implement regulation failed for fears that they would give authorities too much power, such as the bill to create a legal

framework against disinformation and hate speech in 2017 and a protocol for the fight against the spread of fake news via the web during the 2018 general elections. In recent years, the Italian penal code introduced article 612-ter against the illicit dissemination of sexually explicit images or videos (revenge porn) and updated article 612-bis to include formal recognition of cyber-stalking as a crime. The latter constitute minimal legal protection against gender-based disinformation.

The Italian model of media accountability is characterized by the strong role of Italy's independent authority (AGCOM), which has been maintained in the fight against disinformation. At the same time, Italy has shifted toward a coregulation model of social platform accountability, as AGCOM is proceeding with a multistakeholder logic. Social platform operators are invited to join AGCOM in monitoring disinformation and developing a strategy to prevent it. As such, social platforms are conceived as allies in this fight; together with several nongovernmental organizations, they participated in a roundtable—"Pluralism and Platforms"—set up by AGCOM to examine the path forward. The authority has not been left completely alone in this task, as it has partnered with experts and market players, but has also not been supported by the Italian government (which was experiencing turbulence in the aftermath of the March 2018 elections) or parliament. This situation changed somewhat during the Covid-19 pandemic, when the "infodemic" has been regulated by a partnership between the Ministry of Health, Ministry of Education, and Prime Minister's Office and Google, Facebook, and WhatsApp. In the power vacuum that characterized Italy until March 2020, AGCOM published guidelines for political parties rallying for the 2018 elections, extending the legislation regarding the fair representation of political forces that was already applicable to legacy media to online platforms. However, this decision mainly addressed political parties and organizations, rather than platforms as editors (De Blasio & Selva, 2021).

The main actors countering disinformation in Italy are:

- Fact-Checkers

Facta, news, Pagella Politica (specialised in debunking statements by politicians and parties), La Voce (focussing on economy-related content), and Open.online are verified signatories of the IFCN code of principles and Meta's independent fact-checking partners. Other relevant news verifiers are Bufale.net, BUTAC, Smask.online (exclusively exposing disinformation from the League), Giornalettismo's and PolicyMakerMag's fact-checking sections.

- Media Literacy

Italian Digital Media Observatory

- Government Initiatives

During the Covid-19 pandemic (2020-2022), the Ministry of Health compiled a dedicated portal on health-related "fake news"; AGCOM (the national regulator and competition authority for the communication industries) started an observatory on online disinformation, which has not updated since 2020.

Turkey:

In Türkiye, various methodologies and strategies are employed to identify and trace the sources of disinformation. Prominent fact-checking organizations such as Teyit.org and Doğruluk Payı

use rigorous methodologies including reverse-image searches, content verification, and network analysis to identify misinformation and its origins. Collaborative platforms like Teyit.org apply systematic fact-checking procedures to trace disinformation by examining original content sources and distribution pathways. Furthermore, initiatives by digital rights groups and academic institutions in Türkiye, such as the Istanbul Bilgi University's Digital Rights Platform, employ data-driven analysis and machine learning algorithms to map and monitor disinformation flows effectively.

Portugal:

General:

- Some strategies can be used to identify and trace sources of disinformation, namely:
  - The implementation of robust fact-checking mechanisms such as utilising organisations like FactCheck.org and Snopes.org to verify information and debunk false claims (Zacky & Zacky-Eze, 2025).
  - Promoting education campaigns and programmes to increase digital literacy has proven, in previous research, to have been crucial in reducing the sustainability of disinformation (Pennycook & Rand, 2019). Educational programmes can cover topics such as identifying credible sources, recognising bias and understanding the impact of media (Ashikuzzaman, 2024).
  - Machine learning algorithms and artificial intelligence (AI) have become essential tools in identifying and analysing disinformation patterns (Zacky & Zacky-Eze, 2025).

Portugal:

- In Portugal, the most common fact-checks mechanisms are 'Poligráfo', the first Portuguese journal of fact-checking whose main goal is to verify the truth by actively scrutinising the various protagonists on digital platforms; 'Viral', a fact-checking newspaper fully dedicated to checking misinformation about health on social media platforms; and 'Iberifier', an observatory dedicated to fighting disinformation in Spain and Portugal (Pardar, 2023).

Luxembourg: No Answer

4. How can the impact of media literacy initiatives on public awareness and behavior be measured?

France:

To measure the impact of media literacy initiatives on public awareness and behavior, researchers often use a combination of quantitative and qualitative methods. These can include surveys, pre-and post-assessments, focus groups, and behavioral tracking. Surveys assess changes in knowledge, attitudes, and behaviors, while focus groups provide in-depth insights into how media literacy interventions influence critical thinking and decision-making. Behavioral tracking can measure changes in actual media consumption patterns or attitudes toward media sources. Additionally, experimental designs like control groups and intervention

groups are frequently used to compare outcomes and identify specific effects Levine et al., 2009, p. 9-42).

Italy:

Pressures for action against disinformation have been particularly intense in Italy where there is one of the highest percentages of citizens in Europe who deem disinformation a problem for their country and are most likely to say that national authorities should tackle disinformation (Eurobarometer 2018). However, demand for action against disinformation did not translate into platform accountability regulated by law as it happened in France and Germany before the pandemic. Rather, the Italian model of platform accountability has been characterized by the strong role of the media regulator (Autorità per le Garanzie nelle Comunicazioni – AGCOM), which encouraged self-regulation by platforms in the absence of regulatory measures taken by the government or parliament (De Blasio et al., 2021, p. 5).

Turkey:

The impact of media literacy initiatives on public awareness and behavior can be measured through several methodologies. Pre- and post-intervention assessments evaluate participants' critical thinking and media analysis skills before and after educational activities, revealing tangible improvements (Hobbs & Frost, 2003). Surveys and questionnaires assess changes in attitudes, awareness, and practices concerning misinformation and media consumption (Livingstone & Helsper, 2007). Additionally, behavioral observations and focus group discussions can provide qualitative insights into behavioral shifts, while longitudinal studies capture sustained behavioral changes and long-term effectiveness (European Commission, 2021).

Portugal: No Answer

Luxembourg: No Answer

5. What are the main challenges in educating the public about recognizing fake news and disinformation on young adults?

France:

Educating young adults on recognizing fake news and disinformation presents numerous challenges, given the evolving nature of digital media and cognitive biases that affect information processing. These challenges can be categorized into technological, psychological, educational, and legal dimensions.

1. Pervasiveness of Digital Media and Information Overload

The widespread use of digital platforms has facilitated an unprecedented volume of information exchange, making it difficult for young adults to differentiate between credible and misleading sources. Social media algorithms often prioritize engagement over accuracy, leading to the rapid spread of sensationalized and misleading content. Young adults, who are among the most active users of social media, are particularly vulnerable to consuming and sharing misinformation without critical evaluation (Friggeri, et al., 2014, p. 101-110).

Example: Studies show that during major events such as elections or crises, false news spreads faster than factual news, making it harder for individuals to discern truth from fiction.

2. Sophistication of Fake News and Algorithmic Manipulation

Disinformation strategies have evolved with the use of AI-generated content, deepfake technology, and synthetic media. These sophisticated techniques make it more difficult to distinguish between authentic and manipulated content. Additionally, search engine and social media algorithms create "filter bubbles," reinforcing pre-existing beliefs by exposing users to information that aligns with their views.

Example: Deepfake videos have been used in political campaigns to distort public perception by showing fabricated statements from politicians.

3. Cognitive Biases and Psychological Barriers

Human cognition plays a crucial role in how young adults process information. Biases such as confirmation bias (favouring information that supports pre-existing beliefs) and the illusory truth effect (believing repeated misinformation) contribute to the persistence of fake news.

Example: Research has shown that individuals are more likely to believe a false claim if they have seen it multiple times, even if they were initially sceptical.

4. Lack of Comprehensive Media Literacy Education

Despite the increasing prevalence of digital content, media literacy education is often insufficient or inconsistently integrated into school curricula. Many young adults do not receive formal training on critical thinking, source verification, or fact-checking techniques.

Example: In some countries, media literacy is not a mandatory subject, leaving students without the necessary skills to critically analyze information they encounter online.

5. Resistance to Fact-Checking and Misinformation Fatigue

Fact-checking initiatives play an essential role in combating misinformation, yet many young adults exhibit resistance to fact-checks that contradict their beliefs. Additionally, repeated exposure to disinformation debates can lead to "misinformation fatigue," where individuals disengage from news entirely, reducing their ability to stay informed.

Example: Studies indicate that individuals who are repeatedly confronted with corrections to their misconceptions may reject fact-checks due to psychological reactance.

6. Legal and Ethical Challenges in Regulating Disinformation

While various legal frameworks exist to combat disinformation, challenges arise in balancing freedom of expression with censorship concerns. Countries have different approaches to tackling fake news, from imposing strict penalties to promoting voluntary self-regulation among tech companies.

Example: (Albarian, 2018, p. 45-49) discusses how some nations impose extraterritorial sanctions on disinformation campaigns, highlighting the complexity of regulating misinformation across borders.

Addressing these challenges requires a multi-faceted approach that includes integrating media literacy into educational curricula, enhancing algorithmic transparency, fostering critical thinking skills, and promoting responsible digital citizenship. By equipping young adults with the tools to critically assess information, society can mitigate the impact of disinformation and strengthen public discourse.

Italy:

The most widespread practices for countering disinformation include fact-checking (Robertson et al. 2020) and debunking (Chan et al. 2017; Lewandowski et al. 2020), whose effectiveness is a key issue in the debate on the topic (Cotter et al. 2022). Indeed, while fact-checking and debunking initiatives have multiplied (Amazeen 2020), recent studies have concluded that these practices have a number of limitations. First, there is the problem of mismatched publics: very often, the public that consumes fake news is not the same as that which has access to fact-checking sites (Guess et al. 2018). Second, certain "blocking" dynamics, mostly associated with processes of selectivity and hyper-selectivity, have effects that are added to and exacerbate those of confirmation bias. Third, a number of studies have pointed out that there is a gap between the amount of fake news that is shared and the number of fact-checking articles (Vargo et al. 2017). Fourth, studies of social network echo chambers have found little interaction in terms of user engagement with posts containing "corrective" content (Quattrociocchi & Vicini 2016): such interactions are not only few in number, but their tone and the statements they make reflect negative feelings that indicate a resistance to accepting checks or denials of content in line with the user's ideological stance and values. In addition to these limitations, fact-checking practices have several strong points: first, fact-checking (mostly of articles or politicians' statements) is more effective if it is considered not as an end in itself, but as the starting point for the kind of research typical of investigative journalism. Second, these practices are also more effective if they are combined with network dynamics and collaboration among multiple actors. Fact-checking and debunking practices should become part of the information and communication flows, and their content should be circulated outside their websites and digital platforms, leveraging the potential of the Hybrid Media System in the fight against disinformation. This means ensuring that the agenda-setting power of debunked content is greater than that of fake news so that it can reach different types of audience. Given disinformation's pervasiveness, moreover, and its impact on many levels and many systems—politics, the media, and the public sphere—it seems clear that the phenomenon calls for a multilevel approach that focuses equally on fact-checking processes, the public actors, and the policy lines to be adopted (Belluati & Fubini, 2022).

Turkey:

The main challenges in educating young adults about recognizing fake news and disinformation include their tendency to overestimate their abilities, believing they can easily spot false information, leading to complacency (Bulger & Davison, 2018). Additionally, young adults often rely heavily on social media for news, platforms notorious for rapidly spreading misinformation, complicating efforts to build accurate awareness (Wineburg & McGrew, 2016). Further, educational curricula frequently lack systematic media literacy programs, making it challenging to equip young adults with critical evaluation skills to effectively discern credible information (Hobbs, 2017).

Portugal:

General:

- Data showed that one of the main challenges in educating young adults is the speed at which technology and young people's online activity change, making it difficult to keep up to date with recent developments (Europe, Council of, n.d.).

Portugal:

- In Portugal, another pressing challenge verified is the lack of interest from the younger generation regarding news in general, which might make them less willing to learn more about how they could protect themselves from this type of news and disinformation (Público, 2022).

Luxembourg: No Answer

6. Could you identify the audiences and target groups affected by the measures?

France:

1. Young Adults and Students: One of the primary target groups in efforts to educate the public about recognizing fake news and disinformation is young adults, especially those in higher education or early professional careers. As frequent consumers of social media, they are often exposed to misinformation and disinformation and may struggle to distinguish between credible and unreliable sources. The challenges of cognitive biases and the sophistication of disinformation, such as deep fakes, further complicate their ability to assess information critically (Friggeri et al., 2014, p. 101-110). Thus, media literacy programs are crucial in equipping this group with the necessary skills to navigate the digital information landscape effectively.

2. General Public (Digital Citizens): The general population is another key group affected by the spread of disinformation, as it can shape public opinions, behaviours, and even political outcomes. Public health campaigns, for example, may target all citizens to raise awareness of misinformation during events like the COVID-19 pandemic. In this context, media literacy initiatives are essential in reducing the impact of misinformation on public health and behaviour (Levine et al., 2009, p. 9-42).

3. Political and Electoral Audiences: Voters and citizens participating in democratic processes, such as elections, are particularly vulnerable to disinformation campaigns. Research on disinformation in the French presidential election, for instance, highlighted the role of social bots and the spread of false information to influence public opinion (Ferrara, 2017, p. 17-20). Therefore, political audiences are a significant target group for initiatives aiming to prevent disinformation from influencing electoral outcomes. Educational programs on recognizing fake

news and understanding the sources of political disinformation are key measures to protect this group.

4. Social Media Platforms and Technology Users: Individuals engaged in social media, including both content creators and consumers, are directly affected by disinformation. Since social media platforms serve as major channels for disinformation dissemination, strategies like network analysis and machine learning tools that detect and flag fake news are essential in addressing this challenge. Social media users, particularly those who lack media literacy skills, need tailored educational efforts to help them identify and mitigate the spread of disinformation (Friggeri et al., 2014, p. 101-110).

5. Vulnerable and Marginalized Groups: Certain demographic groups, such as older adults or people with lower levels of digital literacy, may be particularly susceptible to disinformation. They may have less experience navigating digital platforms and thus are more likely to fall victim to misleading content. Media literacy programs and disinformation detection measures should be designed with special attention to these groups to reduce their exposure to harmful content.

The audiences and target groups affected by measures to combat disinformation include young adults, the general public, voters, social media users, and vulnerable groups. Each of these populations faces different challenges in addressing disinformation, and as such, tailored interventions are essential for the effectiveness of these measures.

Italy: No Answer Only References

Acampa S. (2022). La disinformazione incontra il populismo. Un'analisi del contesto italiano. Rivista di Digital Politics, 3, 507-522. Doi: 10.53227/106457

Amazeen M. A. (2020). Journalistic interventions: The structural factors affecting the global emergence of fact-checking. Journalism, 21(1), 95-111. Doi: 10.1177/1464884917730217

Barbieri M. (2024). A subtle hypocrisy: insights into the Italian politics' inertia on the counter-disinformation policy. Rivista di Digital Politics, 1, 91-110.

Belluati M. & A. Fubini (2022). Reacting to disinformation. The multilevel EU fact-checking approach. De Europa. European and Global Studies Journal, 5 (2), 55-75. doi: 10.13135/2611-853X/7143

Chan M.S, Jones C.R, Hall Jamieson K & D. Albarracín (2017). Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. Psychol Sci. 28(11), 1531-1546. doi: 10.1177/0956797617714579

Cinelli M., Quattrociocchi W., Galeazzi A., Valensise C.M., Brugnoli E., Schmidt A.L, Zola P., Zollo F., & A. Scala (2020). The COVID-19 social media infodemic. Sci Rep. 10(1),16598.

Cotter K., DeCook J. R., & S. Kanthawala (2022). Fact-Checking the Crisis: COVID-19, Infodemics, and the Platformization of Truth. Social Media and Society, 8(1). Doi: 10.1177/20563051211069048

De Blasio E., & D. Selva (2021). Who Is Responsible for Disinformation? European Approaches to Social Platforms' Accountability in the Post-Truth Era. American Behavioral Scientist, 65(6), 825-846. doi: 10.1177/0002764221989784

Di Mascio F., Barbieri M., Natalini A. & D. Selva (2021). Covid-19 and the Information Crisis of Liberal Democracies: Insights from Anti-Disinformation Action in Italy and EU. Partecipazione e conflitto, 14(1), 221-240. doi: 10.1285/i20356609v14i1p221

Fallis D. (2015). What Is Disinformation? Library Trends, 63 (3), 401-426.

Giglietto F., Iannelli L., Rossi L & A. Valeriani (2016). Fakes, News and the Election: A New Taxonomy for the Study of Misleading Information within the Hybrid Media System (30 November. Convegno AssoComPol 2016)

Giglietto F., Iannelli L., Valeriani A., & L. Rossi (2019). Fake news' is the invention of a liar: How false information circulates within the hybrid news system. Current Sociology, 67(4). doi: 10.1177/00113921198375

Guess A., Nagler J., Tucker J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Sci Adv. 5(1). doi: 10.1126/sciadv.aau4586

Lewandowsky S., Cook J., Ecker U. K. H., Albarracín D., Amazeen M. A., Kendeou P., Lombardi D., Newman E. J., Pennycook G., Porter E., Rand D. G., Rapp D. N., Reifler J., Roozenbeek J., Schmid P., Seifert C. M., Sinatra G. M., Swire-Thompson B., van der Linden S., Vraga E. K., Wood T. J. & M. S. Zaragoza (2020). The Debunking Handbook 2020. doi: 10.17910/b7.1182

Ministero degli Affari Esteri e della Cooperazione Internazionale, come individuare e contrastare operazioni coordinate di disinformazione in Italia, 2023.

Pasquetto I. V.. Olivieri A. F., Tacchetti L., Riotta G. & A. Spada (2022). Disinformation as Infrastructure: Making and Maintaining the QAnon Conspiracy on Italian Digital Media. Proc. ACM Hum.-Comput. Interact, 6, CSCW1

Quattrociocci W. & A. Vicini (2016). Misinformation. Guida alla società dell'informazione e della credulità, Milano: FrancoAngeli.

Robertson C. T., Mourão R. R., & E. Thorson (2020). Who Uses Fact-Checking Sites? The Impact of Demographics, Political Antecedents, and Media Use on Fact-Checking Site Awareness, Attitudes, and Behaviour. The International Journal of Press/Politics, 25(2), 217-237. doi: 10.1177/1940161219898055

Tabasco M., Ruffo G., Flammini A., & F. Menczer (2015). Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks. Proceedings of the 24th ACM International Conference on World Wide Web, of WWW '15 Companion, 977–982. doi: 10.1145/2740908.27425

Vargo C. J., Guo L., & M. A. Amazeen (2018). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. New Media & Society, 20(5), 2028–2049. doi: 10.1177/1461444817712086

Wardle C., & H. Derakhshan (2017). Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking. Council of Europe.

Turkey:

Over-Reliance on social media – Platforms like Twitter and TikTok spread misinformation rapidly, making it difficult to filter credible news.

Limited Media Literacy Education – Schools lack systematic media literacy programs, leaving students without proper critical evaluation skills (Gencel Bek, 2019).

Political Polarization – Biased news consumption reinforces existing beliefs, reducing willingness to question misinformation (Yesil & Sözeri, 2018).

Cognitive Biases & Trust Issues – Young adults tend to believe news aligning with their views and distrust traditional media.

Weak Critical Thinking Training – Education focuses on memorization rather than analytical skills, weakening students' ability to verify sources.

Portugal:

General:

• As media literacy programmes help individuals recognise their rights and responsibilities online, everyone could, to some extent, benefit from these programmes and measures. However, some audiences are mostly affected, such as:

- Young people
- Older adults: data has shown that other people have lower rates of digital and media literacy (Asteriou & Konstantina, 2023). Investing in these programmes might be of added value to help tackle some of the challenges faced by this population.
- Marginalised or vulnerable groups

Luxembourg: No Answer

b) Hate Speech and Online Violence

1. What theoretical approaches are used to understand hate speech and online violence?

France:

Social Cognitive Theory: This theory posits that individuals learn behaviours by observing and imitating others. In the context of online hate speech, exposure to aggressive interactions can influence users to adopt similar hostile behaviours. For instance, studies have shown that repeated exposure to hateful content online can shape users' attitudes and behaviours, leading them to replicate violent or hostile discourse.

Systems Theory: This approach looks at the complex system of interactions within online platforms, considering how user behaviour, algorithms, and societal norms contribute to the spread of hate speech. In other words, the phenomenon of online hate is not only the result of individual actions but also the interaction between users and the technical structures of

platforms. A relevant study on this theoretical framework is by (Barth et al., 2023, p. 209-252), who propose a systems theory of hate communication on social media platforms, analyzing the factors that contribute to the amplification of hate speech in these digital environments (Barth et al., 2023, p. 209-252).

Goffman's Frame Analysis: This approach suggests that individuals engage in hate speech by adopting specific frames that justify their actions in particular contexts. Goffman argues that individuals construct social and psychological realities that legitimize their behavior. For example, in an online context, a user may adopt a frame that portrays a target group as dangerous or deviant, thereby justifying verbal attacks or hate speech. A study illustrating this perspective is by (Kilvington, 2021, p. 256-272), who uses Goffman's work to conceptualize the motivations behind online hate.

Stochastic Terrorism Framework: This concept explains how public figures use inflammatory rhetoric to incite random individuals to commit violent acts. Political leaders or influencers may, through provocative statements, indirectly incite violence by triggering unpredictable reactions from their audience. In this context, the research by (Amman et al., 2024, p. 234-245) on incitement to violence and stochastic terrorism provides an important perspective on how incendiary rhetoric can lead to violent actions (Amman et al., 2024, p. 234-245).

An important reference that addresses the issue of youth and violent extremism online is the study by Alava, Frau-Meigs, and (Hassan, 2017, p. 5-6). This research focuses on how social and digital influences can lead young people to adopt hate-driven and extremist discourse online. Alava, Frau-Meigs, and (Hassan, 2017, p. 5-6) explore the role that social media and digital platforms play in fostering radicalization among youth, discussing how these platforms provide spaces for extremist groups to target vulnerable individuals. The study highlights the ways in which social networks facilitate the spread of extremist ideologies and hateful rhetoric, contributing to the normalization of online violence. It also looks at the influence of peer groups and online communities, which can act as echo chambers, reinforcing radical views and making it easier for young people to become radicalized. This research emphasizes the need for media literacy and interventions to counteract the harmful impact of social media on youth.

Italy:

Online Hate is a global social problem (Malecki, Keating, & Safdar, 2020), but in literature there isn't an univocal definition of "online hate". Hate speech (Waldron, 2012) refers to any expression (written or oral, verbal or non-verbal, explicit or implicit) containing insults, offences or any form of discrimination against specific individuals or social groups (Gheno, 2019; Hammoode et al., 2022). Authors such as Galán del Rey (2017) and Megías et al. (2020) define it as a type of symbolic violence that is detrimental to diversity. Other authors (Amores et al., 2021; Paasch-Colberg et al., 2021) define it as narratives that, depending on their intensity, can be constituted as hate crimes. More generally, it is a harmful speech whose purpose is to cause harm to the recipient (Anindita, Sadiyah, & Khoiriyah, 2022; Civila, Romero-Rodríguez, & Aguaded, 2021; Megías et al., 2020; Paris-Albert, 2021; Richardson-Self, 2018). Anyway, online hate acquires specific characteristics (UNESCO, 2015) such as permanence over time, itinerant nature of the message, anonymity and transnationality of the contents.

The expression "hate speech" was introduced in the 1990s, but in recent years it seems to be linked to new aspects of contemporary society. In Italy, for example, during COVID-19 pandemic, new categories of people targeted by online hate have emerged, such as virologists, who provided medical and life-related advice during the pandemic by participating in TV shows and talk shows (Barile & Panarari, 2020; Nazem et al., 2023), healthcare workers (Pasta, 2021) and supermarket workers.

Psychology reflects on this feeling, considering it as an affective disorder defined by a constant need to hurt someone or something. According to Sternberg's "triangular structure theory of hate" (2007), this feeling is characterized by the same dynamics identifiable in love: (denied) intimacy, passion and commitment (Bagnato, 2020). Passion has to do with the definition of hate as anger or fear, so it's possible to talk about a "warm hate" or about someone who, full of rancor, attacks or flees from the one perceived as harmful. On the other hand, "cold hate" intends hate as an excessive belittling of another person/group through contempt and the perception of others as inferior.

Bazzaco et al. (2018) provide indicators that allow to establish a clear differentiation between hate speech and hate crime, taking into account the message's content and form, the sender and the recipient, the intentionality and the channel through which the message is transmitted, as well as the geographical location and the social, economic and political climate in which it is carried out.

In this way, it's possible to pay more attention in order to not confuse hate speech with hate crime, as legal language has been requiring since some years. It's also useful in order to observe data that is not purely lexical, attempting to construct a multilevel evaluation grid that brings together key words (or indicators) and semantics, frequency of isolated terms and contexts (what in linguistics it's called semantic prosody), hate words and textual, rhetorical and pragmatic elements.

From another perspective, the work of Al Serhan and Elareshi (2019), that is based on Allport's (1954) scale on prejudice and discrimination, provides a gradation of hate levels configured in a pyramidal structure that scales hate speech from prejudicial activities (level 1) to genocide (level 5).

In Italy, however, awareness campaigns lack of national coordination and also legislation's enforcement is often lacking (Jubany & Roiha, 2016; Qozmar et al., 2023).

Turkey:

In Türkiye, several theoretical approaches are used to understand hate speech and online violence, focusing on social, political, and psychological aspects:

- Social Identity Theory (Tajfel & Turner, 1979): Explains how individuals categorize themselves into in-groups and out-groups, leading to discriminatory online behaviors and the spread of hate speech (Tajfel & Turner, 1986).

- Framing Theory (Entman, 1993): Examines how media representations shape public perception of marginalized groups, influencing the spread of hate narratives (Köseoğlu, 2021).

- Networked Publics Theory (boyd, 2010): Describes how digital platforms amplify hate speech by providing a space for anonymous or coordinated attacks (Turan & Çalışkan, 2020).

• Cyberbullying and Anonymity Theory (Suler, 2004): Highlights the "online disinhibition effect," where anonymity emboldens users to engage in hate speech and harassment (Yılmaz & Kaptan, 2021).

Portugal: No Answer

Luxembourg:

Germany

Schmuck, D., Fawzi, N., Reinemann, C., & Riesmeyer, C. P. 313–331 (2021). https://doi.org/10.1080/17482798.2021.1977358

Germany

Obermaier, M. P. 4785-4807 (2024). https://doi.org/10.1177/14614448221125417 and https://doi.org/10.1093/jcmc/zmac012

2. What types of online violent discourse primarily affect your country?

France:

In France, online violent discourse primarily manifests as verbal violence and hate speech on digital platforms. A study by (Gauducheau et al., 2023, p. 79-122) examined how young internet users perceive verbal violence in digital communication, emphasizing the importance of local norms in evaluating such content. The research found that messages considered violent from a normative perspective are not necessarily perceived as such by young people. Additionally, the study indicates that the perception of online verbal violence varies according to gender and familiarity with the forum. This highlights the subjective nature of perceiving verbal violence, influenced by contextual and personal factors, making it challenging to develop universal solutions to address this issue.

In a broader context, it is also important to mention that cyberviolence, cyberbullying, and cyberhate are increasingly prevalent forms of online violence in France (Blaya, 2020, p. 38-43) describes these phenomena as destructive digital behaviours that primarily affect young people and vulnerable groups. The consequences of such violence are numerous, ranging from severe psychological effects such as depression and anxiety to profound social repercussions, including isolation and stigmatization. This digital violence is further complicated by its often taking the form of hateful speech, targeted harassment, and verbal assaults, amplified by the anonymity provided by online platforms.

Italy:

From Vox report: Six large macro-areas (women, migrants, Muslims, homosexuals, Jews, people with disabilities) concern the groups quantitatively most targeted by hate speech (and hate crime) and allow to compare the data collected in 2019 with those obtained from the surveys of the previous 3 years.

From the latest report by Amnesty International, topics:

-Hate Speech on FB – immigration, rom, right to protest, women's rights, Lgbtqia+.

-Hate speech on Twitter – solidarity or humanitarian activities, rom, immigration, women's rights.

Turkey:

Türkiye experiences several forms of online violent discourse, including:

1. Political Hate Speech: social media is often used for politically motivated hate speech, particularly against opposition groups, journalists, and activists (Yesil & Sözeri, 2018).

2. Ethnic and Religious Hate Speech: Kurdish, Armenian, and other minority communities frequently face targeted online abuse.

3. Gender-Based Online Violence: Women, especially activists and journalists, are subjected to cyber harassment and sexist hate speech.

4. Migrant-Related Hate Speech: Syrian refugees and other migrant communities are frequent targets of xenophobic online narratives.

Portugal:

• Online violent discourse in Portugal is an active problem in need of continuous attention. A recent report by Casa do Brasil Lisboa (2025) showed that hate speech against immigrant people has become a pressing concern in Portugal, having escalated in frequency as well as having become more violent in recent years (Casa do Brasil em Portugal, 2025). This report also showed that the topics of hate speech most often used are xenophobia, misogyny and racial Hate (Casa do Brasil em Portugal, 2025).

• This type of violent discourse online is mostly conducted through social media (e.g. Facebook and Instagram). When conducted offline, is mostly conducted on public services (social security, finance and medical centres) (Casa do Brasil em Portugal, 2025).

• Regarding whether anti-immigration political discourses have turned into practices of violence and hate crime against migrants, 66,4% of the respondents indicated a predominant perception that anti-immigration political discourses generate violence and hate crimes (Casa do Brasil em Portugal, 2025).

Luxembourg:

German Federal State of Hesse

https://digitale-schule.hessen.de/gefahren-und-herausforderungen-in-der-digitalen-welt

3. How can the impact of hate speech on communities and individuals be assessed?

France:

Assessing the impact of hate speech on communities and individuals is a complex process that requires an interdisciplinary, multifaceted approach incorporating both quantitative and qualitative research methods. It is essential to adopt a comprehensive framework for understanding how hate speech affects different social groups, individuals, and societal

structures at large. A recent study by (Jihanet al., 2023, p. 18-19) systematically examined the associations between hate speech in both online and traditional media and its effects on individuals, audiences, and communities. This study highlighted several significant findings regarding the harmful consequences of hate speech, such as increased psychological distress, heightened feelings of vulnerability, and a greater propensity to engage in discriminatory or violent behaviours. Additionally, communities that are directly targeted by hate speech often experience a decline in social cohesion, erosion of trust, and an overall weakening of societal solidarity.

A key component in evaluating the effects of hate speech is understanding the psychological toll it takes on individuals and communities. Studies indicate that individuals exposed to hate speech often report experiencing heightened anxiety, depression, and a sense of alienation. For example, exposure to hateful language targeting one's race, ethnicity, or religion can lead to internalized feelings of inferiority or fear, further exacerbating the psychological burden. One prominent study by (Gagliardone et al., 2015, p. 5-6) found that repeated exposure to hate speech contributed to a decline in self-esteem and an increase in perceived discrimination among minority groups, leading to long-term mental health consequences. Furthermore, hate speech is associated with increased feelings of powerlessness, making it more difficult for affected individuals to participate fully in social, economic, and political spheres.

From a community perspective, the effects of hate speech can be equally devastating. When communities are subjected to consistent exposure to hate speech, there is a marked decline in social trust and cohesion. In multicultural societies, for instance, the targeting of one ethnic or religious group through hate speech can lead to heightened polarization, creating divides between groups and fostering an "us versus them" mentality. This division is often exacerbated by social media platforms, where hate speech spreads rapidly and reinforces extremist viewpoints. The decline in social trust in such contexts can result in decreased collaboration, mutual understanding, and the deterioration of community bonds. A notable example of this can be seen in the aftermath of the 2015 refugee crisis in Europe, where increased hate speech directed at refugees led to an increase in social tensions and division between local communities and migrant populations, contributing to the rise of far-right political movements and xenophobic sentiments.

To effectively assess the impact of hate speech, researchers employ various research methodologies. Quantitative approaches, such as surveys, are frequently used to gauge the psychological effects of hate speech, with questions designed to assess emotional distress, levels of anxiety, and experiences of discrimination. In contrast, qualitative methods, such as content analysis, enable researchers to examine the prevalence and nature of hate speech in media, identifying common themes and targeting patterns. For example, content analysis of online platforms such as social media and news outlets can reveal how frequently hate speech appears, its tone, and the demographic groups that are most often targeted. This analysis helps contextualize how hate speech is disseminated and received, shedding light on its impact on public discourse and individual behaviour.

Moreover, statistical methods, including regression analysis, are employed to identify correlations between exposure to hate speech and negative outcomes. For example, studies have shown that communities that experience high levels of hate speech are more likely to report increased rates of hate crimes, social unrest, and reduced trust in governmental and social

institutions. These findings highlight the tangible consequences of hate speech, illustrating its role as a catalyst for real-world harm and social conflict.

In light of these findings, it is essential to develop targeted interventions to mitigate the harmful effects of hate speech. Researchers and policymakers alike emphasize the importance of promoting digital literacy, fostering media regulation to limit the spread of harmful content, and creating safe spaces for marginalized groups. Efforts to promote social harmony and tolerance, such as educational campaigns that emphasize the value of diversity and inclusion, can play a vital role in countering the divisive effects of hate speech. At the individual level, providing support services such as counselling and community-building programs can help those affected by hate speech to recover and rebuild trust in their social environments.

The assessment of hate speech's impact on individuals and communities requires a nuanced, interdisciplinary approach that combines psychological, social, and media studies. By employing a variety of research methodologies, including surveys, content analysis, and statistical analyses, researchers can capture the multifaceted nature of hate speech and its far-reaching effects. Understanding these dynamics is crucial for the development of strategies aimed at minimizing the harmful consequences of hate speech and promoting social cohesion in an increasingly polarized world.

Italy:

(Source: https://www.retecontrolodio.org/news/)

Impact on community:

Hate speech contributes to a narrative that relegates to inferiority those who are traced back (by the hater) to certain social categories (Girard C., 2015). In this way, hate speech creates and fuels conditions that undermine rights and freedom of stigmatized people, their ability to live without obstacles and to defend their civil interests. Hate speech attacks all those rights that guarantee an equal participation in civil life and an open public debate in which everyone's interests are taken into account and everyone has a voice that will be heard (equality in dignity and rights, the right to take part in the direction of the country's public affairs, participation in the cultural life of the community, etc.). Hate speech, in fact, can have a negative impact on the quality of social life.

Impact on individuals:

Hate speech is based on discrimination's elements. For this reason, it strikes straight to attacked people's dignity and undermines their sense of security, as individuals already considered like socially vulnerable and potentially victims of multiple discrimination. When the attack perpetrated through hate speech is personal, it often takes the form of dehumanizing or degrading treatment. Hate speeches become labels assigned to people's identity and this mark have some consequences. The attribution of inappropriate identities and the gap between these and the perception that each person has of themselves damage the psychological well-being and social adaptation of the individual (Huber & Reynolds, 2014).

Turkey:

Psychological Assessments: Studies use surveys and interviews to measure mental health effects such as anxiety, depression, and self-censorship among victims.

• Social Cohesion Indexes: Hate speech weakens social trust and increases polarization. Researchers analyze voting patterns and public attitudes toward minority groups to assess long-term effects.

• Big Data and Sentiment Analysis: Tracking hate speech trends using AI and machine learning tools helps measure how online discourse affects communities over time.

• Legal Case Studies: Analyzing legal cases related to hate speech prosecution in Türkiye offers insights into enforcement effectiveness and societal impact (Yesil & Sözeri, 2018).

Portugal: No Answer

Luxembourg: No Answer

4. What methods are used to monitor and analyze hate speech trends on digital platforms, and by whom? Feel free to provide names and links to relevant sites

France:

Monitoring and analyzing hate speech trends on digital platforms involves a combination of automated tools and human oversight. Organizations such as ProPublica have developed projects like Documenting Hate, which utilizes machine learning and natural language processing to track hate crimes and bias incidents reported in the media.

(wikipedia.org)

Similarly, the Center for Countering Digital Hate (CCDH) employs data analysis to identify and counteract online hate speech. Their reports often focus on the role of social media platforms in disseminating harmful content.

(wired.com)

Additionally, academic researchers have conducted systematic reviews of hate speech detection methods. For instance, a study by (Jahanet al., 2021, p. 24) provides an overview of automatic detection techniques using natural language processing.

Italy:

Task Force Hate Speech – Amnesty International – from 2016 protects group at risk of discrimination for reason related to ethnicity, religion, gender, gender identity, sexual orientation, disability, migration etc. Task Force intervenes on Facebook pages and other main sources of news, as sometimes online newspapers, in order to increase their visibility, provide headlines and articles that are useful for generating hateful comments. Pages are monitored individually, choosing each time which newspapers it's necessary to check and which posts need to focus on. When they detect a post on specific topics in which hateful comments are already present or could appear, TFHS members communicate it to the group and take action:

1. reporting comments that violate Facebook policies, mainly those relating to hate speech;

2. replying to comments.

3. writing comments on the post, to clarify some points and provide information that the title does not provide (and sometimes not even the article), introducing an alternative narrative;

4. supporting TFHS members and other people who express balanced positions, liking, integrating and reiterating.

Turkey:

Automated Sentiment Analysis & AI Tools: Platforms use machine learning algorithms to detect and flag hate speech patterns. Turkish researchers use tools like Hatebase and TextBlob for linguistic analysis.

• Fact-Checking and Monitoring Organizations:

• Teyit.org (https://teyit.org) – A leading fact-checking organization that counters hate speech and misinformation.

• Media Research Association (MEDAR) (https://medar.org.tr) – Conducts digital media monitoring, analyzing polarization and online violence.

• Social Network Analysis: Researchers track hate speech trends by mapping interactions on platforms like Twitter and Facebook using network analysis tools (Bulut, 2021).

• Human Rights and NGO Reports: Organizations like Amnesty International Türkiye (https://www.amnesty.org.tr) publish reports on digital rights violations, including hate speech.

Portugal:

General:

• Here are some methods and tools to monitor and analyse hate speech used by social media platforms across different countries:

- Hootsuite: https://www.hootsuite.com/hootsuite-vs-brandwatch
- Talkwalker: https://www.talkwalker.com/
- Brandwatch: https://www.brandwatch.com/
- Natural Language processing (NLP): https://www.ibm.com/think/topics/natural-language-processing
- Hatebase: https://hatebase.org/
- Human content moderation, usually conducted by people who manually screen every user-generated content posted on digital platforms.

Portugal:

- KNOwHATE: https://huggingface.co/spaces/knowhate/portuguese-hate-speech-detection

Luxembourg:

Germany

https://love-storm.de/

https://love-storm.de/10-tipps-gegen-hass-im-netz/

5. How effective are current strategies for preventing and responding to online violence and hate speech?

France:

Current strategies for preventing and responding to online violence and hate speech have shown limited effectiveness. A systematic review by (Windisch et al, 2022, p.18) analyzed two studies evaluating online interventions aimed at reducing online hate speech and cyberhate. The findings indicated a small, statistically insignificant reduction in the creation of hateful content online, suggesting that the evidence is insufficient to determine the effectiveness of these interventions.

The review identified several gaps in the evaluation literature, including a lack of experimental and quasi-experimental studies, a focus on the accuracy of detection and classification software rather than the actual reduction of hateful content, and the need to include both extremist and non-extremist individuals in future studies. The authors concluded that more rigorous research is necessary to assess the effectiveness of online interventions in combating hate speech and cyberhate.

Italy: **No Answer**

Turkey:

Current strategies to combat online hate speech in Türkiye include legal regulations, content moderation policies, and civil society initiatives. However, their effectiveness is mixed due to limited enforcement, political influences, and the evolving nature of digital hate speech (Yesil & Sözeri, 2018).

- Legal Framework (Limited Effectiveness): Türkiye has laws against hate speech under the Turkish Penal Code (TCK 216) and Law No. 5651 on Internet Crimes, but enforcement is inconsistent. Critics argue these laws are sometimes used to suppress dissent rather than curb hate speech.

- Platform Content Moderation (Partially Effective): Social media companies like Meta and Twitter comply with some Turkish regulations but lack transparency in content removal decisions. Automated detection systems remove obvious hate speech, but nuanced cases (e.g., sarcasm) often bypass filters.

- Fact-Checking & Awareness Campaigns (Moderately Effective): Organizations like Teyit.org and Media Research Association (MEDAR) provide fact-checking services and promote digital literacy, but their reach remains limited in Türkiye's highly polarized media environment.

Portugal:

- Legal frameworks: In Portugal, discrimination and incitement to hate and violence are punished according to the Portuguese Penal Code, article no.240. The Law no. 93/2017, which establishes the legal framework for prevention, prohibition and combating of discrimination on the grounds of racial and ethnic origin, colour, nationality, ancestry and place of origin, makes

it illegal to incite hatred or violence through public speech, whether online or offline. However, research has shown that the mechanisms in place to help tackle these pressing concerns, either formal (public institutions) or informal (social media platforms), have been deemed inefficient, or at least not sufficient to respond to the rise of hate crimes reported in Portugal, Spain and in Europe (BPI, 2023).

Luxembourg: No Answer

6. What methods, tools, or technologies are used to detect and mitigate hate speech?

France:

Detecting and mitigating hate speech involves a combination of technological tools and methodologies. Machine learning (ML) and deep learning (DL) techniques are extensively employed for automated detection. These methods analyze textual data to identify patterns indicative of hate speech. For instance, a systematic review by (Srba et al, 2021, p.15) categorizes existing studies into five method categories: Machine Learning, Deep Learning, Ensemble models, Graph Neural Networks, and Graph Convolutional Networks. These advancements can aid social networking services in identifying hate messages before being posted, reducing the risk of harassment.

Natural Language Processing (NLP) techniques, such as sentiment analysis and semantic analysis, are also utilized to comprehend the context and intent behind messages, enhancing the accuracy of hate speech detection. Additionally, counterspeech strategies responding to hate speech with alternative narratives or empathetic responses are employed to mitigate the impact of harmful content.

It's important to note that while these technologies are effective, they are not foolproof. Challenges such as biases in detection algorithms and the evolving nature of hate speech require continuous refinement of these methods. Ongoing research aims to address these issues and improve the efficacy of hate speech detection and mitigation strategies.

Italy:

A possible way to mitigate and contrast online hate is to respond to the dominant negative narrative by proposing an alternative vision. To eradicate hate speech, in fact, it isn't enough to sanction it or prove it false, but it's also useful to offer an opposing narrative to adhere to.

Turkey:

In Türkiye, hate speech detection relies on AI-powered tools, linguistic analysis, and manual reporting systems:

- Artificial Intelligence & Machine Learning:
    - Hatebase: A global hate speech database used to train AI for detecting offensive content in Turkish and other languages.
    - Google Perspective API: Assesses the toxicity of online comments.
    - TextBlob & FastText: Used for sentiment analysis and natural language processing in Turkish hate speech detection (Bulut, 2021).
- Social Media Monitoring:

- Teyit.org (https://teyit.org) uses manual verification and AI-supported tools to counter disinformation and hate speech.
- MEDAR (https://medar.org.tr) conducts research on algorithmic biases in content moderation.

• Human Moderation & Community Reporting:

- Amnesty International Türkiye and Hate Speech Watch (https://nefretsoylemi.org) collect reports from individuals and analyse trends.

Portugal:

General:

• One powerful method to detect, mitigate and prevent hate speech is online/ in-person campaigns and courses to teach people to recognise hate, embrace diversity and stand up to bias (Windisch, Wiedlitzka, & Olaghere, 2021).

• Content moderation also plays a crucial role in detecting and mitigating hate speech. Even though AI content moderation might be extremely valuable, research has shown that it is equally important to also rely on human content moderation, as AI models sometimes fail to detect contextual meaning and account for localised nuances of hate speech (Hatano, 2023).

Portugal:

• In Portugal, the tool of the project kNOwHATE, a groundbreaking and innovative tool on tackling hate speech online, has shown promising potential in detection cases of online hate speech, either direct or indirect (e-Global, 2024).

Luxembourg: No Answer

7. What forms of cyberbullying are prevalent in your country, and how are they addressed by legal frameworks and authorities?

France:

In France, cyberbullying manifests in various forms, including:

Harassment via social media: Individuals are targeted with offensive comments, threats, and defamatory content on platforms like Facebook, Twitter, and Instagram.

Cyberstalking: Persistent online surveillance and unwanted contact, often leading to emotional distress

Doxxing: The malicious publication of private information, such as addresses or phone numbers, to intimidate or harm individuals.

Exclusion and Ostracism: Deliberate social exclusion or spreading rumors to damage an individual's reputation.

To combat these issues, France has implemented several legal measures:

Avia Law: Enacted in 2020, this law aimed to remove hateful content from the internet within 24 hours. However, the Constitutional Council struck down key provisions, citing concerns over freedom of expression. The remaining measures include the creation of a specialized prosecutor's office for online hate messages and an Online Hate Observatory

Criminal Code Provisions: French law penalizes online harassment, including cyberbullying, under various articles of the Criminal Code. Perpetrators can face fines and imprisonment, depending on the severity of the offense.

Reporting Platforms: Websites like Internet-signalement.gouv.fr allow individuals to report illegal online content, including cyberbullying incidents (Blaya, C, 2018, p. 169-187)

Despite these efforts, challenges remain in effectively addressing cyberbullying, particularly concerning the balance between regulation and freedom of expression.

Italy:

Cyberbullying is a particular form of bullying that takes place through digital devices such as cell phones, computers and tablets. Cyberbullying can occur via SMS text, apps, online social media, forums or games where people can view, participate or share content. Cyberbullying includes sending, posting or sharing negative, harmful, false or mean content about someone else. It can include sharing personal or private information about someone else, causing embarrassment and humiliation. Some instances of cyberbullying cross the line into illegal or criminal behavior. Cyberbullying occurs especially on: social media, such as Facebook, Instagram, Snapchat and Tik Tok; Text messaging and messaging apps on mobile or tablet devices; Instant messaging, direct messaging and online chat on the Internet; online forums; chat rooms and message boards, such as Reddit, e-mails or online gaming communities.

Studies have shown that different types of cyberbullying can be differentiated based on aspects like the covert or overt nature of the acts, the electronic devices/media used to bully others, or specific behaviours. Starting from Willard's listing of cyberbullying behaviours, Nocentini et al. used a classification based on the nature of the attack:

- Written-verbal includes acts that use written or verbal forms of bullying (i.e., phone calls, text messages, and e-mails).

- Visual includes attacks that are perpetrated through the use of visual forms of bullying (i.e., posting compromising images).

- Impersonation refers to more sophisticated attacks that use identity theft (i.e., revealing personal information through another person's account).

- Exclusion is related to the choice of who is an in-group member and who is an outcast (i.e., intentionally excluding someone from an online group).

In Italy, the cyberbullying's phenomenon is defined by the law "Provisions for the protection of minors for the prevention and contrast of the phenomenon of cyberbullying", which came into force on 18 June 2017. This law introduces educational and training measures, with the main purpose to promote a deeper awareness among young people about the disvalue of

persecutory behaviours which, generating isolation and marginalization, can lead to very serious consequences for victims in particularly fragile situations.

Specifically, this law:

- Identifies the purpose of the intervention in the fight against all cyberbullying's manifestations through a strategy that includes preventive and educational measures towards minors (victims and perpetrators of bullying on the web) that need to be implemented in school environment.

- Provides that a minor who has turned 14 years old and who is a victim of cyberbullying (as well as each parent or person exercising legal responsibility over the minor) can submit a request to the manager of the involved Internet site or social media in order to obtain inhibitory and prescriptive measures for his/her protection (blacking out, removal, blocking of any other minor's personal data disseminated on the Internet, with retention of the original data).

- Establishes a technical table for the prevention and fight against cyberbullying. After consulting the Ministry of Justice, it also provides the adoption by the MIUR of specific guidelines - to be updated every two years - for the prevention and fight against cyberbullying in schools.

- Provides the designation, in each school, of a teacher with the role of contact person for any initiative against cyberbullying. This referent has to collaborate with the police forces, associations and youth centres present in the school's area.

- If a cyberbullying incident occurs in a school, it also provides the obligation for the principal to promptly inform parents (or legal guardians) of the minor involved and to activate appropriate educational actions.

- Applies the discipline on the warning of the police commissioner, borrowed from the one about stalking, also to cyberbullying.

Projects in Italy:

In the last years, the MIUR has dealt with the phenomenon of Cyberbullying also following many dramatic episodes that occurred in Italy.

The Generazioni Connesse – Safe Internet Center Italy Project is a co-financed by the European Commission under the "Connecting Europe Facility" (CEF) programme. In this project the Commission promotes strategies useful to make the Internet a safer place for younger users, promoting its positive and conscious use. The project is coordinated by the MIUR, in partnership with the Ministry of the Interior-Postal and Communications Police, the Guarantee's Authority of Childhood and Adolescence, Save the Children Italy, Telefono Azzurro, University of Florence and University of Rome "La Sapienza". The aim of this partnership is providing continuity to the experience developed over the years, improving and strengthening the role of the Italian Safer Internet Centre as a point of reference at a national level for issues relating to Internet safety and the relationship between young people.

The ELISA Project (E-learning Training for Teachers on Anti-Bullying Strategies) was born thanks to the collaboration between the MIUR-General Directorate for Students and the Department of Educational Sciences and Psychology of the University of Florence. The ELISA Platform aims to provide useful tools to schools and teachers in order to effectively intervene

on the issue of bullying and cyberbullying. The basic training consists of 4 progressive courses, for a total of 25 training hours, and an advanced course:

1) "Bullying and cyberbullying: knowledge, assessment and indications for prevention" (8 hours).

2) "Universal prevention paths" (7 hours);

3) "Cyberbullying: legal aspects" (3 hours);

4) "Indicated Prevention Paths" (7 hours).

5) School and teachers.

Each course is divided into several modules organized as lessons. First of all, there is an introductory video lesson. After, a series of operational tools are offered, these can be downloaded and used at school. There are also suggestions for further study, practical exercises and final questionnaires for self-assessment. At the end of the courses, the teacher receive a certificate of participation containing the hours of training carried out on the ELISA Platform.

Another initiative conducted by researchers from the University of Florence is NoTrap! (www.notrap.it). NoTrap! (Let's not fall into the trap!) is a bullying and cyberbullying prevention programme for middle and high school students. It was conceived and tested by the Laboratory of Longitudinal Studies in Developmental Psychology of the University of Florence, coordinated by Professor Ersilia Menesini. This intervention model supposes an active involvement of the school and especially of the students who, based on a peer education/peer support model, become the main characters of change. The students, after an educational training, intervene both in class and online to give their help and support to all the people who feel in difficulty.

Turkey:

In Türkiye, cyberbullying takes multiple forms, with youth, journalists, and marginalized groups being primary targets.

1. Harassment and Online Threats: Frequent on platforms like Twitter and Instagram, where individuals face abusive messages, death threats, and doxxing

2. Trolling and Political Cyberbullying: Coordinated attacks (e.g., AKTrolls) target activists, opposition voices, and journalists with defamation campaigns (Yesil & Sözeri, 2018).

3. Sexual Harassment and Image-Based Abuse: Women experience cyberstalking, revenge porn, and deepfake abuse.

4. Hate Speech Against Minorities: Kurdish, Armenian and refugee communities face organized online harassment.

Portugal:

• As in other countries, cyberbullying has become a pressing concern in Portugal, especially among students. A recent report by Agência Lusa (2024) showed that the most prevalent forms of cyberbullying Portugal are rumours and harassment (Agência Lusa, 2024). Another study conducted with Portuguese students from public middle schools in Lisbon

showed that victims of cyberbullying also received threats and rude messages through the Internet (Coelho, Sousa, Marchante, Brás, & Romão, 2016).

- Even though cyberbullying isn't considered a crime per se in Portugal, the conduct can be penalised under other types of criminal offences.

- In addition, we also highlight the Law no. 109/2009 of 15 September (The Cybercrime Law) that addresses various types of online crimes, like online harassment and cyberbullying.

- Recognising the ongoing challenges, the Portuguese Government recently announced the creation of an anonymous helpline to report cases and provide support to students who are victims of bullying and cyberbullying in public and private schools in Portugal (Governo da República Portuguesa, 2025). This helpline will be available through an app, website or telephone and aims to be a response to the high number of elevated cases that go unreported because of fear of backlash, distrust of the effectiveness of the institution's responses and lack of awareness of proper channels to seek help (Governo da República Portuguesa, 2025).

Luxembourg:

Germany

https://www.schau-hin.info/cybermobbing

Die Initiative „SCHAU HIN! Was Dein Kind mit Medien macht." hilft Familien bei der Medienerziehung. Der Medienratgeber für Familien informiert Eltern und Erziehende über aktuelle Entwicklungen der Medienwelt und Wissenswertes zu den verschiedensten Medienthemen, zum Beispiel Smartphone & Tablet, Soziale Netzwerke, Games, Apps, Medienzeiten und Streaming. SCHAU HIN! gibt Eltern und Erziehenden Orientierung in der digitalen Medienwelt und konkrete, alltagstaugliche Tipps, wie sie den Medienkonsum ihrer Kinder kompetent begleiten können. „SCHAU HIN! Was Dein Kind mit Medien macht." ist eine gemeinsame Initiative starker Partner: Das Bundesministerium für Familie, Senioren, Frauen und Jugend, ARD und ZDF sowie die AOK – Die Gesundheitskasse unterstützen SCHAU HIN! um Eltern und Erziehenden mit zahlreichen alltagstauglichen Tipps und Empfehlungen zu unterstützen, ihre Kinder im Umgang mit Medien zu stärken.

The initiative "SCHAU HIN! What your child does with media." helps families with media education. The media guide for families informs parents and guardians about current developments in the media world and interesting facts on a wide range of media topics, such as smartphones and tablets, social networks, games, apps, media times and streaming. SCHAU HIN! gives parents and guardians guidance in the digital media world and concrete, everyday tips on how they can competently accompany their children's media consumption. "SCHAU HIN! What your child does with media." is a joint initiative of strong partners: The Federal Ministry for Family Affairs, Senior Citizens, Women and Youth, ARD and ZDF as well as AOK - Die Gesundheitskasse support SCHAU HIN! in order to support parents and guardians with numerous tips and recommendations suitable for everyday use to strengthen their children in dealing with media.

Translated with DeepL.com (free version)

8. What methods and tools are used to monitor and analyze trends in hate speech and online violence, and how effective are current strategies for prevention and response?

France:

Monitoring and analysing trends in hate speech and online violence involve a combination of automated tools and human interventions. Automated methods, such as natural language processing (NLP) algorithms, are employed to detect and classify harmful content at scale. For instance, (Garland et al.,2020, p. 22) developed an ensemble learning algorithm that achieved high accuracy in identifying hate and counter speech in a large dataset of tweets.

In addition to automated detection, human-driven initiatives play a crucial role. The #Iamhere movement exemplifies collective moderation, where volunteers actively engage in online discussions to counteract hate speech and promote civil discourse. Studies have shown that such collective efforts can enhance the quality of online conversations by reducing toxicity and extremism.

Regarding the effectiveness of current strategies, research indicates that while automated tools can efficiently identify harmful content, they may not fully capture the nuanced context of language. Therefore, combining automated detection with human moderation and counter-speech initiatives appears to be a more effective approach in combating online hate and violence.

Italy:

With reference to the online hate speech's phenomenon, it's necessary to recall the Additional Protocol to the "Budapest Convention on Cybercrime", which commits States to consider as crimes (when carried out through IT tools): the dissemination of racist and xenophobic material, at least in cases where the material promotes or incites violence (art. 3); racist and xenophobic threats and insults (arts. 4 and 5); the denial, serious minimization, approval or justification of genocide or crimes against humanity (art. 6). Italy signed in 2011, but the Protocol isn't yet ratified.
(Source:https://www.interno.gov.it/sites/default/files/202103/brochure_quando_lodio_diventa_reato.pdf)

Turkey: No Answer

Portugal:

General:

- Research has showed that deep learning models, such as CNNs (Convolution neural networks), RNNs (Recurrent neural networks) and Transformer-based architectures such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer), have demonstrated remarkable abilities in detecting hate speech online and analyse online content (Subramanian, Sathiskumar, Deepalakshmi, Cho, & Manikandan, 2023).

Portugal:

- Tool from the Portuguese Project: kNOwHATE. This tool showed significate positive results in identifying hate speech online on social media platforms. However, the use of

metaphors, irony and humour challenged the identification of hate speech, which might contribute to the low number of hate speech criminal complaints (Gonçalves, 2024).

• A study conducted in the Portuguese setting, verified that BERT-based models are more effective for Hate Speech detection when compared with generative models and other DL models (Ramos, et al., 2024).

Luxembourg: No Answer

c) Fake News

1. What are the key theories and methods used to study the spread and impact of fake news or info, including trolls and pro-denial false or misleading information?

France:

The study of the spread and impact of fake news, including trolls and pro-denial false or misleading information, employs several key theories and methods drawn from multiple disciplines. These approaches aim to analyze how misinformation propagates, the factors influencing its spread, and the psychological and social consequences of its dissemination. Below are some of the key theories and methods used to study fake news:

- Inoculation Theory:

Inoculation theory, rooted in psychology, suggests that exposing individuals to weakened forms of misinformation can help build resistance against future exposure to false information. The process is like vaccination: by understanding and recognizing common misinformation techniques, individuals become better equipped to critically evaluate and reject future misinformation. This theory has been emphasized as an effective strategy to combat misinformation, enabling individuals to become more discerning consumers of information. Research by (Lewandowsky et al., 2017, p. 353-369) demonstrates that inoculation can be an effective way to reduce the impact of misinformation.

- Social Network Analysis:

Social network analysis is a method that examines how information spreads across social networks by mapping out relationships and interactions between individuals or groups. By understanding the flow of information, researchers can identify influential nodes, such as trolls or automated bots, that significantly contribute to the dissemination of misinformation. A study by Friggeri, (Adamic, and Eckles, 2014, p. 9-17) explored how rumors spread on social media, revealing the role of network structure in the viral propagation of misinformation.

- Cognitive Biases and Heuristics:

Cognitive biases are crucial in understanding why individuals are susceptible to believing and spreading fake news. Heuristics, or mental shortcuts, such as the illusory truth effect, can lead individuals to believe misinformation after repeated exposure. The illusory truth effect suggests that the more frequently a piece of false information is encountered, the more likely it is to be accepted as true. (Pennycook and Rand, 2018, p. 353-369) demonstrated this effect in the context of online misinformation, highlighting how repeated exposure to fake news influences people's judgment and decision-making processes.

- Computational Modelling:

Computational modelling is used to simulate the spread of misinformation across networks and assess the dynamics of its propagation. These models help predict how misinformation might spread and identify interventions that could slow or stop its dissemination. For instance, (Vosoughi, Roy, and Aral, 2018, p.7-9) used computational methods to examine the spread of fake news on Twitter, finding that false information spreads more rapidly and widely than true information.

- Experimental Studies:

Experimental studies, often conducted through surveys or controlled online platforms, allow researchers to assess how individuals interact with misinformation and the impact of various interventions. These studies can explore how fact-checking, media literacy programs, or other interventions might mitigate the effects of fake news. For example, (Ruffo and Semeraro,2021, p. 4) conducted an experimental study that examined how cognitive biases affect individuals' ability to distinguish between true and false news, demonstrating that training individuals in bias detection can improve their ability to discern misinformation.

- The Role of Media Literacy:

Media literacy is a critical strategy for combating fake news. By teaching individuals how to critically assess the sources and credibility of the information they encounter, media literacy programs can reduce the spread of misinformation. (Amiel, 2020, p. 5-9) argued that improving media literacy and encouraging scepticism toward unreliable sources is key to limiting the impact of fake news. Media literacy programs help people recognize the signs of misinformation, such as sensationalist language, lack of verifiable sources, and emotional manipulation, making them more resilient to fake news.

- The Role of Journalism and Fact-checking:

Journalism plays an essential role in combating misinformation by adhering to rigorous editorial standards and fact-checking procedures. Journalistic integrity ensures that only accurate, verifiable information is disseminated to the public. In this context, (Amiel, 2020, p. 5-9) also pointed out the responsibility of the press to fight fake news and its "virus," stressing the importance of accurate reporting and the accountability of social media platforms for content moderation.

The study of fake news and misinformation relies on a combination of interdisciplinary theories and methods, such as inoculation theory, social network analysis, cognitive biases, computational modelling, and experimental studies. Understanding how misinformation spreads, the factors influencing its dissemination, and how individuals can be trained to recognize and resist it is crucial for addressing its societal impact. Media literacy offers a promising strategy to reduce the effectiveness of misinformation campaigns.

Italy:

The term “fake news” is not new, it has a long legacy reaching back centuries, especially in times of war and during interwar periods to promote interests (Schudson Zelizer, 2017, p. 2). In the digital age, the problem has become far more acute, as news became distributed online instead of offline and social media rose as channels of news distribution, multiplying the

influence and impact of fake news. They can be defined as news "either wholly false or containing deliberately misleading elements incorporated within their content or context" (Bakir and McStay, 2018: 154, 155). Fake news can be deconstructed into seven categories: false connection (articles in which headlines, visuals or captions do not support the content); false context (genuine content shared with false contextual information); manipulated content (genuine imagery/information manipulated to deceive); misleading content (misleading use of information to frame an issue or individual); imposter content (genuine sources are impersonated); fabricated content (100 % false, designed to deceive and harm); and satire/parody (with potential to fool but no intention to cause harm). The

definitions of "fake news" can include "news articles that are intentionally and verifiably false, and could mislead readers" (Allcott Gentzkow, 2017: 213), and which appropriate "the look and feel of real news; from how websites look; to how articles are written; to how photos include attributions" (Tandoc, Lim, Ling, 2018: 11), hiding "under a veneer of legitimacy" as they take on "some form of credibility by trying to appear like real news" (ibidem), and extend to the entire cascading process that generates distribution and dissemination through a hybrid information system (Giglietto et al., 2019). It is noteworthy the heterogeneity of definitions in the literature, but it is also interesting to note that what is shared across current definitions is how fake information appropriates the look of factual news and its effects. A journey through definitions might also represent, in this perspective, a recognition of news' place and role in our society, taking into account that, unlike in the past, «by misappropriating news' credibility, fake news might also undermine journalism's legitimacy, especially in a social media environment, when the actual source of information often gets removed or at least perceived at a distance» (Edson et al., 2017: 11). All these factors are present in the definition proposed by Lazer et al. which refers to fake news as «fabricated information that mimics news media content in form but not in organizational process or intent. Fake-news outlets, in turn, lack the news media's editorial norms and processes for ensuring the accuracy and credibility of information» (2018, p.1096). Consequently, the category of truth as a socially perceived standard of judgment is currently weakening. On the contrary, empirical truths, often based on individual perception and construction, multiply thanks to social media. In the post-truth era, one of the consequences of this phenomenon is a progressive strengthening of impermeable ideological spheres, increasingly apart from dissent and diversity, with a dangerous shift from free and different views to many absolute individual truths, without any chance for democratic debate. According to Bracciale and Grisolia, since it is not easy to understand who can correctly be considered as responsible for disinformation, the most promising perspective is that disinformation is a 'collaborative work' where social media users may unwillingly spread partially or totally false contents (Bracciale Grisolia, 2020: 66). Undoubtedly, the global crisis brought about by Coronavirus has enforced some information dynamics linked to social media and made many traditional weaknesses of the Italian information system more evident (see point 3).

Although there is still not a clear definition of this false information, or so-called "fake news", it is possible to affirm that such articles are born from a piece of news that is deliberately built from false and mendacious information but looks plausible and factual to the eyes of the reader. The aim is to manipulate and mold public opinion. While a factual news story is created by a journalist, fake news is a product of both the writer and the audience, since it only exists if it is perceived to be real (Tandoc et al., 2017). The Internet and social media are remarkably fertile grounds for alternative facts and fake news (Tandoc et al., 2018), since there is not a sufficiently coherent fact checking system on the Internet (Yun et al. 2017). Real and fake news blend

together and it is up to the user to filter through a massive mound of news stories according to one's own capabilities, often influenced by: (1) cognitive biases (confirmation bias, cherry picking), (2) the lack of willingness to fact check, (3) rarely adequate digital literacy, and (4) incohesive or entirely absent health literacy (i.e., the ability to obtain, read, understand, and use healthcare information in order to make appropriate health decisions and follow instructions for treatment) (Agenda Digitale; Sorensen et al., 2015). According to UNESCO, the term "fake news" does not have a straightforward or commonly understood meaning. In fact, "news" means verifiable information in the public interest; information that does not meet these standards does not deserve the label of "news". Therefore, "fake news" is an oxymoron that lends itself to undermining the credibility of the "real" news.

Thus, it would be better to use other terms to describe a phenomenon where there is a spreading of untruths via social media and the Internet, involving disinformation (false information deliberately created to harm a person, social group, organization, or country), misinformation (false but not created with the intention of causing harm), and mal-information (based on reality, used to inflict harm on a person, social group, organization, or country). The centrality of the topic of fake news in the Italian information system has aroused an increasing interest among scholars, who have proposed many definitions of the different gradations of 'information disorder' (Vaccaro et al., 2019; Bracciale & Grisolia, 2020), such as misinformation, disinformation or propaganda, focusing on the presence of manipulative, ironic or false elements in the news coverage (Riva, 2018). The evolution of this phenomenon, although linked to other open questions, such as the conceptualization of the post-truth era (Corner, 2017; Maddalena Gili, 2017; Lorusso, 2018), is also closely connected to two aspects which seem to characterize contemporary hybrid media ecosystem: the 'acceleration' of journalistic practices, due to the presence of digital technologies (Antenore, Splendore, 2017), and the success of the spectacular logic in newsmaking (Rizzuto, 2012). These aspects have brought about new informative formats, with a peculiar mix of factual and totally (or partially) false or incorrect content (Ireton, Posetti, 2018).

The Fake News Discourse is organised around a post-truth semiosphere, that predicts a

dishomogeneous distribution of credibility between the various actors. Fake news sites/actors look for credibility, and use heterogeneous tactics, i.e:

- Documentality (tactics employing "exhibition of proof"): the credibility is constituted by

the paper texture of the document displayed and the "bureaucratic" style

- Storytelling, usually employed by dramatizing minor details.

- Visual and textual exaggeration of messages. The use of capital letters and close-up

photographs have a very powerful emotional effect on the reader. This passionate charge

often constitutes the lowest common denominator of "indignant" online communication.

- No call for interpretation, that call for the reader not to try to interpret the news

or use their critical abilities

At the centre of all narration within the fake news discourse lies the problem of the relationship between words and things. Most of the pages spreading propaganda and fake news decry the

split between the world depicted through the mainstream media, and the reality of the facts, a world in which being and seeming do not coincide, a world dominated by secrets and lies. This representation translates into

a call to action to a mass (supposedly a majority) of defenceless citizens. «Let the people speak», as is often heard in populist propaganda, means asking that those we have called "silent majority", characterised by the coincidence of assumption and non-predication, may transform themselves into activists, agents of the revolution. The lowest level of activism is sharing content from propaganda pages with one's own contacts.

In this way, the individual asserts themselves as a (committed) activist (Mangiapane, 2018).

Turkey:

The spread and impact of fake news, trolls, and pro-denial misinformation in Türkiye are studied through several key theories. Agenda-Setting Theory explains how media influences public focus, shaping perceptions of reality. Framing Theory highlights how misinformation is structured to manipulate public opinion, particularly in political narratives. Social Identity Theory describes how group affiliations create echo chambers that reinforce false beliefs. Network Propaganda Theory. examines how coordinated networks spread disinformation, amplifying political and ideological bias. Researchers in Türkiye use big data analysis, AI-powered tools, social network analysis, and fact-checking initiatives to study and combat fake news. Platforms like Teyit.org and Doğruluk Payı play a critical role in verifying misinformation.

Portugal: No Answer

Luxembourg:

https://doi.org/10.1093/jcmc/zmac024

Abstract: As a significant source of misinformation, out-of-context visual misinformation refers to visuals presented in an unrelated context. This study explores whether a digital media literacy intervention that features reverse image search tools has significant effects on participants' message credibility judgment, discernment of visual misinformation, and intention of using reverse image search tools. Data were collected from a pre-registered, web-based, between-subjects experiment (N = 905). Results revealed a significant difference on intention of using reverse search tools among three experimental conditions: active intervention that involved both knowledge and behavior, passive intervention that involved knowledge only, and a control condition. Specifically, active intervention significantly increased intention of using reverse search tools, compared to the passive intervention and the control. Neither active nor passive intervention had an effect on credibility judgment or misinformation discernment. We discuss the implications for future digital media intervention designs and journalism practice that aim to combat visual misinformation.

Keywords: visual misinformation, "cheapfakes"

2. What are the distinctions between misinformation, disinformation, and fake news, including elements such as rumors and propaganda?

France:

Misinformation, disinformation, and fake news are three distinct but often interrelated phenomena that have significant implications for public understanding and democratic processes. Misinformation refers to the dissemination of false or inaccurate information, which occurs without any intent to deceive. People may unknowingly share incorrect facts or misinterpret information, leading to its spread (Polewka, 2023, 12-16). This can be as simple as a misunderstanding or an unintentional mistake, with no malice involved.

Disinformation, by contrast, involves the intentional spread of false information with the deliberate aim of misleading others. It is a more targeted form of manipulation, often used to sway public opinion or achieve political, social, or economic goals. Disinformation campaigns are typically strategic and carefully crafted to influence perceptions, especially in contexts like elections or geopolitical conflicts (Dauphin, 2023, p. 31-45). In this sense, disinformation goes beyond mere error; it is a calculated effort to deceive and manipulate.

Fake news represents a broader and more generalized category that includes intentionally fabricated content designed to deceive or mislead. While fake news is often associated with sensationalist headlines or fabricated stories, it also encompasses a wide range of misleading information, even if the original intent was not always malice. Fake news is particularly prevalent in digital media, where it can spread rapidly, attracting attention through sensationalism and the promise of shocking or controversial content. It is often tied to commercial interests, such as generating clicks or ad revenue, but can also serve ideological purposes (Rana et al., 2022, p.4). In contemporary discourse, "fake news" has become a catch-all term for any form of misleading or distorted information.

In addition to these terms, rumors and propaganda are important elements in the landscape of misleading information. Rumors are unverified pieces of information that circulate among people, often distorting facts or exaggerating details. These can be particularly dangerous in times of crisis, as they can exacerbate confusion and panic. Propaganda, on the other hand, refers to the strategic use of media to manipulate public opinion and shape perceptions. Propaganda often seeks to influence people emotionally, pushing political or ideological agendas through selective presentation of information (Polewka, 2023, p. 12-16).

The growing concerns about these phenomena have sparked critical discussions around the ethical implications of information manipulation. Scholars like (Piguet,2022, p. 15) have critiqued how certain "nudges" or subtle manipulations in media can erode individual autonomy, especially in the context of misinformation and disinformation. By strategically guiding people's choices and perceptions, these manipulative practices can undermine the foundations of democratic decision-making and citizen engagement.

Italy:

The term “fake news” falls within the broader concept of “misinformation”, defined as “information whose inaccuracy is unintentional” (Jack, 2017, p. 2). However, as Wardle and Derakhshan (2018) emphasize, this term is “inadequate to explain the scope of information pollution” and “problematic, thus we should avoid using it” (ibid., p. 43). They propose a

differentiation among various forms of incorrect information: misinformation, which, when intersecting with malinformation, becomes disinformation, creating a scenario of production and reproduction of informational disorder (Melchior & Strizzolo, 2024). The broad range of issues related to disinformation, including fake news, propagandistic content, satirical expressions, conspiracy theory constructions, deceptive fabrications, and virally disseminated visual content, represents a significant obstacle in our ongoing effort to interpret the information landscape we engage with on a daily basis. In order to fully understand the situation, it would be necessary to analyse and define them individually, but also, and above all, to consider their interconnectedness and their contribution to the overall complex and thorny communicative dynamics of the digital era. Indeed, we have already recognized the significance of categorizing fake news as an emerging form of online disinformation and as part of a broader issue. However, it is our responsibility to apprehend the intricate overlapping factors that fuel its production, distribution, and reception. In fact, it is imperative to embrace an integrated approach to media studies, or more precisely, to the study of the media circuit, in order to grasp all the interconnected facets of various forms of online disinformation (D'acci, 2004). This model compels us to analyse how ideas, discourses, and

practices exert influence, intersect, manifest incongruities, and disseminate within permeable

spheres of production, socio-historical contexts, reception, and cultural manifestations. Effectively addressing the current issue would necessitate a cultural mutation in our approach to understanding information, in the selection of reliable sources, and in the cognitive assimilation of our role in information management (Boyd, 2017).

Disinformation often thrives on polarizing and emotionally impactful topics (Humprecht, 2019), as demonstrated by numerous studies, including the European Union transnational survey conducted in July 2019 (Hameleers et al., 2021). According to this research, citizens are more exposed to false or inaccurate news regarding topics such as immigration, economy, environment, as well as international politics, terrorism, welfare, and the European Union. the different types of false information present on social networks, which can be classified based on their content, level of manipulation, and intentionality. Some types of false information are represented by entirely fabricated stories that have no connection to real events (Rubin et al., 2015). However, it is important to note that this type of false news is not attributed to a new phenomenon but dates back to the origins of journalism. Another type of false information is propaganda (Garth & O' Donnell, 2014), which represents a specific example of fabricated stories aimed at damaging the aspirations of a specific group or a political leader. This category of erroneous information is not a novelty, as it found widespread application during World War II and the Cold War era. Propagandistic narratives are frequently employed within political contexts with the aim of disorienting public opinion, pursuing the overarching goal of inflicting harm upon specific political parties or national entities.

This tactic is grounded in the strategic dissemination of misleading or distorted content, with the intention of influencing collective perceptions and undermining the reputation of political

adversaries. As a result, propaganda assumes a role of particular peril within the landscape of false information, as it possesses the capability to significantly influence the course of human history. This capacity becomes even more relevant when considering the potential implications within electoral competition contexts. The extent of the impact in shaping public opinion and molding critical citizens; decisions is an aspect that requires careful and thorough consideration.

The interconnection between propaganda and electoral dynamics, in fact, further underscores the necessity of fully comprehending the manipulative power of distorted and deceptive narratives in the realm of democratic choices. Another categorization includes Conspiracy Theories (Feldman, 2011), that pertain to narratives attempting to justify a situation or an event by invoking a conspiracy, without, however, providing concrete evidence in support of such a hypothesis.

Typically, such narratives involve illicit practices committed by governments or powerful

individuals holding prominent roles within society. The term Falsehood; (Battista, 2023) it refers to the dissemination of false or inaccurate information presented under the guise of authentic facts. However, within the research environment, this genre is also identified as half-truth; (Merriam- Webster, 2018). Another genre of false information consists of highly biased or one-sided narratives, also known as hyper partisan news (Potthast et al. 2017). These narratives exhibit a strong bias against individuals, parties, situations, or events and are often propagated in the political domain through unofficial channels such as fringe web communities.

Turkey:

Misinformation, disinformation, and fake news are related but distinct concepts. Misinformation refers to false or misleading information shared without intent to deceive, often spreading due to misunderstanding or lack of verification. Disinformation, on the other hand, is deliberately crafted false information designed to manipulate public opinion, frequently used in political propaganda. Fake news is a broader term referring to fabricated stories that mimic real journalism but serve political or financial motives. Additionally, rumors are unverified claims that gain traction through informal social sharing, while propaganda is systematically produced content intended to influence public perception, often state sponsored. In Türkiye, all these forms of misleading information are prevalent, especially in political discourse and crisis situations.

Portugal:

- Fake news, disinformation and misinformation are related concepts that differ mostly in the impact produced and felt. While misinformation can be defined as false information spread without the intent to mislead people (Costa, 2024), disinformation and fake news are spread with the deliberate intention to create misleading content on social media and traditional media, online or office and by deceiving people by mimicking the form of legitimate news (Centro Internet Segurança, n.d.; Chen, Xiao, & Kumar, 2023; Costa, 2024; Stepney & Lally, 2024).

- Rumours can be labelled as statements or stories that circulate, based on true or false affirmations (Chen, Xiao, & Kumar, 2023). For this, we can conclude that rumours, when shared innocently, can be a form of misinformation, but when shared with bad intentions, might pose as disinformation.

- Propaganda can be defined as the spreading of information, which can be true or false, destined to influence the public opinion on a certain topic (Costa, 2024). It's important to mention that just like rumours, propaganda might not always have a negative end.

Luxembourg: No Answer

3. Have there been instances of conspiracy theories and misinformation campaigns in your country?

France:

In France, as in many other countries, conspiracy theories and misinformation campaigns have been prominent, particularly in recent years. These phenomena have been particularly notable in the realms of politics and public health. A significant instance occurred during the COVID-19 pandemic, where the spread of misinformation surrounding the virus, its origins, and the vaccines became widespread. These falsehoods contributed to a climate of confusion, fear, and public mistrust, especially when combined with political conspiracy theories that undermined governmental measures. This complex interplay of misinformation has had serious consequences for public health efforts and political discourse.

(Barkun and Rollet, 2016, p. 168-176) argue that conspiracy theories, particularly in times of crisis, often gain traction because they provide simple explanations for complex events, appealing to those seeking certainty and control. In this context, the COVID-19 pandemic became fertile ground for such theories, as misinformation about the virus's origin and potential cures spread rapidly, amplified by social media platforms. These conspiracies are not just harmless fabrications but are part of a broader issue of knowledge stigmatization, which affects public understanding and trust in science and institutions.

(Compton et al., 2021, 22) further explore how conspiracy theories and misinformation thrive in the "post-truth" era, where contested science and digital media amplify false narratives. They suggest that inoculation theory, which involves pre-emptively debunking misinformation to prevent its spread, could be a useful tool in combating these harmful ideologies. In the case of France, where misinformation about vaccines and public health measures proliferated, this approach could help mitigate the widespread impact of conspiracy-driven narratives.

Italy:

In the Italian information system, often presented by scholars as an anomaly among Western

countries (Hallin, Mancini 2004; Rizzuto, 2009), there has always been a peculiar relationship between newsmedia and politics, considering media as means of political mobilization and using them not to inform but to participate to politics. In the last three decades, the emergence of commercial television and the success of its language and formats have made a radical change possible, bringing about a profound crisis of the social role of journalism, strongly debated by scholars (Murialdi, 2014; Morcellini, 2011). As a consequence, the contemporary communication circuit between journalists and citizens outlines a new information ecosystem (Valentini, 2012), a framework in which there is a changed perception of the meaning of journalism itself, of its ability to 'tell stories' about real life as a serious activity connecting people to the world. A softer and more recreational use of news is privileged and imposed by infotainment, with its dramatizing and emotion-based logic, which has made more and more problematic the traditional separation between factual and false in news reporting. In this perspective, with its narrative styles and lexical choices, infotainment can be linked to the problematic topic of the increasing presence of fake news, damaging journalism and putting into crisis its social responsibility. Due to the predominance

of the entertainment perspective in the selection of events and the definition of them

newsworthiness, journalism tends to offer dramatized reconstructions of parts of reality, focusing on the visual impact of the events and the narrative strategy, too often used to confirm prejudices and stereotypes (Zelizer, 2004). Consequently, perceptions of social problems and priorities are more and more influenced by these media-narratives that are appealing but also exaggerated and distorted (Polesana, 2010; Rizzuto, 2012), since many events are selected and presented only because of their dramatic content or because images are available. In the last decades, the role played by Italian newsmedia in the perception of the difference between real and false facts has weakened due to many technological and social factors (Antenore, Splendore 2017). With the spread of fake news even more so, causing an increasing interest from experts and scholars in the academic debate (Albright, 2017; Edson et al., 2017; Gili Maddalena, 2017). As we noted earlier, the evolution of this phenomenon is not new. Still, it can be easily connected also to the digital 'acceleration' of journalistic practices and the informative immediacy of social media, which have been creating a digital context where false or distorted information is favoured. The lack of professional control by newsrooms in online information and the increasing viral spread of disinformation have become a cultural and political problem in the Coronavirus age, with the risk that news could have contributed to reinforcing distortions of topics such as the origin of the virus or wrong medical remedies, making invented facts as tools to interpret the world and even legitimating dangerous behaviours through specific discursive forms, narrative and rhetorical structures. In an emergence period, fake news spreads quickly, too often functioning as fuel for an information overload process, which obliges users to make unprecedented cognitive efforts for tackling enormous amounts of data (Giglietto et al.2018). A UNISOB (2020) study claims that 82% of Italians are not able to recognize fake information. In the 2024 edition of the Digital Decade report, Italy has achieved 45.8% basic digital skills coverage, compared to the EU average of 55.6%. The score in Italy has slightly improved comparing to 2023 results (45.6%). According to the Digital Decade report 2024, Italy performs slightly below the EU average in both digital skills indicators. A pilot study in 2019 showed that in Italy only 62.3% of the studied population had adequate skills concerning reasoning and autonomy, while the rest had inadequate (11.2%) or problematic skills (26.5%) when it came to health literacy (Palmieri et al., 2020). Fake news" has inevitably and indubitably influenced health communication in the Covid-19 emergency, as part of the phenomenon called infodemia (Zarocostas, 2020). Digital and health illiteracy can account for most of the inaccurate news stories depicting the cause of the pandemic of Covid19, whether suggesting it is a virus engineered purposely in a Wuhan laboratory, a biological military weapon, or a result of 5G (5 th Generation) technology being implemented in many countries, or news stories claiming that as-yet undiscovered vaccines for SARS-CoV-2 are harmful and useless. Such fake stories were intentionally built to rapidly and tumultuously spread over the Internet and social media, and their popularity rapidly increased, such that retractions were always published too late.

In Italy there are many non-profit organisations, foundations and fact-checking projects that deal with combating disinformation and fake news on different levels. Some examples of active fact- checking organisations in Italy are the fact-checking section part of Agenzia Giornalistica Italiana (AGI); the fact-checking portal Pagella Politica, which has obtained IFCN certification and

collaborates with Facebook; FACTA news, a spinoff of Pagella Politica; the fact-checking portal Bufale.net; the fact-checking blog Butac.it; the independent magazine Lavoce.info; and the fact- checking portal of the project OpenMigration, amongst others.

Turkey:

Türkiye has experienced several large-scale misinformation campaigns and conspiracy theories. One prominent example is the Gezi Park protests (2013), where the government and pro-government media framed demonstrators as foreign-backed agents attempting to destabilize the country. Another example is COVID-19 disinformation, where false claims about vaccine microchips, secret government experiments, and exaggerated death tolls circulated widely on social media. Political misinformation is also common, with fabricated news reports targeting opposition figures, often spread by political bots and trolls. These examples demonstrate how misinformation in Türkiye is frequently used to influence public perception and reinforce political narratives.

Portugal:

- In Portugal, there have been cases where conspiracy theories, disinformation and misinformation, particularly in recent years, have spread to the general population. One of the more prominent disinformation and conspiracy campaigns was the information around the pandemic COVID-19 that was spread by the now extinct negationist Spanish group that later spread to Portugal 'Médicos pela Verdade'. This group criticised the lockdowns, the use of masks and the use of PCR tests to identify the presence of COVID-19, using mainly opinion articles and studies that were globally refuted by the scientific community (Sábado, n.d.). Another disinformation campaign related to COVID-19 pandemic was spread by a former judge called Rui Fonseca e Castro, who through his actions and words, promoted civil disobedience against confinement, defending theories that denied the seriousness and the existence of the global pandemic. The former judge also appeared on an ultraconservative television channel 'El Toro TV' where he stated that Portuguese men were at risk of becoming sterile due to the COVID-19 vaccine (Pardar, 2023).

- On a different topic, a Facebook post spread with the intent of spreading disinformation considered Portugal the 33rd most corrupt country in the world. However, a fact-check conducted by Poligráfo showed that Portugal holds the 33rd position as the least corrupt country in the world (Pardar, 2023).

Luxembourg:

Portugal and Spain

https://iberifier.eu/2024/05/15/iberifier-reports-disinformation-consumption-patterns-in-spain-and-portugal/

IBERIFIER Reports – "Disinformation consumption patterns in Spain and Portugal" (2024)

In this report we conduct a systematic analysis of the matters pertaining to the spread and evolution of disinformation in Spain and Portugal, in terms of forms of consumption, platforms, narratives and permanent adaptation to unfolding events in both countries. The analysis is based on a thorough overview of IBERIFIER reports over the course of four years, third-party data sources such as the Digital News Report and EC Eurobarometer and third-party reports tackling the issues in both countries.

As future recommendations, we suggest that forthcoming initiatives on these matters should follow a strict continuity and innovation logic. Continuity as in the:

Adoption of psychological-social driven frameworks which subjectively frame the perception of audiences and their behaviours towards information and disinformation, and exploration of negative behavioral patterns (drop in interest for news or news avoidance).

Thorough approaches surrounding specific events such as national and regional electoral processes, armed conflicts, etc.

A deeper understanding of the circulation of disinformation content on social networks, from a supply and demand holistic perspective. Persistence of both multidisciplinary and multi-institutional approaches in detriment of isolated, "island"-logic approaches.

Increased investment in media and news literacy among the population.

Convergence on political and regulatory frameworks.

4. From a general perspective, how could the algorithms play a role in the dissemination of fake news, and how do they impact public perception?

France:

Algorithms, particularly those utilized by social media platforms and news aggregators, play a crucial role in the dissemination of fake news. These algorithms are primarily designed to maximize user engagement, and this goal often results in prioritizing content that evokes strong emotional reactions or sensationalism, characteristics commonly associated with fake news. By relying on engagement metrics such as likes, shares, and comments, these algorithms tend to promote content that attracts attention, regardless of its factual accuracy. As a result, misinformation can quickly gain traction, spreading far beyond its initial audience through recommendation systems that amplify highly engaging yet misleading or false content. This process creates a feedback loop that fuels the proliferation of fake news, as users are continually presented with content that aligns with their emotional responses, whether accurate or not

Additionally, algorithms often contribute to the creation of filter bubbles, which are spaces where users are repeatedly exposed to similar types of content that reinforce their preexisting beliefs. This selective exposure significantly limits the diversity of information that users encounter, reducing the likelihood that they will encounter corrective or accurate narratives. In this context, individuals are less likely to come across fact-checking efforts or rebuttals to the misinformation they are exposed to. The result is a distortion of public perception, where users become increasingly susceptible to believing false narratives or conspiracies. This not only impacts individual beliefs but also undermines collective understanding by perpetuating a skewed version of reality.

In a broader context, the algorithmic amplification of misinformation can be understood as part of a power struggle in the public discourse surrounding the "truth." As Fake-News! (n.d.) discusses, the dissemination of fake news is not simply a technical issue, but a conflict over who has the authority to define what is true. Algorithms, by promoting content based on engagement rather than veracity, effectively empower those who can manipulate these

platforms to spread false or misleading information, thus influencing public opinion and shaping the political and social narratives of the digital age.

The impact of these algorithmic systems on public perception has been widely examined. For example, (Friggeri et al.,2014, 101-110) discuss how rumor cascades, a concept closely linked to the spread of fake news, occur when misinformation rapidly propagates through networks, facilitated by algorithms that prioritize sensational content. The rapid dissemination of false narratives not only distorts individual perceptions but can also shape societal norms and behaviors, reinforcing divisions and misconceptions within the public sphere.

Italy:

Fake news also needs to be related to the increasing commodification of the internet (fake news as clickbait are highly lucrative) and the algorithms' power to customize people's online environments trapping them into homophilic circles, as said. Research shows that, in fact, people want to remain in such circles where they can comfortably share news (no matter if fake) that appeals to their existing worldviews (hence generating more income for social media companies) (Seargeant & Tagg, 2014). Such findings not only question long-established positions that blame the media without considering the uses (or misuses) people make of them but also highlight once again the critical role of education. In other words, the idea that algorithms are responsible for filter bubbles suggests it should be easy to fix the problem by simply getting rid of them. However, this perspective ignores how users themselves effectively create their own filter bubbles by withdrawing from political controversies and hiding opinions they disagree with. Moreover, we need to understand why the algorithms that govern these differences are mostly invisible to ordinary users. We need to understand that they are governed by a logic that we cannot hastily settle as a purely technical question as they are in fact the result of specific choices as to what the algorithm must do, what types of data must be collected, how different data must relate to each other and, ultimately, what kind of profit (economic or political) can be done (and by whom) from these data.

Paraphrasing two of Melvin Krantzberg's laws (1986): 1) the algorithm is neither good nor bad; nor is it neutral; 2) although algorithms might be a prime element in many public issues, nontechnical factors take precedence in algorithm-policy decisions. Techno-giants (such as Google or Facebook) have always defended the neutrality of their infrastructures and the freedom of expression they allow, blaming people for any incorrect, if not openly illegal, use of that freedom. As Buckingham states, they «are not merely technology companies: they are also media companies. It is via the internet – and via these digital platforms and services in particular – that we are

increasingly accessing media of all kinds. These companies are not merely supplying us with

technical devices or tools, hardware, or software. They are also increasingly providing the means of representation and communication that are indispensable to modern life» (2019, p.13-14). Using one communication tool rather than another (whether new or old) inevitably implies resorting to a certain system of mediation and construction of knowledge. In this sense, all communication technologies are not 'neutral' vehicles of knowledge; they are also, inevitably, media that 'mediate' the relationship between individuals, knowledge, and power.

Turkey:

Algorithms play a crucial role in spreading fake news by prioritizing engagement over accuracy. Social media platforms like Facebook, Twitter, and TikTok use machine-learning algorithms that promote content based on user interactions, often amplifying sensational or misleading information. Echo chambers and filter bubbles emerge when users are repeatedly exposed to content aligning with their existing beliefs, reinforcing misinformation. In Türkiye, politically motivated bot networks and coordinated disinformation campaigns take advantage of these algorithms to push propaganda and suppress dissenting voices. As a result, fake news spreads faster than factual corrections, significantly shaping public opinion and increasing societal polarization. To mitigate these effects, fact-checking initiatives, media literacy education, and algorithmic transparency are necessary to limit the spread of disinformation.

Portugal: No Answer

Luxembourg:

Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators (2022) – European Commission

https://op.europa.eu/en/publication-detail/-/publication/d81a0d54-5348-11ed-92ed-01aa75ed71a1/language-en

These ethical guidelines on AI and data usage in teaching and learning are designed to help educators understand the potential that the applications of AI and data usage can have in education and to raise awareness of the possible risks so that they are able to engage positively, critically and ethically with AI systems and exploit their full potential.

Includes a very interesting and simple glossary

III. Ethics and Digital Responsibility

a) Personal Data Protection

1. What measures are currently implemented in your country to ensure the protection of personal data in digital environments?

France:

In France, several legislative, regulatory, and institutional measures have been implemented to safeguard personal data in digital environments, reflecting the growing importance of privacy and data protection in the digital age. These measures are grounded in European Union law, national legislation, and independent oversight bodies, ensuring a robust framework for protecting citizens' personal data.

1. General Data Protection Regulation (GDPR): As a member state of the European Union, France adheres to the General Data Protection Regulation (GDPR), which represents one of the most comprehensive and stringent data protection frameworks globally. The GDPR, which came into effect on May 25, 2018, imposes strict guidelines on data collection, processing, storage, and transfer. It prioritizes individuals' rights to privacy and data protection, establishing clear protocols for consent, transparency, and accountability. Among its core provisions are the

right to access, rectify, and erase personal data, as well as the right to object to certain types of data processing. It also mandates data protection impact assessments and requires organizations to appoint Data Protection Officers (DPOs) to ensure compliance.

2. Loi pour une République numérique (Digital Republic Act): Enacted on October 7, 2016, the Digital Republic Act (Loi pour une République numérique) is a significant piece of national legislation aimed at enhancing transparency, promoting open data, and strengthening individual rights in the digital environment. This law emphasizes user consent, the right to be forgotten, and the need for greater control over personal data by citizens. For example, it introduces provisions that ensure individuals can request the deletion of their data from online platforms under certain conditions. The law also aims to enhance user awareness of how their personal data is collected and used, reinforcing the transparency required under the GDPR.

3. Commission Nationale de l'Informatique et des Libertés (CNIL): The CNIL is France's independent regulatory authority for data protection and privacy. Established in 1978, the CNIL is tasked with ensuring compliance with both the GDPR and the Digital Republic Act. The CNIL plays a critical role in upholding privacy rights by offering guidance to both individuals and organizations, conducting investigations, and enforcing regulations through sanctions. The CNIL has the power to impose fines and corrective measures on entities that fail to comply with data protection laws, making it a key player in the enforcement of personal data protection in France. Furthermore, the CNIL provides educational resources to the public, fostering a deeper understanding of data protection rights and practices.

4. Privacy by Design: The principle of "Privacy by Design" is integral to data protection measures in France and is enshrined in both the GDPR and national legislation. This principle mandates that privacy considerations be embedded into the design and operation of digital systems and processes from the outset. In practice, this means that organizations must incorporate data protection features, such as anonymization and encryption, during the development phase of digital services and products. By proactively addressing privacy risks, "Privacy by Design" aims to minimize potential threats to personal data before they materialize, rather than responding reactively to breaches or violations.

5. Public Awareness and Practices: Public awareness of data protection issues has significantly increased in France in recent years, driven in part by the implementation of the GDPR and the Digital Republic Act. Surveys indicate that French citizens are increasingly proactive in managing their personal data online, employing various tools and practices to safeguard their privacy. Many individuals use ad-blockers to prevent targeted advertising and data tracking, adjust privacy settings on social media platforms, and adopt password managers to secure their online accounts. This growing awareness is indicative of a societal shift towards greater digital literacy and personal responsibility regarding data privacy.

France has established a comprehensive framework for the protection of personal data in digital environments. By combining strong legal protections, independent regulatory oversight, and a focus on public awareness, these measures aim to ensure that personal data is handled with the utmost care and respect for individuals' privacy rights. However, as digital environments continue to evolve, ongoing adaptation and enforcement will be necessary to address emerging challenges in data protection.

Italy:

In Italy, the regulation of personal data protection in digital environments is primarily governed by the General Data Protection Regulation (GDPR), an EU-wide framework that was formally incorporated into Italian law through Legislative Decree No. 101/2018. This decree brought amendments to the Italian Data Protection Code, originally established by Legislative Decree No. 196/2003, ensuring its alignment with the GDPR's stringent requirements.

At the national level, the primary authority overseeing compliance with these regulations is the Italian Data Protection Authority, known as the Garante per la Protezione dei Dati Personali. This independent body is responsible for ensuring that both public and private entities adhere to the principles of data protection. Among the core measures in place to safeguard personal data, organizations are required to adhere to the principle of data minimization, meaning that they may only collect and process information that is strictly necessary for the intended purpose. Additionally, individuals are granted extensive rights over their personal data, including the ability to access, rectify, erase, and transfer their information.

To further enhance protection, certain high-risk data processing activities are subject to mandatory Data Protection Impact Assessments (DPIAs) to evaluate and mitigate potential risks. Organizations must also implement robust security standards through both technical and organizational measures, ensuring the confidentiality, integrity, and availability of personal data. In cases where data breaches occur, regulations require organizations to notify the Garante within 72 hours if there is a risk to individuals' rights and freedoms. Encryption and other security measures are also mandated to safeguard sensitive information.

Failure to comply with these regulations can lead to severe financial penalties. Companies found in violation of data protection laws may face fines of up to €20 million or 4% of their global annual turnover, depending on which amount is higher. Beyond these general provisions, Italy has also implemented sector-specific regulations that impose additional restrictions on data processing, particularly in areas such as healthcare, where the Electronic Health Record (Fascicolo Sanitario Elettronico) plays a crucial role, and finance, where anti-money laundering measures further regulate data usage.

In addition to adhering to national regulations, institutions such as the University of Calabria (UNICAL) have developed their own internal policies to reinforce data protection within their structures. These policies provide detailed information on personal data processing, defining the responsibilities of key figures such as the Data Controller and the Data Protection Officer (DPO). They also specify the types of personal data processed and outline the rights available to data subjects in accordance with both GDPR provisions and university-specific policies.

To effectively manage data breaches, UNICAL has established a comprehensive set of guidelines that outline procedures for detecting, assessing, and reporting incidents. These guidelines categorize breaches into three main types: confidentiality breaches, which involve unauthorized access to personal data; availability breaches, which occur when data is lost or made unavailable due to accidental or unauthorized actions; and integrity breaches, where data is accidentally or unlawfully modified. In line with national requirements, all breaches must be documented in a Data Breach Register, and if deemed necessary, reported to the Garante per la Protezione dei Dati Personali within the legally prescribed timeframe of 72 hours.

By implementing these stringent national and institutional measures, Italy continues to strengthen its commitment to data protection, ensuring that personal information is handled responsibly and securely across various sectors.

Turkey:

In Türkiye, personal data protection is primarily governed by the Personal Data Protection Law, which aligns with the EU General Data Protection Regulation (GDPR). The Personal Data Protection Authority (KVKK Kurumu) is responsible for overseeing compliance and investigating data breaches. Businesses and organizations are required to obtain explicit user consent before processing personal data, ensure data minimization, and implement security measures such as encryption and access controls. Additionally, Türkiye mandates Data Controllers Registry (VERBİS) registration, requiring companies to declare their data-processing activities to increase accountability.

Portugal:

- GDPR Compliance: Portugal has been a key player in implementing GDPR (General Data Protection Regulation) in the European Union. This regulation gives citizens control over their personal data and strengthens trust in digital platforms.

- The National Data Protection Commission (Comissão Nacional de Proteção de Dados - CNPD): The CNPD is an independent administrative body with a legal personality that has administrative and financial autonomy to monitor and enforce compliance with General Data Protection Regulation, aiming to protect the rights and guarantees of individuals concerning the processing of their personal data (Comissão Nacional de Proteção de Dados, n.d).

Luxembourg: No Answer

2. What measures are in place to ensure transparency and obtain user consent regarding data collection and usage?

France:

To ensure transparency and obtain user consent regarding data collection and usage, organizations must adhere to several key principles, as outlined by legal frameworks such as the General Data Protection Regulation (GDPR). The GDPR mandates that entities handling personal data provide clear and accessible information about the data collection process, the purposes for which the data is being used, and the rights of the users. This is typically achieved through the provision of comprehensive privacy policies, user consent forms, and data processing notices, all of which are essential for ensuring that users can make informed decisions about their data.

As highlighted by (Voigt and Von dem Bussche,2017, p. 17-33), the GDPR requires that users are informed about their data rights and how their data will be processed, stored, and shared. This information must be communicated in a way that is understandable and easily accessible to users, ensuring that consent is truly informed. The regulation emphasizes the importance of obtaining explicit consent from users before data collection begins, particularly when sensitive

data is involved. Additionally, it stipulates that user must be able to withdraw consent at any time, reinforcing the concept of ongoing control over their personal information.

Moreover, the GDPR mandates that organizations implement user control mechanisms, such as opt-in consent for data sharing and customizable privacy settings, allowing individuals to tailor the level of data access they grant to various services. These mechanisms not only ensure transparency but also give users greater autonomy over their personal data, which is crucial for maintaining trust in digital platforms.

In addition to these legal frameworks, (Andoulsi, 2022, p.17-33) emphasizes the importance of practical implementation in sectors such as healthcare, where the GDPR's principles are applied to sensitive data like health information. He argues that medical practices and organizations dealing with personal health data must pay particular attention to obtaining consent and ensuring data protection compliance. Andoulsi highlights that, in these contexts, professionals must provide clear and detailed information regarding the processing of medical data, ensure that patients' rights are respected, and create user-friendly mechanisms for consent. His work underscores the necessity of not only adhering to legal requirements but also integrating them into day-to-day operations in a manner that is both practical and comprehensible for users.

Together, these measures promote a high level of transparency and user empowerment, as individuals are not only informed but also given substantial control over their personal data. This combination of legal requirements and practical strategies forms the backbone of privacy protection under the GDPR and is crucial in fostering a culture of data security and trust in both the public and private sectors.

Italy:

In Italy, ensuring transparency in data collection and processing is a fundamental requirement under data protection regulations. Organizations must obtain explicit and informed consent from users before processing their personal data, particularly in cases involving direct marketing, profiling, or the use of biometric data. This requirement ensures that individuals have full awareness of how their information is being handled and can make informed decisions about granting or withholding their consent. Additionally, companies and institutions are obligated to provide clear and accessible privacy notices that specify what data is collected, how it is used, and what rights users have in relation to their personal information.

To further enhance transparency, Italy adheres to the ePrivacy Directive, which regulates the use of cookies on websites. Under this framework, websites must obtain explicit user consent before tracking online behavior through cookies, except in cases where the cookies are strictly necessary for basic site functionality. Users also have the right to opt out of data processing at any time without needing to provide a justification, ensuring they retain control over their personal information.

In the public sector, the Italian government has implemented digital services such as SPID (Sistema Pubblico di Identità Digitale), a secure authentication system that enables citizens to access online public services in a transparent and efficient manner.

At the University of Calabria (UNICAL), transparency in data protection is reinforced through specific policies that apply to students, employees, and online users. The university provides detailed privacy information regarding the collection and processing of personal data from

students and staff, covering sensitive areas such as health information, union membership, and financial data. These policies also clarify the legal bases for data processing, ensuring that individuals understand the justification behind the collection and use of their data.

To address online privacy concerns, UNICAL has implemented an extended cookie policy that explains how cookies are used on its website and provides guidance on managing cookie preferences. In the field of digital learning, the university has established an E-Learning Privacy and Copyright Notice, which outlines how personal data is handled on its educational platforms. This policy also strictly prohibits the unauthorized sharing or commercial use of educational materials, safeguarding both users' data and intellectual property.

Additionally, UNICAL has developed a structured approach to data breach management, which includes internal mechanisms for students and staff to report privacy concerns. Clear protocols are in place to ensure that affected individuals are promptly notified in the event of a data breach. To foster a culture of digital responsibility, the university also offers training sessions on GDPR compliance, helping students and staff understand their rights and responsibilities in relation to data protection.

By implementing these measures, both at the national and institutional levels, Italy and its academic institutions continue to uphold high standards of transparency and accountability in data protection, ensuring that individuals can trust how their personal information is being used.

Turkey:

Transparency and user consent are enforced through privacy policies, cookie consent mechanisms, and opt-in requirements. Organizations operating in Türkiye must provide clear and accessible privacy notices, informing users about what data is collected, why it is processed, and who it is shared with. Websites and digital platforms must implement cookie banners that allow users to accept or reject tracking cookies. Additionally, individuals have the right to request data deletion (right to be forgotten) and the right to access their stored personal data. However, enforcement varies, and some companies do not fully comply with transparency obligations.

Portugal:

- Article 7 'Conditions for consent' of the GDPR: Before collecting personal data, organisations must ensure that users give explicit, informed consent. This means that users must be provided with clear and accessible information about how their data will be used, why it is being collected, and for how long it will be retained (European Union, 2016).

- Cookie Consent: In line with both GDPR and the ePrivacy Directive, websites in Portugal (and the broader EU) are required to obtain explicit consent before using cookies or other tracking technologies. This is typically done via a cookie banner or pop-up that clearly informs users about the use of cookies and allows them to accept or reject them (GDPR, n.d).

Luxembourg: No Answer

3. What are the challenges involved in balancing data protection with service functionality?

France:

Balancing data protection with service functionality presents a range of significant challenges, particularly in the context of the increasing reliance on digital services and the complex regulatory environment that governs data privacy. A primary challenge lies in maintaining user privacy while ensuring that services remain efficient, effective, and user-friendly. Many digital services, such as personalized recommendations, location-based services, and social media platforms, require access to large volumes of personal data in order to function optimally. However, this can conflict with stringent privacy regulations, such as the General Data Protection Regulation (GDPR) in the European Union, which imposes strict guidelines on how personal data should be collected, processed, and stored. Organizations must find ways to balance the need for personal data to enhance service functionality with the requirements to protect that data and maintain user privacy.

A further challenge is ensuring that the implementation of robust security measures—such as encryption, secure access controls, and data anonymization—does not undermine the performance of the service. This is particularly important in real-time or high-demand environments, where delays or slowdowns caused by excessive security protocols can impact the user experience. In addition, the integration of privacy-enhancing technologies must be carefully designed so as not to create friction in user interactions or degrade the overall service quality. For example, the increased use of encryption and authentication processes can lead to longer processing times, which may hinder the seamless operation of services.

Moreover, organizations face the ongoing challenge of navigating the complexities of obtaining informed consent from users and ensuring transparency in how personal data is collected, stored, and used. The concept of informed consent, central to data protection laws, requires that users are fully aware of the implications of sharing their personal data and how it will be used. This process can be cumbersome and may negatively affect the user experience, especially when it involves lengthy terms and conditions or complicated consent mechanisms. (Rabagny-Lagoa,2022, p. 9-10) highlights that, in the context of data protection, organizations must not only inform users of their rights but also ensure that data practices are clear and understandable. This transparency is crucial for fostering trust between users and service providers and ensuring compliance with data protection laws.

Another layer of complexity arises when organizations operate across multiple jurisdictions, each with its own set of data protection regulations. As-(Lagoa,2022, p. 9-10) points out, the evolving nature of data protection laws worldwide means that businesses must continuously update their data processing practices to stay compliant with new legal frameworks. The diversity of these regulations creates challenges in harmonizing privacy measures across services, particularly when global operations are involved.

(Zhou et al., 2018, p. 122, 1-15) discuss the challenges involved in maintaining data privacy while providing cloud-based services, focusing on the balance between service functionality and data security. Their research emphasizes the technical and regulatory aspects that impact the design of secure and efficient cloud services, shedding light on the ongoing struggle between maintaining robust data protection and ensuring high service availability and functionality.

In conclusion, balancing data protection with service functionality requires a multifaceted approach. It involves addressing technical, regulatory, and operational challenges to ensure that services can function as intended without compromising user privacy or data security. As data protection laws continue to evolve, organizations must remain agile and responsive to these challenges to maintain both compliance and a high-quality user experience.

Italy:

While data protection regulations significantly enhance user privacy, they also introduce a range of challenges that organizations must navigate. One of the most pressing issues is the operational complexity associated with GDPR compliance. Ensuring adherence to these regulations requires continuous monitoring, regular assessments, and substantial administrative efforts. For businesses, particularly small and medium-sized enterprises (SMEs), the financial and logistical burdens of compliance can be particularly overwhelming, as they often lack the resources needed to implement robust data protection measures while maintaining efficiency.

Cross-border data transfers have also become a complex issue, particularly following the invalidation of the Privacy Shield Agreement between the European Union and the United States. This legal development has created significant uncertainty for multinational companies operating in Italy, as they must now navigate stricter requirements when transferring personal data across borders. Without a stable legal framework for transatlantic data flows, businesses face increased compliance risks and operational challenges.

On a broader scale, Italy's digital transformation efforts face obstacles, particularly in the public sector. The country ranks relatively low in the Digital Economy and Society Index (DESI), reflecting persistent challenges in digital literacy and awareness among citizens. This gap hinders the effective implementation of digital privacy tools and security measures, making it more difficult to ensure that individuals fully understand and exercise their data protection rights.

At the University of Calabria, these challenges are addressed through a dedicated Incident Response Team (IRT), which plays a critical role in safeguarding digital privacy and security within the institution. In the event of a data breach or security incident, the team rapidly assesses risks to determine the severity of the situation. Targeted mitigation strategies are then implemented to minimize service disruptions while maintaining strict privacy standards. Additionally, the university ensures that any breaches are reported promptly to the Garante per la Protezione dei Dati Personali and affected individuals, in full compliance with regulatory requirements.

By proactively addressing these challenges, both at a national level and within institutions like Unical, efforts continue to strengthen data protection frameworks while maintaining a balance between security, usability, and ethical considerations.

Turkey:

The main challenge in balancing data protection and service functionality lies in security versus user experience. Strict regulations may limit how companies use data for personalization, affecting customer experience, targeted advertising, and AI-based recommendations (Yıldız, 2022). Additionally, businesses face costly compliance requirements, such as data anonymization and storage regulations. Another issue is government access to personal data, as

concerns remain about state surveillance and content censorship under cybersecurity laws (Yesil & Sözeri, 2018). Striking a balance between privacy rights, innovation, and security remains an ongoing challenge.

Portugal:

General:

- One of the main challenges faced by service providers in managing data protection is tackling and/or preventing data breaches or cyberattacks. This is an urgent and pressing concern for service providers, as the increased use of online personal data across digital platforms in people's daily lives has created significant challenges. The main problem lies in the fact that it is difficult for an individual or an organisation to understand the magnitude of the implications of a data breach before it occurs (Zhang, et al., 2023).

- Another pressing challenge, transversal to all countries, is the fact that laws and regulations around data protection vary from country to country, making it difficult for services or companies that operate in multiple countries to comply with all of them (Mishova, 2024).

- Another challenge is ensuring that companies and service providers manage people's data and the utilisation of AI mechanisms while guaranteeing respect for people's rights (Group, FI, 2025).

Luxembourg: No Answer

4. Can you share any good practices or case studies related to data protection in your organization?

France:

One good practice for data protection in organizations is data anonymization, which helps protect sensitive personal information while still enabling analysis. For instance, the European Union's General Data Protection Regulation (GDPR) emphasizes anonymizing data as a key approach to ensure privacy and reduce the risks associated with data breaches. A case study conducted by (Nguyen et al.,2022, p. 1-37) discusses how various organizations implemented anonymization techniques to protect personal data while still achieving valuable insights for decision-making processes.

Italy:

Italy has implemented several best practices to enhance data protection and digital security, particularly in the public sector. One notable initiative is PagoPA, a government-backed digital payment platform designed to facilitate secure and GDPR-compliant transactions for public services. By incorporating strong authentication mechanisms, PagoPA ensures that users' financial and personal data remain protected during online transactions with public institutions.

Another crucial element in Italy's data protection framework is the active enforcement of regulations by the Garante per la Protezione dei Dati Personali. This regulatory body consistently monitors compliance with GDPR and acts against companies that fail to adhere to data protection laws. Recent enforcement measures have included significant sanctions against major corporations such as TIM (Telecom Italia) for improper telemarketing practices and Meta

(formerly Facebook) for unlawful data processing in targeted advertising. These actions reinforce the importance of data privacy and serve as a deterrent against violations.

Additionally, the Italian Agency for Digitalization (AgID) has established guidelines that set minimum ICT security standards for public administrations. These regulations ensure that Italian institutions implement robust cybersecurity measures, safeguarding both government and citizen data from potential threats.

At the University of Calabria, specific practices have been developed to uphold data protection standards and mitigate security risks. A key aspect of Unical's approach is its structured Data Breach Response Framework, which ensures swift action in the event of a security incident. The university provides a dedicated email system for reporting security breaches, allowing incidents to be promptly assessed and addressed. To evaluate the severity of breaches, Unical follows the guidelines established by the European Union Agency for Cybersecurity (ENISA), ensuring that responses align with best practices at the European level. Moreover, the university employs encrypted backups as a preventive measure to safeguard against data availability breaches, ensuring that critical information remains protected even in cases of cyber incidents.

Transparency in data processing is another essential pillar of Unical's data protection strategy. The university actively communicates its privacy policies to both students and staff, ensuring that individuals understand how their personal data is collected, stored, and used. In addition, the institution places significant emphasis on securing student data within its e-learning platforms, guaranteeing compliance with privacy regulations and maintaining the confidentiality of academic records.

References

General Data Protection Regulation (GDPR) (Regulation (EU) 2016/679): https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

Italian Data Protection Code (Legislative Decree No. 196/2003): https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/1311248

Legislative Decree No. 101/2018: https://www.gazzettaufficiale.it/eli/id/2018/09/04/18G00129/sg

Garante per la Protezione dei Dati Personali (Italian Data Protection Authority): https://www.garanteprivacy.it/

University of Calabria Data Breach Management Guidelines: https://www.unical.it/portale/portaltemplates/view/view.cfm?10115

University of Calabria privacy policies and regulations: https://www.unical.it/privacy/?lang=en

Agenzia per l'Italia Digitale (AgID): https://www.agid.gov.it/

PagoPA: https://www.pagopa.gov.it/

European Union Agency for Cybersecurity (ENISA): https://www.enisa.europa.eu/

Turkey:

A successful example of data protection in Türkiye is the Turkish Banking Sector, where strong encryption, two-factor authentication, and data minimization practices are widely implemented.

Banks must adhere to both KVKK and Banking Regulation and Supervision Agency (BDDK) guidelines, ensuring secure storage and processing of customer data. Another good practice comes from e-Government (e-Devlet) services, where two-step verification and data access logs provide citizens with greater transparency over who accesses their data (BTK, 2022). Additionally, Teyit.org, a fact-checking organization, follows ethical data collection policies, ensuring user privacy in research.

Portugal:

- At IPS, all employees had to attend an e-learning course around Data Protection Regulations in Portugal, to stay up to date with the correct regulations and laws that regulate data protection in the country.

Luxembourg: No Answer

b) Inclusion and Diversity in Digital Spaces

1. What strategies are implemented to promote digital inclusion and diversity in online spaces?

France:

Strategies to promote digital inclusion and diversity in online spaces typically focus on increasing accessibility, education, and representation. One key approach is improving digital literacy, which empowers individuals from diverse backgrounds to effectively engage with technology and online platforms. Another strategy is ensuring affordable access to technology, such as affordable internet and devices for marginalized communities. Inclusive design of online platforms, ensuring they are accessible to individuals with disabilities, is another critical strategy. Furthermore, policy frameworks that encourage inclusivity, such as supporting equal internet access, play a significant role in bridging the digital divide (Loader, B. D., 2017, 3-9).

Italy:

Italy represents a difficult case for successful digital transformation in the public sector for three reasons. First, according to European reports (European Commission, 2021), the country can be considered a laggard in the digital innovation process. Secondly, Italy has always faced a high level of government instability at the national level and decades of incoherent devolution processes geared toward regions and local governments, which the literature associates with a poor policy capacity in several sectors (Capano, 2018). Third, since the 1990s, national governments have constantly created agencies to design and implement digitalization in the public sector (see Di Giulio & Vecchi, 2019; Mele, 2008). However, all these attempts have brought about few results. However, the initiatives designed and implemented by a new task force created by the national government in 2016 has thus far produced a relevant impact, which consists of the implementation of some of the digital platforms imagined in the past years. Thus, this outcome represents the notion of "partial success"

In the area of digital skills (Digital Economy and Society Index – DESI - 2021), the data see Italy in the last positions in Europe: this is reflected in the low use of online services and active participation in decision-making processes. Moreover, Eurostat survey (2021) highlights that

the level of digital skills of Italian citizens is dangerously low: 42% of citizens do not reach basic skills and more than one million of Italians (3%) have no digital skills at all. In the 4th National Action Plan 2019-2021 (4NAP), it is declared that "a scarce awareness of the rights of digital citizenship adds up to the scarce diffusion of digital competences". In addition, "the technologic evolution and the subsequent request of digital services by the citizen and the firms brings the necessity of promoting tools and initiatives for the diffusion of digital citizenship. In this sense, it is important to activate tools and initiative to spread awareness among citizens with respect to the right of digital citizenship and of value added provided by digital services, in particular:

- Supporting the implementation of digital citizenship protecting his right/duty: right of the digital citizen to use the services; duty of the citizen to comply with the procedures required by e-Government to exercise the right.

- Enabling the civil society to understand and know projects of digital transformation of the administrations involving young people in digital projects and promoting a platform dedicated to them".

The 4NAP shows different commitments and activities to guarantee the digital citizenship and digital competences: some Italian administrations have arranged interventions to ensure a widespread digital citizenship by offering training but also revising regulations, processes and tools in use. Consultations, participative processes, re-designing of websites and online services and platforms to allows for the participation and interaction between PA and citizens are some of the implemented interventions: however, still too few local public administrations are engaged directly and actively. In general, it is necessary to direct an extensive activity of awareness towards political decision-makers and administrative officers in order to enhance the consciousness and the interventions aimed at ensuring digital citizenship in the whole country. These objectives should include timing and measurability to guarantee another essential principle of OGP: the accountability. All these considerations have been incorporated during the process of co-creation of the 5th National Action Plan 2022-2023, which has included an action dedicated to inclusive digital innovation with the aim of planning projects related to the digital support and facilitation, moderating at the same time the technologic and socio-economic-cultural gaps.

Research show that accelerating the digital transformation of public services has positive effects on citizens' perception of the activity of the Public Administration (PA). Recent research shed light on relevant issues and challenges to be implemented in terms of: 1) new technologies integration and digital strategies needs; 2) hybridization of digital services and citizens' assistance; 3) the pivotal role of age: elderly citizens and socialization; and 4) digital skills gap reduction strategies (Sestino et al., 2021). At the Italian level, digitization strongly impacted the public sector (Datta, 2020; Tangi et al., 2021). Specifically, the results derived from the implementation of the provisions introduced by the Agency for Digital Italy (AgID), which coordinates various projects – including electronic invoicing, open data, digital security, Electronic Health Record, citizens digital identity, payments, etc. (Datta, 2020). The results of these projects vary according to the complexity and longevity of the implementation path. Some actions currently record positive results. For instance, more than 30% (20 millions) of Italians has a digital identity (SPID) through which using one unique set of credentials to access the online services offered by all Italian PAs. Moreover, 18,147 out of roughly 23,000 Italian PAs

adhered to PagoPA, a digital platform allowing citizens to pay for all public services in a more natural, fast, and modern way, relieving administrations from the costs and delays of handling digital payment methods. Finally, 11 million of Italians downloaded and regularly use IO, a mobile app designed to comply with the provisions of the simplification Decree (Legislative Decree 76/2020), which allow simple, safe, and fast communication between the citizen and the various PAs. The mobile app offers several advantages and simplifications to the citizen such as: information, communications, notices near deadlines and reminders; possibility to obtain self-certifications, documents and submits applications or declarations with a simple click.

In Italy, Law 92/2019, with art. 5, introduces the teaching of Digital Citizenship within the subject Civic Education. Educational institutions are not fully aligned to this new challenge while young people need to learn a conscious and responsible way to express their digital citizenship. In this context, the role of teachers is of crucial importance. They are called upon not only to be digitally aware citizens, but also to teach this competence to their students, with leadership that helps their students learn a responsible and active way of living the digital transformation. For this purpose, it is necessary, first of all, to fill any existing gaps in the digital skills needed to fully play thei rrole as Digital Citizens and, subsequently, to teach this to their students with innovative and effective teaching methodologies in the digital transition. In order to define the concept of Digital Citizenship, the Ministry of Education guidelines (Ministerial Decree n.35 -22 June 2020), identifies it as the ability to consciously and responsibly use digital tools. The competences needed for teachers to fully perform their task, according to Digicomp EDU (2017) cover six specific areas:

- professional engagement: using digital technologies to communicate, to collaborate with colleagues and for personal development.

- digital resources: identifying, creating and sharing digital resources.

- teaching and learning: managing the use of digital technologies for teaching and learning.

- assessment: reinforcing assessments through digital technologies; empowering learners: using digital technologies for inclusion and active involvement of learners.

- facilitating learners' digital competence: facilitating learners' creative and responsible use of digital technologies for sharing, communication, content-creation and problem-solving activities.

In order to stimulate and develop in students a model of Digital Citizenship that encompasses these aspects, it is necessary for teachers to use a circular learning process, aimed at providing not only knowledge but also skills for everyday life. In such a process the main steps are: •awareness: developing awareness about the need to be digitally literate, understanding needs and distinguishing between appropriate and inappropriate behaviours.

- guided practice: experimenting with digital technologies under adult guidance in protected environments, where there is the possibility of taking risks and making mistakes.

- modelling & demonstration: observing appropriate digital citizenship behaviours from adults.

- feedback & analysis: discussing and sharing reflections and experiences with peers and adults, receiving constructive feedback on the effective use of technologies in school and society.

The educational programme developed and delivered by our research team aims to develop the Digicomp EDU skills in teachers and to provide a teaching model for them to follow with their pupils to enhance learning, experimenting it directly during the proposed training course.

Turkey:

In Türkiye, digital inclusion and diversity are promoted through a combination of government policies, NGO initiatives, and private sector programs aimed at bridging the digital divide. The Ministry of Transport and Infrastructure runs projects like "e-Turkiye" (e-Government Portal) to make essential public services accessible online for all citizens, including disabled individuals and low-income groups. NGOs such as Habitat Association work to increase digital literacy and tech skills for women, youth, and marginalized communities (Habitat, 2022). Additionally, coding and STEM education programs targeting rural areas help reduce inequalities in digital participation.

Portugal:

General:

• A lot of individuals that live in rural and low-income areas are at risk of isolation. Providing affordable services is a positive strategy to tackle the isolation felt, as well as increase the levels of happiness and digital literacy on the population (Nguyen, 2021; Awad, 2023).

• Another important strategy is to invest in individuals' digital literacy, providing them with the tools to use technology safely and to find, evaluate, utilise and create information effectively and efficiently (Nguyen, 2021).

• Creating a culture of inclusion, by maintaining safe and welcoming online environments (Awad, 2023).

Portugal:

• In Portugal, programmes like Inclusão and Diversidade (Inclusion and Diversity) aim to promote gender equality, give women access to training in digital competencies, as well as develop competencies to allow women to follow a digital career.

Luxembourg: No Answer

2. How are issues of representation and accessibility addressed in digital platforms?

France:

Digital platforms address issues of representation and accessibility through various strategies aimed at inclusivity. For individuals with visual impairments, ensuring compatibility with assistive technologies (AT) is crucial. This includes providing alternative text for images, ensuring keyboard navigability, and avoiding reliance solely on visual elements. Additionally, offering customization options, such as adjustable text sizes and color contrasts, allows users to tailor the platform to their specific needs. Involving people with disabilities in the design and

evaluation processes further enhances the accessibility and relevance of digital platforms Kiambati et al., 2024, p. 25)

Italy:

Scholarly interest in the relationship between disability and the internet have recently morphed into a lively field of studies. One trend in digital disability studies has been to focus on digital affordances. People with disabilities are more likely than non-disabled people to be socially isolated and lack opportunities to interact with other people with disabilities in person due to distance and a myriad other barriers. This is not limited to online communities that are disability- or illness-specific (Vicari & Cappai, 2016). Instead, it is also relevant for virtual groups based on the intersection of disability with other identities such as gender and disability, race and disability, sexual orientation and disability, and so on (Cole & Nolan, 2019). This has fuelled the rise of new instances of digital disability activism, sometimes in open opposition to the work of established advocacy organizations, which were criticized by a new generation of tech savvy self-advocates for not expressing a disabled leadership and adequately representing grassroots grievances (Parsloe & Holton, 2018). Other work has also shown that, on a personal level, uses of social media and other platforms can boost interpersonal support and participation in social and political life, although at the same time they can also generate psychological stress and anxiety where online context collapse requires users to navigate difficult relationship issues (Trevisan, 2020). Within these broader international trends, Italy constitutes an interesting case study. While other Western countries such as the U.S. and the UK pioneered anti-disability discrimination legislation aimed at promoting the inclusion of people with disabilities in all aspects of social life in the early 1990s with the Americans with Disabilities Act (1990) and the Disability Discrimination Act (1995) respectively, Italy waited until 2006 to pass similar legislation. Recent studies have revealed alarming rates of social isolation for Italian people with disabilities with nearly 22% living in complete isolation and almost 25% whose few relationships are exclusively based on help and support (ISTAT, 2019: 95-96). These, however, are only rough measures of sociality and do not account specifically for online relationships and interaction. The more recent and less established nature of legislative frameworks and protections for Italian disabled people, in addition to cultural differences, makes it an interesting country for studying digital disability spaces. Moreover, in Italy people with disabilities have built significant online resources through websites such as “Disabili.com” and social media pages such as “Ability Channel.” Yet, an in-depth investigation of these spaces is still missing (Melchior & Trevisan, 2022).

Turkey:

Representation and accessibility remain challenges in Türkiye, but progress has been made through policy regulations and corporate practices. The Law on the Rights of Disabled People requires websites and digital services to be accessible, leading platforms like e-Devlet and some banking applications to incorporate screen reader compatibility, voice command features, and simplified interfaces (BTK, 2022). However, issues remain in media and online content representation, where gender bias, ethnic stereotypes, and limited representation of disabled individuals persist in digital platforms and advertisements.

Portugal:

Portugal:

• In the words of Botelho (2021) accessibility considerations 'should be part of the research and development process of all entities developing new digital tools and environments. In this sense, some digital platforms in Portugal try to mitigate these issues by utilising/relying on:

- Web Content Accessibility Guidelines (WCAG): The main goal of these guidelines is to ensure that everyone, including individuals with disabilities (either visual, motor, auditory or cognitive) have access to digital content that is accessible to them. These guidelines are followed by some public and some private institutions in Portugal (World Wide Web Consortium, s.d.).
- Portugal Digital Agency (Agência para a Transformação Digital – ATD): The Portuguese government has been actively working on digital transformation, focusing on making public services more accessible online.
- Portuguese Observatory for Web Accessibility (Observatório Português de Acessibilidade Web): This observatory aims to assess, identify and promote de application of the best practices of accessibility in using websites and mobile APPs and it also helps to raise-awareness, support education and alert the decision-makers on the public sector, contributing to more inclusive digital services (Observátorio Português da Acessibilidade Web, n.d).

Luxembourg: No Answer

3. What steps are taken to ensure equal access to digital resources and services for all users?

France:

Ensuring equal access to digital resources and services involves implementing several key strategies:

Providing Affordable Services and Products: Offering low-cost or subsidized internet services and digital devices helps bridge the digital divide, particularly for low-income individuals. Programs like the Federal Communications Commission's Lifeline Program in the U.S. provide discounts on internet services to eligible households.

Promoting Digital Literacy: Enhancing digital skills through training programs enables individuals to effectively use digital technologies. Libraries and community centers often offer workshops to improve digital literacy, empowering users to navigate online platforms confidently.

Creating Accessible Content: Designing digital platforms with accessibility in mind—such as incorporating closed captions, alt-text, and audio descriptions—ensures that all users, including those with disabilities, can access and benefit from digital resources.

Establishing a Culture of Inclusion: Fostering inclusive digital environments by moderating online spaces to promote diversity and prevent discrimination helps create welcoming spaces for all users.

Implementing Public Policies: Governments can enact policies that support digital inclusion, such as investing in infrastructure to provide reliable internet access in underserved areas and offering financial assistance for technology access.

By adopting these strategies, societies can work towards ensuring equitable access to digital resources and services for all users (Anderson, O., 2019, p.1)

Italy:

Digitization produces, among other effects, the phenomenon of the digital divide, excluding certain categories of subjects from the use of digital public services. For example, the elderly and, more generally, all those individuals who do not possess the knowledge suitable for the correct use of technological artifacts; individuals who do not have the ability to have an internet network or a device that allows access to the digital world. To prevent this phenomenon from persisting with the risk of frustrating the efforts pursued by the PA in the digital field, a successful strategy could consist in the maximum simplification of applications and websites that allow accessing and using digital public services. There is also a need to consult more intuitive and easier platforms, reducing the dispersiveness of digital interfaces and inserting practical guides allowing a clear display of the features built into the platform. To this end, it could be useful to implement additional forms of assistance by public officials to users in difficulty, or, to assist those with minimal digital skills, to prepare online tools that allow them to achieve the same goal without the physical operator help, such as simplified guides, information pages, information in the app, and so on. Indeed, by leveraging on the concept of socialization strategies (Guido et al., 2019) whereby young family members' assistance can improve both the cognitive (elderly's knowledge and interest) and behavioral (elderly's purchase intentions and technology usage) dimensions, thus facilitating the acquisition of complex skills that cannot be assimilated independently, several initiatives have been projected.

Turkey:

To promote equal access, Türkiye has implemented free internet access points in public areas, including municipal libraries, universities, and digital literacy centers. Additionally, the "Internet for All" initiative by the BTK (Information and Communication Technologies Authority) provides affordable broadband packages for low-income families. Education-focused programs, such as the FATİH Project, aim to bridge the gap in schools by providing tablets and digital infrastructure for students in underprivileged regions (MEB, 2023). However, rural and low-income communities still face challenges in stable internet connectivity and affordability.

Portugal:

- In Portugal, there are some initiatives/strategies destined to provide equal access to digital resources to all. We highlight the National Strategy for Connectivity in Very High-Capacity Electronic Communication Networks (2023-2030). With this initiative, the Portuguese Government aims to guarantee access to very high-capacity networks for the entire population, to ensure coverage of the entire national territory, guaranteeing coverage of all households by a Gigabit network until 2030 (Diário da República, 2022).

4. Can you provide examples of initiatives or policies that have successfully fostered diversity and inclusion in digital contexts?

France:

A successful initiative that fosters diversity and inclusion in digital contexts is highlighted in the study by (Triplett,2023, p. 37-41). The research explores the role of technology in promoting diversity and inclusion, emphasizing how digital tools help connect individuals and create more inclusive environments. The study demonstrates that technology, when leveraged properly, can bridge gaps related to economic, social, and cultural backgrounds, enabling organizations to develop diverse and inclusive teams. The findings suggest that promoting complete access to digital technology is essential for fostering a more equitable and inclusive digital landscape.

Italy:

Digital inclusion is a challenge to inequalities and allows to guarantee to the citizen all the tools (digital skills and technologies) necessary to participate in society, communicate, work and use the digital services of the Public Administration. To this purpose, the Open Government Partnership promotes the development of public platforms for service delivery, participation, increase of digital citizenship and sharing of ideas, projects and information. As planned in the project "Competenze digitali per la PA" by the Public Administration Department, it is necessary to consolidate digital competences for public servants to boost propensity to change and to innovate in public administration. The process innovation has the aim to review procedures, norms and regulations to guarantee simplification of processes also through the adoption of more adequate technologies up-to-date. Public administration has to guarantee to everyone the possibility to use digital services, for this reason, from the stage of designing, it has to consider accessibility of platforms where services, contents and documents, activities of interaction between the administration and citizens are provided. The project addresses accessibility for informatic applications regardless of the device used (PC, smartphone, tablet, digital TV, Totem, etc.). Law n.4 of 2004 (Provisions to favour and simplify access to users and, in particular, to users with disabilities to informatic tools) and the Legislative Decree n.106 of 2018 (Implementation od Directive UE 2016/2102 related to the accessibility of websites and mobile applications of public entities) are the normative baselines to whom "Guidelines on the informatic tools accessibility" by the Italian Agency for Digital (AgID) refers, supporting public administrations by defining technical requirements to guarantee accessibility. As written by the AgID, "the usability measures the degree of ease and satisfaction with whom users interact through the website or app dashboard; it will end up being the more usable, the more the baseline analysis of the design are closer to the expectations of the user that will interact with the system". The issue is "citizen at the heart" even in the design of services in order to guarantee usability, accessibility and easy consumption of the service to every type of user.

The Italian Strategy for Digital Skills is aimed to fight the cultural digital divide affecting the Italian population through supporting digital inclusion, to support the e-skills development throughout the higher education and training cycle, to promote key competences development for the future and increase the percentage of ICT specialists (especially in emerging technologies). Another objective is to ensure that the entire working population has basic digital skills for the new needs and ways of working. This Strategy has been drafted jointly with the collaboration of Ministries, Regions, Provinces, municipalities, universities, research institutes, companies, professionals, the National Public Broadcasting, associations and the various public

sector organizations, the organizations belonging to the National Coalition, and with the informal exchanges with the European Commission, under the direction of the Technical Steering Committee of the 'Repubblica Digitale' initiative and the coordination of the Department for Digital Transformation. It develops some of the objectives of the 2026 Digital Italy strategy of the Minister for innovation and digital transition. The strategy identifies four lines of intervention, coherently with the four pillars of the European Coalition for Digital Skills and Jobs:

- Higher Education and Training - for the development of e-skills for young people within the mandatory education cycles (working group is coordinated by the Ministry of Education and the Ministry of University and Research)

- Active workforce - to ensure adequate e-skills in both the private and public sectors, including e-leadership skills (working group is coordinated by the Ministry of Economic Development and Department for Public Administration)

- ICT specialist skills - to enhance the country's ability to develop skills for new markets and new jobs, with a specific focus on emerging technologies and key competencies for future jobs (working group is coordinated by the Ministry of University and Research and the Ministry of Economic Development)

- Citizens - to develop the digital skills needed to exercise citizenship rights and promote active participation in the democratic life (working group is coordinated by the Department for Digital Transformation).

For each line of intervention are associated priorities and lines of action. The Strategy has then been complemented by an Operational Plan including targets for 2025, based mainly on both DESI and Eurostat indicators like the following:

- Equip 70% of the population with at least basic digital skills and bridge the gender skills gap in the ICT sector.

- To double the rate of Italian citizens with advanced digital skills (78% of young people with higher education, 40% of workers in the private sector and 50% of civil servants).

- Increase the number of graduates in ICT three times; and four times for the number of female graduates and increasing twofold the share of companies active in the field of big data.

- Double the share of companies that uses big data; and double the employment of digital experts and ICT specialists in small and medium sized enterprises (SMEs).

- Increase five times in the share of the population using public digital services (64%).

- Increase in the use of the Internet and the ability to use computers also for older people: 84% in the 65-74 age group.

The Plan addresses the 41 lines of action identified in the Strategy through 111 actions and provides a dashboard of over 60 indicators to monitor the impact on the 4 lines of intervention. Each action also includes appropriate milestones, result indicators and target values. The dashboard is based on the indexes included in the Digital Economy and Society Index (DESI) of the European Commission and the Digital Maturity Indexes (DMI) elaborated by the Digital Agenda Observatory. The monitoring model of the Plan allows to update the operational plan

on a six-monthly basis, evaluating the general impact of the whole plan and the effectiveness of the actions and improving them where necessary, and considering the ever-changing context according to the Plan-Do-Check-Act (PDCA) continuous improvement model.

On the basis of the National Strategy for digital skills and the related Operational Plan, a series of actions are previsedin the fourth axis aimed to 1) provide digital and IT contents in those formal training courses for adults offered by educational institutions; 2) introduce training courses within the non-formal educational circuit, based on the enhancement of lifelong learning, e.g., by online learning platforms; 3) digital facilitation path to support and facilitate the access to public digital services; 4) communication paths, increasing individuals' awareness toward the relevance digital skills and competences with a systematic support deriving from the mass media (e.g., television, radio); and 5) inclusion initiatives aimed to those disadvantage social groups (e.g., elderly, immigrants). Importantly, there are also numerous digital skills initiatives supported by different actors from small third sector companies to large players in the deep-tech sector (e.g., as for Amazon, Fastweb). These digital skills initiatives are part of the National Coalition instituted by Repubblica Digitale. The aim of the National Coalition is to systematize and centralize the existing and virtuous initiatives, fostering collaboration and creating new synergies. Thus, by creating a synergistic network of active stakeholders, those successful initiatives even at local level may be shared and know by a wider audience in the attempt to provide more digital skills improvement stimuli for citizens.

Turkey:

Several initiatives in Türkiye have successfully promoted diversity and digital inclusion:

Habitat Association's Digital Literacy & Coding Programs: Empowers women, youth, and refugees by teaching coding, cybersecurity, and digital entrepreneurship skills (Habitat, 2022).

Engelsiz e-Devlet (Barrier-Free e-Government): Offers accessible online government services for disabled users, including voice-controlled navigation and sign language assistance (BTK, 2022).

FATİH Project (Education Technologies Enhancement): Provides interactive smartboards, free tablets, and digital learning platforms to reduce educational inequalities in rural schools (MEB, 2023).

Women in Tech Türkiye: A corporate-led initiative that mentors and supports women in digital entrepreneurship and STEM fields.

Portugal:

- In Europe:
    - European Commission Digital Education Action Plan (2021-2027): This initiative aims to improve digital education in Europe, by supporting the adaptation of the education and training systems of member states to the digital age.
- In Portugal:
    - InCode2030 Programme (2018-2030): One of the objectives of this national programme is to include the promotion of digital inclusion and literacy for all

and to foster digital education and awareness of technology from a very young age (Jākobsone, 2021).

Luxembourg: No Answer

c) Ethical Governance of Technologies

1. What ethical frameworks or guidelines are employed to guide the development and deployment of digital technologies?

France:

The development and deployment of digital technologies are guided by various ethical frameworks and guidelines to ensure responsible and beneficial outcomes. One prominent approach is Value Sensitive Design (VSD), which integrates human values into the design process through a structured methodology. VSD involves three key investigations:

1. Conceptual Investigation: Identifying and analyzing the values at stake.

2. Empirical Investigation: Understanding how stakeholders perceive and are affected by these values.

3. Technical Investigation: Examining how technological features can support or hinder these values.

This comprehensive approach ensures that ethical considerations are embedded throughout the technology's lifecycle.

Another significant framework is the Ethical Guidelines for Trustworthy AI established by the European Commission. These guidelines emphasize seven key requirements:

1. Human Agency and Oversight: Ensuring AI systems support human autonomy and decision-making.

2. Technical Robustness and Safety: Guaranteeing AI systems are secure and reliable.

3. Privacy and Data Governance: Protecting personal data and ensuring privacy.

4. Transparency: Providing clear information about AI systems' operations.

5. Diversity, Non-Discrimination, and Fairness: Promoting inclusivity and preventing bias.

6. Societal and Environmental Well-Being: Considering the broader societal and environmental impacts.

7. Accountability: Establishing mechanisms for accountability in AI systems.

These guidelines aim to foster trust and ensure that AI technologies are developed and used responsibly.

dl.acm.org

Additionally, the OECD Principles on Artificial Intelligence provide a set of recommendations for governments and stakeholders to promote the responsible development and use of AI. These principles include:

1. Inclusive Growth, Sustainable Development, and Well-Being: Ensuring AI contributes to economic growth and societal well-being.

2. Human-Centered Values and Fairness: Upholding human rights and promoting fairness.

3. Transparency and Explainability: Making AI systems understandable and transparent.

4. Robustness, Security, and Safety: Ensuring AI systems are secure and operate as intended.

5. Accountability: Holding stakeholders accountable for AI systems' outcomes.

These principles serve as a foundation for creating policies and practices that guide the ethical development and deployment of AI technologies.

Ethical frameworks like Value Sensitive Design, the European Commission's Ethical Guidelines for Trustworthy AI, and the OECD Principles on Artificial Intelligence provide structured approaches to integrating ethical considerations into the development and deployment of digital technologies. These frameworks aim to ensure that technological advancements align with societal values and contribute positively to human well-being.

Italy:

The increasing integration of digital technologies and Artificial Intelligence (AI) into various fields of human and social life calls for a structured ethical framework to ensure fairness, inclusivity, and human well-being. Ethical digital governance must address risks such as digital divide, discrimination, loss of autonomy, and the need for structured education on digital and AI literacy. Among the ethical frameworks, here we refer to the AI Act (2023), the White Paper on Trustworthy AI (2020), the United Nations' Sustainable Development Goals (SDGs), guidelines from the World Health Organization (WHO), the insights from "AI for Teachers" (Pressbooks) edited by Colin de la Higuera, and the ethical considerations outlined in the Onlife Manifesto by Floridi et al.

The AI Act highlights the need for AI systems to be developed in a way that benefits all individuals, regardless of socio-economic status. This model has also been dsicussed in The Principle of Bomedical Ethics by Beauchampp and Childress (1979) and readapted by Luciano Floridi in his Ethics of Artificial Intelligence (2022).

The SDGs, particularly Goal 9 (Industry, Innovation, and Infrastructure) and Goal 10 (Reduced Inequalities), emphasize bridging the digital divide by ensuring that technological advancements are inclusive and do not exacerbate disparities. Additionally, the WHO recognizes digital health equity as a crucial component of global well-being, advocating for universal access to health-related digital services.

The Onlife Manifesto underscores the transformation of human experience in the digital age, calling for policies that ensure technology remains a tool for inclusivity rather than

exclusion. This should be stressed as the assumption of the Manifesto refers to an anthropological change in the digital and AI Era.

Following a risk-based and human-centred approach, the AI Act mandates that AI systems be designed to prevent bias and discrimination, especially against marginalized communities, including women and minorities. This happens in many ways in the digital era, with specific reference to AI systems, such as Generative AI and Large Language Models. The AI Act emphasizes that AI should be used to support, rather than replace, human decision-making in critical areas such as healthcare, justice, and employment. At the same time, the WHO underscores the importance of mental and physical health, advocating for responsible technology use that enhances, rather than diminishes, quality of life. Additionally, SDG 3 (Good Health and Well-being) aligns with these principles, urging the development of AI-driven health solutions that respect patient rights and privacy.

Similarly, the document which precedes the AI Act, White Paper on Trustworthy AI, stresses fairness and transparency, requiring AI systems to be audited for algorithmic bias. These concerns align with SDG 5 (Gender Equality), which calls for eliminating discrimination against women in technological advancements.

Ethical use of new technologies must not compromise human autonomy or well-being.

For ethical digital transformation, societies must prioritize education on AI and digital technologies. The AI Act calls for greater transparency and accountability in AI systems, including user education to ensure informed interactions with technology and promoting the idea the working on competences helps people in managing new technologies.

Turkey: No Answer

Portugal:

General:

- General Data Protection Regulation (GDPR)
- Organisation for Economic Co-operation and Development (OECD)
- The Ethics Guidelines for Trustworthy AI
- The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

Luxembourg: No Answer

2. How are potential biases or unintended consequences addressed in algorithmic decision-making processes?

France:

To address potential biases or unintended consequences in algorithmic decision-making processes, researchers have suggested various strategies, including the implementation of fairness-aware algorithms, transparency in data collection, and ongoing monitoring and evaluation of algorithmic outputs. Biases can arise at different stages of the process, such as data collection, model development, and deployment. Therefore, addressing these biases

requires both technical and ethical considerations. For instance, fairness interventions, such as re-weighting training data or adjusting the decision thresholds, can help to mitigate algorithmic bias. Additionally, enhancing transparency through explainability and auditability allows stakeholders to better understand how decisions are made and to identify any potential sources of bias or unintended outcomes.

A study by (Binns et al., 2018, p. 1-14) discusses these issues in depth, emphasizing the need for addressing algorithmic bias and providing strategies to ensure more equitable and ethical decision-making.

Italy:

Algorithmic decision-making systems are increasingly integrated into fields such as healthcare, employment, education, and law enforcement; and, obviously, in individual life through social media. While AI offers efficiency and scalability, it also presents risks of bias and unintended consequences. Ethical AI governance, as outlined in the AI Act, the White Paper on Trustworthy AI, the UN Sustainable Development Goals (SDGs), WHO guidelines, "AI for Teachers," and the Onlife Manifesto, mandates proactive strategies to mitigate these risks.

Bias and unintended consequences can be addressed by emphasizing structured education, accountability, and interdisciplinary collaboration.The AI Act imposes strict obligations on high-risk AI systems; it mandates thorough testing and risk assessments to identify and mitigate biases before deployment.

The White Paper on Trustworthy AI highlights the necessity of explainability and oversight to ensure that AI systems do not reinforce existing inequalities. So, the SDG 10 (Reduced Inequalities) reinforces these principles stressing on inclusive technology that does not exacerbate social disparities. WHO guidelines similarly stress the ethical deployment of AI in healthcare, ensuring that decisions prioritize human well-being over algorithmic efficiency.

A key method to address algorithmic bias is through structured and interdisciplinary education. Indeed, both the AI Act and the White Paper emphasize digital literacy and responsible AI training as crucial for both developers and users. For instante, the "AI for Teachers" supports this by advocating for AI education in academic curricula to ensure that students and professionals understand algorithmic decision-making and its ethical implications.

The White Paper on Trustworthy AI suggests independent audits and regulatory oversight to ensure compliance with ethical guidelines. Additionally, interdisciplinary review boards and teams including ethicists, sociologists, bioethicists, philosophers, and legal experts—should be established to assess the social impact of AI systems and bias mitigation: in order to prevent underrepresentation or overrepresentation of certain groups. Bias in AI can take multiple forms, including the exclusion of minority groups from datasets or the disproportionate emphasis on certain demographics. SDG 16 (Peace, Justice, and Strong Institutions) calls for equitable AI systems that promote justice rather than perpetuate disparities. The AI Act mandates rigorous testing to detect and correct skewed datasets. For this reason, interdisciplinary teams – different competences - play a crucial role in identifying and rectifying these biases during the design phase. Additionally, continuous monitoring and feedback loops should be implemented to detect and correct biases post-deployment.

Turkey: No Answer

Portugal: No Answer

Luxembourg: No Answer

3. What measures are taken to ensure that digital technologies support social good and do not contribute to harm?

France:

Protection Regulation (GDPR) in the European Union, which aims to protect individual privacy and ensure that personal data is used responsibly. Another key measure is promoting fairness and transparency in AI systems, which can help mitigate risks of bias and discrimination. Many scholars emphasize the importance of ethical AI and responsible data use, and suggest that technologies should be designed to prioritize human rights, equity, and inclusivity.

One example is the development of frameworks that address AI ethics, such as the guidelines proposed by the European Commission, which advocate for transparency, accountability, and data protection (European Commission, 2019). Another study discusses how integrating ethical considerations in the design of machine learning models can reduce potential harm and promote fairness (Mehrabi et al., 2021, p. 1-35). Additionally, the creation of policies aimed at preventing misinformation and ensuring the responsible use of technology has become critical in mitigating harm to societies (Vosoughi et al., 2018, p.11).

Italy:

As digital technologies become increasingly embedded in daily life, ensuring they contribute positively to society while minimizing harm is a priority. As said above, Ethical frameworks such as the AI Act, the White Paper on Trustworthy AI, the UN Sustainable Development Goals (SDGs), WHO guidelines, "AI for Teachers," and the Onlife Manifesto provide an approach oriented to the promotion of social good.

The AI Act establishes a comprehensive regulatory structure, categorizing AI systems based on risk levels and imposing strict requirements on high-risk applications for all the industries that produce AI systems. The principles of transparency, accountability, and human oversight are necessary to prevent misuse, and to promote a good cooperation between humans and machines.

In the SDGs, we find Goal 10 (Reduced Inequalities) and Goal 16 (Peace, Justice, and Strong Institutions) – they advocate for the responsible deployment of technology that promotes inclusivity and safeguards fundamental rights. The Onlife Manifesto urges a rethinking of governance structures in the digital transition to protect human agency and prevent the erosion of public trust in technological systems. This is important for the future of education and work that could be human-centred, avoiding inerruption of tasks in human-computer interaction.

One of the most important practice in the digital age, in order to preserve the social good, is accountability. For instance, the AI Act mandates thorough documentation and impact assessments for AI systems, allowing stakeholders to challenge algorithmic decisions.

As argued in point (2), the interdisciplinary review boards play a crucial role in identifying potential harms and ensuring that technology aligns with ethical principles, iht specific stages

where AI is founded and programmed. This aligns with SDG 16, which promotes strong institutions and justice by advocating for mechanisms that prevent technological abuse and reinforce accountability.

Again, the role of education is crucial: The AI Act and the White Paper emphasize the need for widespread AI literacy to empower individuals to critically engage with digital systems. This means that the problem about AI misuse is not only linked to founders and programmers, but also to users. This is the policy supported by the Onlife Manifesto about digital literacy and lateral thinking. Education must extend beyond technical expertise to include ethical, philosophical, and sociological dimensions, ensuring a holistic understanding of AI's impact on society. For this reason, the interdisciplinary approach is every day more necessary to be applied in use of digital tools. For instance, the Onlife Manifesto argues that breaking disciplinary silos is crucial for fostering ethical digital innovation.

Turkey: No Answer

Portugal:

- Accountability and Legal Frameworks, such as GDPR regulations.
- In Portugal, the National Digital Strategy (Estratégia Digital Nacional) helps to simplify the digital interaction between citizens, companies and the state, ensuring that their fundamental rights are protected and the universal accessibility of digital technology for all.

Luxembourg: No Answer

4. What are the key considerations for ensuring the ethical governance of digital public spaces?

France:

To ensure the ethical governance of digital public spaces, several key considerations need to be considered:

Transparency: It is essential to establish clear rules and policies about how digital platforms collect, use, and share data. This transparency fosters trust and ensures users understand their rights within these spaces (Van Dijck et al., 2018, p.18).

Accountability: Digital platforms should be accountable for the content and actions occurring within their spaces. This includes establishing mechanisms for addressing harmful behavior, such as hate speech or misinformation, while protecting freedom of expression (Gillespie, 2018).

Privacy: Protecting user privacy is critical in ensuring that personal data is not exploited or misused. This involves implementing strong data protection measures, such as encryption, and minimizing the collection of sensitive information (Andrejevic, 2013, p. 18-40).

Inclusivity and Accessibility: Ensuring that digital spaces are inclusive and accessible to diverse groups, regardless of their socio-economic status, cultural background, or ability, is another important ethical consideration (Gillespie, 2018, p.21).

Fairness and Non-Discrimination: Algorithms and automated systems used in digital spaces must be designed to avoid discrimination or bias, ensuring that all users are treated regardless of their background (O'Neil, 2016, p. 17-21).

Italy:

The ethical governance of digital public spaces requires a balance between innovation, privacy protection, and social well-being. As outlined in the AI Act, the White Paper on Trustworthy AI, the SDGs, WHO guidelines, and the Onlife Manifesto, safeguarding democratic values while mitigating privacy risks and discrimination is paramount.

Firstly, some key considerations about the AI Act will be provided. The AI Act introduces a risk-based classification of AI systems, categorizing them into minimal, limited, high, and unacceptable risk levels. High-risk AI applications in digital public spaces, such as facial recognition and content moderation algorithms, must adhere to strict transparency and accountability measures. Aligned with this classification, some key practices could be:

- Impact Assessments: Mandatory assessments evaluating the societal and privacy risks of AI deployments.
- Regulatory Oversight: Independent auditing bodies to monitor AI compliance with legal and ethical standards.
- Explainability and Transparency: Ensuring algorithmic decision-making is understandable and challengeable by users.

About the the White Paper on Trustworthy AI which are guidelines, not a law as the AI Act, some key-good practices can be derived. Necessary to be rimended: the White Paper is linked to the GDPR principles:

- Data Minimization: Limiting data collection to what is strictly necessary for intended purposes.
- User Consent Mechanisms: Empowering individuals to control how their data is used.
- Secure Data Handling: Implementing encryption, anonymization, and secure data storage to prevent breaches.

About The Onlife Manifesto to redefine privacy in the digital age, it urges the adoption of decentralized data governance models to enhance user autonomy. To ensure digital public spaces are free from bias and discrimination, a multi-faceted approach must be adopted:

- Diverse and Interdisciplinary Teams: Integrating sociologists, ethicists, legal experts, and AI researchers to design inclusive systems.
- Bias Audits and Fairness Checks: Regular evaluations to identify and rectify biases in algorithmic decision-making.
- Algorithmic Diversity: Avoiding over-reliance on homogeneous datasets that may reinforce societal inequalities.

About the SDG 10 (Reduced Inequalities), it underscores the importance of ensuring that digital technologies do not disproportionately disadvantage marginalized groups Ensuring ethical digital public spaces requires educating users about AI-driven processes and their implications. For this reason, the references wwere abou the AI Act and "AI for Teachers", which advocate for:

- Public Awareness Campaigns: Informing citizens about their digital rights and AI decision-making processes.
- Educational Initiatives: Integrating AI ethics into school curricula to build critical digital literacy skills.
- Open AI Governance Platforms: Encouraging participatory governance where users can provide feedback on AI policies and implementations.

In conclusion, ensuring the ethical governance of digital public spaces requires a comprehensive approach that integrates risk-based AI oversight, privacy safeguards, inclusivity measures, digital literacy promotion, and adaptive regulation. By adhering to the AI Act, the White Paper on Trustworthy AI, SDGs, WHO guidelines, and the Onlife Manifesto, policymakers can create digital environments that protect individual rights while fostering democratic values and societal well-being. Ethical AI governance is not just a regulatory necessity but a fundamental pillar of trust in the digital age.

Turkey: No Answer

Portugal:

General:

• Digital spaces, such as social media platforms and other types of digital technologies have become vital for public discourse and social interaction. That being said, it is important to make sure that they follow and abide by fundamental human rights, promote well-being and prevent harm. For this, and to ensure ethical governance, it's crucial to find a balance between freedom of speech and hate speech/misinformation and cyberbullying. According to Walker (2019), a way to reach this balance is through transparency. Online platforms must provide users with guidelines to inform them about the company's standards, as it informs the users about what should be posted/ shouldn't be posted, reducing the harmful content that needs to be moderated (Walker, 2019).

• As digital platforms collect important and personal data, they need to ensure that the data is protected (abiding by the GDPR). They also need to provide ways to anonymise data, to minimize risks of harm in case of data breaches.

Portugal:

• In Portugal, the Portuguese Letter of Human Rights in the Digital Era, is an important mechanism to ensure that the rights of human beings are protected, especially in the online world.

Luxembourg: No Answer

Synthesis and Discussion of the Results

Can you discuss any challenges or lessons learned in implementing ethical governance in the digital age?

France:

Balancing Free Speech and Regulation: A fundamental challenge in the realm of ethical governance is the need to strike a delicate balance between safeguarding free speech and curbing harmful content, such as hate speech and disinformation. Governments and digital platforms must craft regulations that protect individuals from harm, ensuring that speech is not used to incite violence, spread false information, or cause societal division, while still respecting the right to free expression. This balance becomes particularly complex due to the cultural and legal variability of definitions of hate speech and disinformation. Different nations, for example, may interpret what constitutes harmful content differently, which makes it difficult to create universally accepted laws that adequately protect individuals while upholding the principle of free speech. Furthermore, the dynamic nature of online platforms and the rapid spread of information complicate enforcement, as content regulation often requires real-time action. Governments and organizations must navigate these tensions carefully, ensuring that their regulatory frameworks are both effective in preventing harm and fair in preserving individual freedoms.

Algorithmic Transparency: Another significant issue in ethical governance concerns the use of algorithms for content moderation. Algorithms designed by digital platforms to filter and moderate content may unintentionally suppress a wide range of diverse viewpoints, resulting in a homogenized information environment that reflects only a narrow set of perspectives. This can lead to the marginalization of certain voices or the unintentional promotion of biased narratives. One crucial lesson in addressing this issue is the need for greater transparency in algorithmic decision-making processes, especially regarding how these algorithms determine content visibility. The opacity of algorithms where decisions about what content is shown or hidden are often shielded from public view can lead to a breakdown in public trust and accountability. As digital platforms grow increasingly influential in shaping public discourse, it is essential for tech companies to open up their algorithms to scrutiny and involve a wide range of stakeholders in shaping the policies governing content moderation. Transparency not only helps ensure fairness but also provides users with the information necessary to understand why certain content is being promoted or suppressed, thus fostering a more informed and engaged user base.

Data Protection and Privacy: The challenge of balancing data protection with service functionality is another central issue in ethical governance. Digital services increasingly rely on access to personal data to provide personalized experiences, such as tailored recommendations, location-based services, and customized content. However, the ethical management of this data is fraught with complications, especially in light of increasing concerns about privacy and the growing body of data protection laws worldwide. Platforms must ensure that they obtain informed consent from users and remain transparent about how personal data is collected, stored, and used. Moreover, they must respect user privacy while still delivering functional services that meet user expectations. Frameworks such as the General Data Protection Regulation (GDPR) in Europe aim to address these challenges by setting strict guidelines for the handling of personal data. However, the implementation of such regulations often presents challenges in terms of balancing the need for data to improve service functionality with the need to protect privacy. Striking this balance is an ongoing and evolving challenge, as companies must continually adapt to new legal standards and technological advancements while ensuring that user data is treated ethically and securely.

Inclusivity in Digital Spaces: Digital inclusion is a critical issue in ensuring that all individuals, particularly those from marginalized groups, have equitable access to the opportunities that the digital world offers. The lessons learned in this domain highlight the importance of making digital platforms accessible to all users, regardless of their socioeconomic background, disability, or geographic location. Ensuring digital literacy and making technology affordable are foundational steps toward achieving this goal. However, the path to digital inclusion remains challenging, as there are significant barriers to accessing digital resources, especially in underserved communities. Discrimination and bias in online spaces further exacerbate these challenges, creating environments where some individuals may feel excluded or unsafe. Addressing these issues requires a multifaceted approach, which includes expanding access to affordable technology, improving digital literacy programs, and fostering inclusive digital communities that actively work to counteract discrimination and marginalization.

Mitigating Harmful Content: The rapid spread of harmful content online, such as cyberbullying, hate speech, and graphic violence, has become one of the most pressing issues in digital governance. Despite efforts to combat these forms of online violence, responses to date have yielded limited success, particularly when it comes to curbing the spread of harmful content across social media platforms. One of the primary challenges is the sheer volume of content generated by users, which makes it difficult for platforms to effectively monitor and address harmful behavior in real-time. Many platforms rely on a combination of automated tools and human moderation to identify and remove harmful content, but these strategies often fall short. Automated tools may struggle to accurately assess the context of content, leading to false positives or the overlooking of subtle forms of harm. Human moderators, on the other hand, are limited in number and resources, which makes it difficult to keep up with the sheer scale of content being generated. Research calls for more rigorous and comprehensive studies on the effectiveness of current interventions, as well as for continuous refinement of detection tools and moderation practices. The challenge lies in developing more sophisticated and nuanced methods to identify and counter harmful content while still respecting free expression and ensuring that platforms remain spaces for open dialogue.

Ethical governance in the digital age involves a careful balancing act that requires policymakers, tech companies, and users alike to consider a wide range of complex issues. From free speech to algorithmic transparency, data privacy to digital inclusion, and the mitigation of harmful content, addressing these challenges requires a nuanced and collaborative approach to ensure that the digital landscape remains both functional and fair for all participants.

Italy: No Answer

Turkey:

Implementing ethical governance in the digital age presents several challenges, particularly in balancing privacy, transparency, and regulation while fostering innovation and inclusivity. In Türkiye, key difficulties include data protection enforcement, misinformation regulation, and ethical AI governance.

One of the main challenges is balancing privacy with national security and surveillance concerns. While Türkiye's Personal Data Protection Law (KVKK - Law No. 6698) aligns with GDPR principles, enforcement remains inconsistent. Government access to digital data under cybersecurity laws raises concerns about surveillance and data misuse (Yesil & Sözeri, 2018).

Lessons from this show the need for greater transparency and independent oversight to prevent data privacy violations.

Another critical issue is regulating misinformation and digital ethics without restricting freedom of expression. Recent social media laws in Türkiye mandate platforms to remove content flagged as disinformation, yet critics argue these laws can be used to silence opposition voices. A key lesson here is that misinformation policies should be transparent, evidence-based, and not weaponized for political purposes.

In the realm of AI governance, Türkiye faces challenges in algorithmic bias and ethical AI implementation. AI-driven decisions in finance, recruitment, and law enforcement risk reinforcing societal biases. As seen in global cases, biased AI training data can lead to unfair outcomes, demonstrating the need for diverse data sets, ethical AI auditing, and inclusive policymaking.

Finally, corporate digital ethics and accountability remain a concern, as many businesses in Türkiye lack structured ethical frameworks for handling consumer data and digital rights. International companies face pressure to comply with GDPR and KVKK, but smaller businesses often lack awareness and resources. Successful governance models emphasize stronger corporate responsibility, clear ethical guidelines, and stakeholder engagement in decision-mak.

Portugal: No Answer

Luxembourg: No Answer

Bibliographical References

France :

Alava, S., Frau-Meigs, D., & Hassan, G. (2017). Youth and Violent Extremism on Social Media: Mapping the Research. UNESCO Publishing, 5-6.

Albarian, p. 45-49 (2018). Bref aperçu du traitement juridique de la désinformation en droit comparé : de la mise en place de sanctions strictement internes au recours à de véritables sanctions extraterritoriales. LEGICOM, N° 60(1), 45-49. https://doi.org/10.3917/legi.060.0045

Amiel, P. (2020). Aider la presse à lutter contre le virus des fake news (pp. 5-9). Presses universitaires de Grenoble.

Amman, M., & Meloy, R. (2024). Incitement to violence and stochastic terrorism: legal, academic, and practical parameters for researchers and investigators. Terrorism and Political Violence, 36(2), 234-245.

Anderson, O. (2019). Addressing digital divide: Strategies for equity in online learning. Frontiers of Educational Review, 1(1), 1.

Andoulsi, I. (2022). Aspects pratiques du règlement général sur la protection des données personnelles pour les cabinets d'expertise médicale. Consilio manuque, 49(1), 17-33.

Andrejevic, M. (2013). Surveillance in the digital enclosure. In The New Media of Surveillance (pp. 18-40). Routledge.

Avia Law (France). (2020, June 24). In Wikipedia. Retrieved from https://en.wikipedia.org/wiki/Avia_Law_(France)

Badouard, R. (2020). La régulation des contenus sur Internet à l'heure des « fake news » et des discours de haine. Communications, (1), 161-173.

Barkun, M., & Rollet, B. (2016). Les théories du complot comme connaissance stigmatisée. Diogène, 249-250(1), 168-176. https://doi.org/10.3917/dio.249.0168

Barth, N., Wagner, E., Raab, P., & Wiegärtner, B. (2023). Contextures of hate: Towards a systems theory of hate communication on social media platforms. The Communication Review, 26(3), 209-252.

Beddiar, D. R., Jahan, M. S., & Oussalah, M. (2021). Data expansion using back translation and paraphrasing for hate speech detection. Online Social Networks and Media, 24, 100153.

Binns, R., Van Kleek, M., Veale, M., Lyngs, U., Zhao, J., & Shadbolt, N. (2018, April). 'It's Reducing a Human Being to a Percentage' Perceptions of Justice in Algorithmic Decisions. In Proceedings of the 2018 Chi conference on human factors in computing systems (pp. 1-14).

Blaya, C. (2018). Cyberbullying in France. International Perspectives on Cyberbullying: Prevalence, Risk Factors and Interventions, 169-187.

Blaya, C. (2020). Cyberviolence, cyberbullying, and cyberhate: Consequences and protective factors. Le Journal des psychologues, 382(10), 38-43.

Caramizaru, A., & Uihlein, A. (2020). European Commission: Joint Research Centre. Energy communities: an overview of energy and social innovation.

Cetro, R., & Sini, L. (2020). Fake news, rumeurs, intox... : stratégies et visées discursives de la désinformation. L'Harmattan.

Collectively, these organizations and studies employ a range of methods, including machine learning algorithms, data analysis, and collaborative reporting, to monitor and analyze hate speech trends on digital platforms.

Compton, J., van der Linden, S., Cook, J., & Basol, M. (2021). Inoculation theory in the post-truth era: Extant findings and new frontiers for contested science, misinformation, and conspiracy theories. Social and Personality Psychology Compass, 15(6), e12602. 22.

Dauphin, F. (2023). Le debunking sur YouTube : une nouvelle pratique de lutte contre la désinformation en marge du journalisme. Les Enjeux de l'information et de la communication, (S1), 31-45.

De Blasio, E., & Selva, D. (2019). Implementing open government: A qualitative comparative analysis of digital platforms in France, Italy and United Kingdom. Quality & Quantity, 53(2), 871–896. https://doi.org/10.1007/s11135-018-0793-7

De Montbrial, T., & David, D. (2017). Ramses 2018 : La guerre de l'information aura-t-elle lieu ? https://doi.org/10.3917/ifri.demon.2017.01. 14.

Educational Programs :

Facilitated Workshops :

Fake-News ! (N.d.). Pouvoirs et conflits autour de l'énonciation publique du « vrai ». Études de communication, 53(2), 218.

Ferrara, E. (2017). Disinformation and social bot operations in the run up to the 2017 French presidential election. arXiv preprint arXiv:1707.00086. 17-20.

Fries, F. (2021). L'emprise du faux : désinformation, le temps du combat. Éditions de l'Observatoire. 3.

Friggeri, A., Adamic, L. A., & Eckles, D. (2014). Rumor cascades. Proceedings of the 2014 ACM Conference on Computer Supported Cooperative Work & Social Computing, 9-17. https://doi.org/10.1145/2531602.2531608

Friggeri, A., Adamic, L., Eckles, D., & Cheng, J. (2014, May). Rumor cascades. In Proceedings of the International AAAI Conference on Web and social media (Vol. 8, No. 1, pp. 101-110). 10.

Friggeri, A., Adamic, L., Eckles, D., & Cheng, J. (2014, May). Rumor cascades. In Proceedings of the International AAAI Conference on Web and social media (Vol. 8, No. 1, pp. 101-110).

Friggeri, A., Adamic, L., Eckles, D., & Cheng, J. (2014, May). Rumor cascades. In proceedings of the international AAAI conference on web and social media (Vol. 8, No. 1, pp. 101-110).

Friggeri, A., Adamic, L., Eckles, D., & Cheng, J. (2014, May). Rumor cascades. In Proceedings of the International AAAI Conference on Web and social media (Vol. 8, No. 1, pp. 101-110).

Gagliardone, I., Gal, D., Alves, T., & Martinez, J. (2015). Countering online hate speech. UNESCO Publishing. 5-6.

Garland, J., Ghazi-Zahedi, K., Young, J. G., Hébert-Dufresne, L., & Galesic, M. (2020). Countering hate on social media: Large scale classification of hate and counter speech. arXiv preprint arXiv:2006.01974. 22.

Gauducheau, N., & Marcoccia, M. (2023). Verbal violence in a discussion forum for 18–25-year-olds: How do young people judge messages? Reseaux, 241(5), 79-122.

Gazagne-Jammes, V. (2020). La citoyenneté numérique à l'épreuve des fausses informations. La Revue des droits de l'homme. Revue du Centre de recherches et d'études sur les droits fondamentaux, (17).

Gelfert, A. (2018). Fake News: A Definition. Informal Logic, 38(1), 84–117.

Géré, F. (2018). Sous l'empire de la désinformation : la parole masquée. Economica. 9.

GILLESPIE, T. (2018). Custodians of the Internet: Platforms, Content Moder ation, and the Hidden Decisions that Shape social media. Yale University Press.21.

Gorwa, R., Binns, R., & Katzenbach, C. (2020). Algorithmic content moderation: Technical and political challenges in the automation of platform governance. Big Data & Society, 7(1), 2053951719897945.12.

Grosbois, P. de. (2022). La collision des récits : le journalisme face à la désinformation. Éditions Écosociété.

Guess, A. M., Lerner, M., Lyons, B., Montgomery, J. M., Nyhan, B., Reifler, J., & Sircar, N. (2020). A digital media literacy intervention increases discernment between mainstream and false news in the United States and India. Proceedings of the National Academy of Sciences, 117(27), 15536-15545.

Hampton, K. N., Goulet, L. S., Rainie, L., & Purcell, K. (2016). Social Media and the 'Spiral of Silence': How social media Contributes to Political Silence. Journal of Computer-Mediated Communication, 21(4), 178-190.

Hassan, G., Rabah, J., Madriaza, P., Brouillette-Alarie, S., Borokhovski, E., Pickup, D., & Danis, E. (2022). PROTOCOL: Hate online and in traditional media: A systematic review of the evidence for associations or impacts on individuals, audiences, and communities. Campbell Systematic Reviews, 18(2), e1245.

Heinze, E. (2016). Hate Speech and Democratic Citizenship. Oxford University Press, 4-5.

Howard, P. N. (2020). Lie Machines: How to Save Democracy from Troll Armies, Deceitful Robots, Junk News Operations, and Political Operatives. Yale University Press, 11-13.

Howard, P. N., & Bradshaw, S. (2018). The Global Disinformation Order: 2019 Global Inventory of Organised Social Media Manipulation. Project on Computational Propaganda, Oxford Internet Institute.

Kiambati, F. G., Juma, S. W., & Wawire, B. A. (2024). Accessibility of digital systems in information retrieval by users with visual impairment. Quality Assurance in Education.25.

Kilvington, D. (2021). The virtual stages of hate: Using Goffman's work to conceptualize the motivations for online hate. Media, Culture & Society, 43(2), 256-272.

Kross, E., & Ayduk, O. (2011). Making meaning out of negative experiences by self-distancing. Current directions in psychological science, 20(3), 187-191.

Levine, M. P., & Murnen, S. K. (2009). "Everybody knows that mass media are/are not [pick one] a cause of eating disorders": A critical review of evidence for a causal link between media, negative body image, and disordered eating in females. Journal of Social and Clinical Psychology, 28(1), 9-42.

Levine, M. P., & Murnen, S. K. (2009). "Everybody knows that mass media are/are not [pick one] a cause of eating disorders": A critical review of evidence for a causal link between media, negative body image, and disordered eating in females. Journal of Social and Clinical Psychology, 28(1), 9-42.

Lewandowsky, S., Ecker, U. K. H., & Cook, J. (2017). Beyond Misinformation: Understanding and Coping with the "Post-Truth" Era. Journal of Applied Research in Memory and Cognition, 6(4), 353-369. https://doi.org/10.1016/j.jarmac.2017.07.008. p. 353-369.

Lewandowsky, S., Ecker, U. K., & Cook, J. (2017). Beyond misinformation: Understanding and coping with the "post-truth" era. Journal of applied research in memory and cognition, 6(4), 353-369.

Loader, B. D. (2017). Digital divide: The internet and social inequality in international perspective. 3-9.

Lukasik, S., & Salou, A. (2023). La réinformation ou l'amplification socionumérique de la désinformation. Balisages, 7. https://doi.org/10.35562/balisages.1200. 20.

Mazzoleni, G. (2015). Towards an inclusive digital public sphere. In Can the Media Serve Democracy? Essays in Honour of Jay G. Blumler (pp. 174-183). London: Palgrave Macmillan UK.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. ACM computing surveys (CSUR), 54(6), 1-35.

Monneret, J. (2020). Dissidence, dissonance : contre la désinformation à propos de la guerre d'Algérie. Fauves éditions.

Muirhead, R., & Rosenblum, N. L. (2020). A Lot of People Are Saying: The New Conspiracism and the Assault on Democracy. Princeton University Press, 9-14.

Nguyen, D. C., Pham, Q. V., Pathirana, P. N., Ding, M., Seneviratne, A., Lin, Z., ... & Hwang, W. J. (2022). Federated learning for smart healthcare: A survey. ACM Computing Surveys (Csur), 55(3), 1-37.

O'Neil, C. (2017). Weapons of math destruction: How big data increases inequality and threatens democracy. Crown.17-21.

Ostian, E. (2019). Désinformation. Plon. 8.

Pennycook, G., & Rand, D. G. (2018). Fighting misinformation on social media using crowdsourcing: The role of cognitive biases. Proceedings of the National Academy of Sciences, 115(10), 2573-2581. https://doi.org/10.1073/pnas.1718832115

Perspective-Taking:

Piguet, J.-G. (2022). Nudges, désinformation et autonomie citoyenne. Une critique de Sunstein. Éthique Publique, 24(2). https://doi.org/10.4000/ethiquepublique.7583.16.

Piguet, J.-G. (2022). Nudges, désinformation et autonomie citoyenne. Une critique de Sunstein. Éthique Publique, 24(2). https://doi.org/10.4000/ethiquepublique.7583. 15.

Polewka, A. (2023). Outiller les élèves pour combattre la désinformation. Revue internationale d'éducation de Sèvres, (93), 12-16. https://doi.org/10.4000/ries.14003

Rabagny-Lagoa, A. (2022). Fiches de Droit du traitement et de la protection des données personnelles. Editions Ellipses. 9-10.

Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake detection: A systematic literature review. https://doi.org/10.1109/ACCESS.2022.3154404. 4.

Richardson, J., & Milovidov, E. (2019). Digital Citizenship Education Handbook: Being Child in the Age of Technology. Council of Europe Publishing.

Smith, A. (2013). Civic engagement in the digital age. Pew Research Center, 25, 307-332.

Smuha, N. A. (2019). The EU approach to ethics guidelines for trustworthy artificial intelligence. Computer Law Review International, 20(4), 97-106.

Srba, I., Lenzini, G., Pikuliak, M., & Pecar, S. (2021). Addressing hate speech with data science: an overview from computer science perspective. Hate Speech-Multidisziplinäre Analysen und Handlungsoptionen: Theoretische und empirische Annäherungen an ein interdisziplinäres Phänomen, 317-336.

Subjective Phrasing:

Suherlan, S. (2023). Digital Technology Transformation in Enhancing Public Participation in Democratic Processes. Technology and Society Perspectives (TACIT), 1(1), 10-17.

These efforts are complemented by initiatives like Hatebase, which collects and analyzes instances of hate speech worldwide to predict potential regional violence.

Trevisan, F. (2022). Beyond accessibility: Exploring digital inclusivity in US progressive politics. New Media & Society, 24(2), 496-513.

Triplett, W. J. (2023). The role of technology in promoting diversity and inclusion. Cybersecurity and Innovative Technology Journal, 1(1), 37-41.

Tufekci, Z. (2017). Twitter and tear gas: The power and fragility of networked protest. Yale University Press, 15.

van Dijck, J. (2018). The platform society: Public values in a connective world.18.

Van Zomeren, M., Postmes, T., & Spears, R. (2008). Toward an integrative social identity model of collective action: a quantitative research synthesis of three socio-psychological perspectives. Psychological bulletin, 134(4), 504.

Vasist, P. N., Chatterjee, D., & Krishnan, S. (2024). The polarizing impact of political disinformation and hate speech: a cross-country configural narrative. Information Systems Frontiers, 26(2), 663-688.

Voigt, P., & Von dem Bussche, A. (2017). The EU General Data Protection Regulation (GDPR): A practical guide (1st ed., pp. 10-5555). Cham: Springer International Publishing. 18.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146-1151. 7-9. https://doi.org/10.1126/science.aap9559

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. science, 359(6380), 1146-1151.11.

Vraga, E. K., & Tully, M. (2015). Media literacy messages and hostile media perceptions: Processing of nonpartisan versus partisan political information. Mass Communication and Society, 18(4), 422-448.

Windisch, S., Wiedlitzka, S., Olaghere, A., & Jenaway, E. (2022). Online interventions for reducing hate speech and cyberhate: A systematic review. Campbell systematic reviews, 18(2), e1243.18.

Zhou, L., Fu, A., Yu, S., Su, M., & Kuang, B. (2018). Data integrity verification of the outsourced big data in the cloud environment: A survey. Journal of Network and Computer Applications, 122, 1-15.

Italy:

Abebe, R., Barocas, S., Kleinberg, J., Levy, K., Raghavan, M., Robinson, D. G. (2020), "Roles for computing in social change". In Proceedings of the 2020 Conference on Fairness, Accountability, and Trasparency.

Acampa S. (2022). La disinformazione incontra il populismo. Un'analisi del contesto italiano. Rivista di Digital Politics, 3, 507-522. Doi: 10.53227/106457

Al Serhan, F., & Elareshi, M. (2019). University Students' Awareness of Social Media Use and Hate Speech in Jordan. International Journal of Cyber Criminology, 13(2), 548-563. doi:10.5281/zenodo.3709236

Alaieri, F., Vellino, A. (2016), "Ethical decision making in robots: Autonomy, trust and responsibility". In International conference on Social Robotics

Albright, J. (2017). Welcome to the era of fake news. Media and Communication, 5(2): 87-89. doi: 10.17645/mac. v5i2.977

Alghamdi, J., Luo, S., & Lin, Y. (2023). A comprehensive survey on machine learning approaches for fake news detection. Multimedia Tools and Applications, 83(17), 51009–51067. doi:10.1007/s11042-023-17470-8

Allcott, H., & Gentzkow, M. (2017). Social Media and Fake News in the 2016. Election. Journal of Economic Perspectives, 31(2), 211-36. doi: 10.1257/jep.31.2.211

Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. Journal of economic perspectives, 31(2), 211-236.

Allport, G. W. (1954). The Nature of Prejudice. Addison-Wesley.

Amazeen M. A. (2020). Journalistic interventions: The structural factors affecting the global emergence of fact-checking. Journalism, 21(1), 95-111. Doi: 10.1177/1464884917730217

Amores, J. J., Blanco-Herrero, D., Sánchez-Holgado, P., & Frías-Vázquez, M. (2021). Detectando el odio ideológico en Twitter. Desarrollo y evaluación de un detector de discurso de odio por ideología política en tuits en español. Cuadernos. info, 49, 98-124. Doi:10.7764/cdi.49.27817

Anindita, R., Sadiyah, A. A., & Khoiriyah, N. (2022). Income and price elasticities of animal food demand and welfare in Indonesian urban: an application of the LA-AIDS. Future of Food: Journal on Food, Agriculture and Society, 11(1). doi:10.17170/kobra-202210056939

Anjum, & Katarya, R. (2024). Hate speech, toxicity detection in online social media: a recent survey of state of the art and opportunities. International Journal of Information Security, 23(1), 577–608. doi:10.1007/s10207-023-00755-2

Annany, M., Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. New Media & Society, 20 (3), 973-989. doi: 10.1177/1461444816676645

Anneser, E., Levine, P., Lane, K. J., & Corlin, L. (2024). Climate stress and anxiety, environmental context, and civic engagement: A nationally representative study. Journal of Environmental Psychology, 93(102220), 102220. doi: 10.1016/j.jenvp.2023.102220

Antenore M. & S. Splendore (Eds) (2017). Data Journalism. Guida essenziale alle notizie fatte coni numeri. Milano, Mondadori.

Bagnato, K. (2020). Online hate speech: responsabilità pedagogico-educative. Annali online della Didattica e della Formazione Docente, 12(20), 195-211. doi:10.15160/20381034/2274

Bakir, V., & McStay, A. (2018). Fake News and The Economy of Emotions: Problems, causes,solutions. Digital Journalism, 6 (2): 1-22.

Barbieri M. (2024). A subtle hypocrisy: insights into the Italian politics' inertia on the counter-disinformation policy. Rivista di Digital Politics, 1, 91-110.

Barile, N., & Panarari, M. (2020). Viralpolitik. Pandemia, populismi e altre infodemie. Mediascapes journal, 15, 211-221. https://bit.ly/4bMGAXn

Bartels, L. (2024). Challenges of Applying GDPR to Large Language Models. Journal of Data Protection, 18 (2), 45–67. doi: 10.1093/grurint/ikae060

Battista, D. (2023). Disinformation as a danger to international security: An exploration of the implications in the Italian context. Geopolitical, Social Security and Freedom Journal, 6 (1-2).

Baum, S. D. (2020). Social choice ethics in artificial intellingence. AI & Society, 35, 165-176. doi: 10.1007/s00146-017-0760-1

Baumer, E. P.S. (2017). Toward human-centered algorithm design. Big Data & Society, 4 (2). doi: 10.1177/2053951717718854

Bazzaco, E., García-Juanatey, A., Lejardi, J., Palacios, A., & Tarragona, L. (2018). ¿Es odio? Manual práctico para reconocer y actuar frente a discursos y delitos de odio. Institut de Drets Humans de Catalunya, SOS Racisme Catalunya. https://bit.ly/3Y5CLZY

Beauchamp, T. L., Childress, J. F. (1979). Principles of Biomedical Ethics. Oxford University Press, New York.

Belluati M. & A. Fubini (2022). Reacting to disinformation. The multilevel EU fact-checking approach. De Europa. European and Global Studies Journal, 5 (2), 55-75. doi: 10.13135/2611-853X/7143

Benkler, Y., (2019). Don't let industry write the rules for AI. Nature, 569 (7754), 161-162. doi: 10.1038/d41586-019-01413-1

Bentivegna S., Boccia Artieri G, (2021). Voci dalla democrazia. Il futuro del dibattito pubblico, Il Mulino, Bologna.

Boccia Artieri G., (2012). Stati di connessione. Pubblici, cittadini e consumatori nella (Social) Network Society, FrancoAngeli, Milano.

Borriello, G. (2023). «Spid vote»: fast democracy and digital identity for citizens in Italy. Rivista di Digital Politics, 1, 49-72, doi: 10.53227/107476

Botha, J., & Pieterse, H. (2020, March). Fake news and deepfakes: A dangerous threat for 21st century information security. In ICCWS 2020 15th International Conference on Cyber Warfare and Security. Academic Conferences and publishing limited (p. 57).

Bracciale, R., & Grisolia, F. (2020). Information disorder: acceleratori tecnologici e dinamiche sociali. Federalismi.it.

British Academy, The Royal Society (2017), "Data management and use: Governance in 21st century – A joint report by the British Academy and Royal Society

Broda, E., & Strömbäck, J. (2024). Misinformation, disinformation, and fake news: lessons from an interdisciplinary, systematic literature review. Annals of the International Communication Association, 48(2), 139–166. doi:10.1080/23808985.2024.2323736

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., Dafoe, A. (2020). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. doi: 10.48550/arXiv.1802.07228

Buckingham, D. (2017). Fake news: is media literacy the answer?

Capano, G. (2018). Policy design spaces in reforming governance in higher education: The dynamics in Italy and The Netherlands. High Education, 75, 675–694. https://www.jstor.org/stable/26449275

Cath, C., Wachter, S., Mittelstadt, B. (2018). Artificial Intelligence and the 'Good Society': the US, EU, and UK approach. Sci Eng Ethics, 24, 505–528. doi: 10.1007/s11948-017-9901-7

Chan M.S, Jones C.R, Hall Jamieson K & D. Albarracín (2017). Debunking: A Meta-Analysis of the Psychological Efficacy of Messages Countering Misinformation. Psychol Sci. 28(11), 1531-1546. doi: 10.1177/0956797617714579

Cinelli M., Quattrociocchi W., Galeazzi A., Valensise C.M., Brugnoli E., Schmidt A.L, Zola P., Zollo F., & A. Scala (2020). The COVID-19 social media infodemic. Sci Rep. 10(1),16598.

Civila, S., Romero-Rodríguez, L. M., & Aguaded, I. (2021). Symbolic-discursive violence and new media: An epistemological perspective. Catalan Journal of Communication & Cultural Studies, 13(1), 43-61. doi:10.1386/cjcs_00038_1

Cogan, J. J. (2000). Citizenship education for the 21st century: Setting the context. Citizenship for the 21st century: An international perspective on education, 1-21.

Cole, J., & Nolan, J. (2019). GimpGirl: Insider perspectives on technology and the lives of disabled women. In K., Ellis, G., Goggin, B., Haller, & R., Curtis, The Routledge Companion to Disability and Media. (pp.233–242). London, Routledge.

Corner, J. (2017). Fake news, post-truth and media–political change. Media, Culture & Society, 39 (7): 1100– 1107. doi:10.1177/0163443717726743

Cotter K., DeCook J. R., & S. Kanthawala (2022). Fact-Checking the Crisis: COVID-19, Infodemics, and the Platformization of Truth. Social Media and Society, 8(1). Doi: 10.1177/20563051211069048

Cowls, J., Tsmados, A., Taddeo, M., Floridi, L., (2019). "Designing Ai for social good: Seven essential factors". In SSRN Electronic Journal.

D'Acci, J. (2004). Cultural studies, television studies, and the crisis in the humanities. In L. Spigel & J. Olsson (Eds.). Television after TV: Essays on a Medium of Transition (pp.418-446). Durham, Duke University Press.

Danks, D., London, A. J., (2017). Algorithmic Bias in Autonomous Systems". Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence AI and autonomy track, pp. 4691-469

Datta, P. (2020). Digital transformation of the Italian public administration: A case study. Communications of the Association for Information Systems, 46, 11. doi: 10.17705/1CAIS.04611

De Blasio E., & D. Selva (2021). Who Is Responsible for Disinformation? European Approaches to Social Platforms' Accountability in the Post-Truth Era. American Behavioral Scientist, 65(6), 825-846. doi: 10.1177/0002764221989784

De Fortuna, A. M., & De Luca Picione, R. (2024). Searching for meaning through conspiracy theories. Considerations on the state of the art of psychological literature and definition of a research agenda from a semiotic dynamic cultural perspective. Culture & Psychology. doi:10.1177/1354067x241246760

democracies. Information, Communication & Society. 22(13), 1973-1988. doi: 10.1080/1369118X.2018.1474241

Di Giulio, M., & Vecchi, G. (2019). Multilevel policy implementation and the where of learning: The case of the information system for school buildings in Italy. Policy Sciences, 52(1), 119–135. doi: 10.1007/s11077-018-9326-4

Di Mascio F., Barbieri M., Natalini A. & D. Selva (2021). Covid-19 and the Information Crisis of Liberal Democracies: Insights from Anti-Disinformation Action in Italy and EU. Partecipazione e conflitto, 14(1), 221-240. doi: 10.1285/i20356609v14i1p221

Edson, C., Tandoc Jr., Lim, Z.W., & Ling, R. (2018). Defining Fake news. Digital Journalism, 6(2): 137 -153. doi: 10.17157/mat.10.3.7281

European Commission. (2012) The Digital Agenda for Europe—Driving European growth digitally, COM 2012/784.

European Commission. (2018). A Multi-Dimensional Approach to Disinformation; Report of the Independent High-Level Group on Fake News and Online Disinformation. Publications Office of the European Union.

European Parliament and Council. (2016). General Data Protection Regulation (GDPR). Regulation (EU) 2016/679.

European Parliament and Council. (2024). Regulation on Artificial Intelligence (EU AI Act). Proposal for a Regulation laying down harmonized rules on artificial intelligence.

Fallis D. (2015). What Is Disinformation? Library Trends, 63 (3), 401-426.

Farhangian, F., Cruz, R. M. O., & Cavalcanti, G. D. C. (2024). Fake news detection: Taxonomy and comparative study. An International Journal on Information Fusion, 103(102140), 102140. doi: 10.1016/j.inffus.2023.102140

Feldman, L. (2011). Partisan differences in opinionated news perceptions: A test of the hostile media effect. Political Behavior, 33(3), 407–432. doi: 10.1007/s11109-010-9139-4

Fenn, N., Sacco, A., Monahan, K., Robbins, M., & Pearson-Merkowitz, S. (2024). Examining the relationship between civic engagement and mental health in young adults: a systematic review of the literature. Journal of Youth Studies, 27(4), 558–587. doi:10.1080/13676261.2022.2156779

Fong, C.P., & To, S.M. (2022). Civic engagement, social support, and sense of meaningfulness in life of adolescents living in Hong Kong: Implications for social work practice. Child & Adolescent Social Work Journal: C & A, 41(1), 1–13. doi:10.1007/s10560-022-00819-7

Galán del Rey, Á. (2017). El papel de los medios en el imaginario social y en la construcción del discurso del odio'. In RICCAP (Ed.), Contar sin odio, odio sin contar: Visibilidad y contra-narrativas del discurso del odio en los medios para futuros periodistas (pp. 40-44). Madrid, Sevilla and Barcelona: RICCAP. https://bit.ly/3WnaWey

Gheno, V. (2019). Cos'è lo hate speech e come si combatte, "Semplice come", 9/9/2019. https://bit.ly/4bLU8lJ

Giglietto F., Iannelli L., Rossi L & A. Valeriani (2016). Fakes, News and the Election: A New Taxonomy for the Study of Misleading Information within the Hybrid Media System (30 November. Convegno AssoComPol 2016)

Giglietto F., Iannelli L., Valeriani A., & L. Rossi (2019). Fake news' is the invention of a liar: How false information circulates within the hybrid news system. Current Sociology, 67(4). doi: 10.1177/00113921198375

Giglietto, F., Iannelli, L., Rossi, L., Valeriani, A., Righetti, N., Carabini, F., Marino, G., Usai, S. & Zurovac, E. (2018). Mapping Italian News Media Political Coverage in the Lead-Up to 2018 General Election. Urbino, Università di Urbino, Carlo Bo.

Giglietto, F., Iannelli, L., Valeriani, A., & Rossi, L. (2019). 'Fake news' is the invention of a liar: How false information circulates within the hybrid news system. Current Sociology, 67(4). Doi: 10.1177/0011392119837536

Gillis T. B., Spiess J. L. (2019). Big Data and discrimination. University of Chicago Law Review, 86(2), 4. https://chicagounbound.uchicago.edu/uclrev/vol86/iss2/4

Girard C. (2015). Pourquoi punir les discours de haine? Esprit. https://esprit.presse.fr/article/charles-girard/pourquoi-punir-les-discours-de-haine-38476

Guess A., Nagler J., Tucker J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Sci Adv. 5(1). doi: 10.1126/sciadv.aau4586

Guido, G., Pichierri, M., Pino, G., & Conoci, R. (2018). The segmentation of elderly consumers: A literature review. Journal of Customer Behaviour, 17, 257-278. doi: 10.1362/147539218X15445233217805

Hagendorff, T. (2024). Mapping the Ethics of Generative AI: A Comprehensive Scoping Review. Minds & Machines, 34, 39. doi: 10.1007/s11023-024-09694-w

Hager, G. D., Drobnis, A. W., Fang, F., Ghani, R., Greenwald, A., Lyons, T., Parkes, D. C., Schultz, J., Saria, S., Smith, S. F., Tambe, M. (2018). Artificial Intelligence for Social Good. Computing Community Consortium. doi:10.48550/arXiv.1901.05406

Hallin, D., & Mancini, P. (2004). Modelli di giornalismo. Bari, Laterza.

Hameleers, M. (2021). Prospect theory in times of a pandemic: The effects of gain versus lossframing on risky choices and emotional responses during the 2020 coronavirus outbreak. Evidence from the US and the Netherlands. Mass Communication and Society, 24(4), 479-499. doi: 10.1080/15205436.2020.1870144

Hammoode, J. A., Neamah, N. R., Hasoon, A., Ali, M. H., Qusai, N., Al-Muttar, M. Y. O., et al. (2022). Empirical Evidence on the Impact of Project Management Capabilities on Projects Success of Construction in Iraq: a Pls-Sem Approach. The Journal of Modern Project Management, 10(2), 384-395.

https://rosa.uniroma1.it/rosa03/mediascapes/article/view/18490

Huber J., Reynolds C., 2014. Développer la compétence interculturelle par l'éducation, Serie Pestalozzi n° 3, Consiglio d'Europa, 2014 https://rm.coe.int/developingintercultural-enfr/16808ce258

Humprecht, E. (2019). Where 'fake news' flourishes: a comparison across four Western

Hutson, M., (2019). Bringing machine leraning to the masses. Science, 365 (6452), 416-417. doi: 10.1126/science.365.6452.416

Ireton, C., & Posetti, J. (Eds) (2018). Journalism, 'Fake News' & Disinformation. Handbook for journalism education and training. UNESCO.

Jubany, O., & Roiha, M. (2016). Backgrounds, Experiences and Responses to Online Hate Speech: A Comparative Cross-country Analysis. Online Report. Barcelona: University of Barcelona. https://bit.ly/4cH0iFi

Jun, Y., Meng, R., & Johar, G. V. (2017). Perceived social presence reduces fact-checking. Proc. Natl. Acad. Sci.114 (23), 5976-5981. doi: 10.1073/pnas.1700175114

Katell, M., Young, M., Dailey, D., Herman, B., Guetler, V., Tam, A., Binz, C., Raz, D., Krafft P. M., (2020). Toward situated intervention for algorithmic equity: Lessons frome dhe field. FAT '20: Conference on Fairness, Accountability, and Trasparency, 27 gennaio.

Kranzberg, M. (1986). Technology and History: 'Kranzbergs Laws'. Technology and Culture, 27(3),544–560. doi:10.2307/3105385

Larson, B., (2017). Gender as a variable in natural-language processing: Ethical considerations. In Proceedings of the first ACI. Workship on Ethics in Natural Language Processing, pp. 1-11.

Lazer, D. M., Baum, M.A., Benkler, Y., Berinsky, A. J., Greenhill, K. M, Menczer, F. & Schudson, M. (2018). The science of fake news. Science, 359(6380): 1094– 1096. Doi: 10.1126/science.aao2998

Lee, M. K., Kim, J. T., Lizarondo, L. (2017). A human-centerd approach to algorithimic services. Considerations for fair and motivating smart community service management that allocates donations to non-profit organizations. In Proceedings of the 2017 CHI Conference on Human Factors in Computing System-CHI", pp. 3365-3376. ACM Press.

Lee, M. S. A., Floridi, L., Denev, A. (2020). Innovating with confidence. Embedding AI Governance and fairness in a financial services risk management framework. Berkeley technology Law Journal, 34 (2), pp. 1-19. https://btlj.org/wp-content/uploads/2020/01/Lee_Web.pdf

Lepri, B., Oliver, N., Letouzé, E., Pentland, A., Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes: The premise, the proposed solutions, and the open challenges. Philosophy & Technology, 31 (4), 611-627. https://hdl.handle.net/1721.1/122933

Lewandowsky S., Cook J., Ecker U. K. H., Albarracín D., Amazeen M. A., Kendeou P., Lombardi D., Newman E. J., Pennycook G., Porter E., Rand D. G., Rapp D. N., Reifler J., Roozenbeek J., Schmid P., Seifert C. M., Sinatra G. M., Swire-Thompson B., van der Linden S., Vraga E. K., Wood T. J. & M. S. Zaragoza (2020). The Debunking Handbook 2020. doi: 10.17910/b7.1182

Lorusso, A. M. (2018). Postverità. Fra reality tv, social media e storytelling. Bari, Laterza.

Lu, C., & Gu, M. M. (2024). A systematic review and meta-analysis of factors and outcomes of digital citizenship among adolescents. Asia Pacific Journal of Education, 1–16. doi:10.1080/02188791.2023.2296352

Maddalena, G. & Gili, G. (2017). Chi ha paura della postverità? Effetti collaterali di una parabola culturale. Genova, Marietti.

Malecki, K. M. C., Keating, J. A., & Safdar, N. (2020). Crisis Communication and Public Perception of COVID-19 Risk in the Era of Social Media. Clinical Infectious Diseases, 72(4), 697-702. https://doi.org/10.1093/cid/ciaa758

Mangiapane, F. (2018). The Discourse of Fake News in Italy. A Comparative Analysis. Versus, Quaderni di studi semiotici, 2, 291-306.

Marcos-Vílchez, J. M., Sánchez-Martín, M., & Muñiz-Velázquez, J. A. (2024). Effectiveness of training actions aimed at improving critical thinking in the face of disinformation: A

systematic review protocol. Thinking Skills and Creativity, 51(101474), 101474. doi: 10.1016/j.tsc.2024.101474

Megías, I., Amezaga, A., García, M. C., Kuric, S., Morado, R., & Orgaz, C. (2020). Romper cadenas de odio, tejer redes de apoyo:los y las jóvenes ante los discursos de odio en la red. Madrid: Centro Reina Sofia sobre Adolescencia y Juventud, Fad. https://doi.org/10.5281/zenodo.4288486

Melchior, C. & Trevisan, F. (2022). Use of digital resources in the Italian disability community: An exploratory study. Proceedings 20th CIRN Conference, Monash Centre, Prato Italy, 9-11 November

Melchior, D., & Strizzolo, N. (2024). Fake news. Origins, connotations (and decline?) of a successful expression. Mediascapes, 23.

Mele, V. (2008). Explaining Programmes for change. The policy cycle of electronic government in Italy (1993–2002). Public Management Review, 10(1), 21–49. doi: 10.1080/14719030701763179

Menesini E. Nocentini A. (2009). Cyberbullying definition and measurement: some critical considerations. Zeitschrift fur Psychologie/Journal of Psychology,217, 230–232. doi: 10.1027/0044-3409.217.4.230

Miconi, A. (2020). Teorie e pratiche del web. Il Mulino.

Ministero degli Affari Esteri e della Cooperazione Internazionale, come individuare e contrastare operazioni coordinate di disinformazione in Italia, 2023.

misinformation, disinformation, and mal-information. In C. Ireton, & J. Posetti, Journalism, 'Fake News' Disinformation Handbook for Journalism Education and Training, (pp.43-53).

Mökander J., Floridi L., (2021). Ethics-based auditing to develop trustworthy AI. Minds and Machines, 31, 323-327. doi: 10.1007/s11023-021-09557-8

Molina, M. D., Sundar, S. S., Le, T., & Lee, D. (2021). "Fake News" Is Not Simply False Information: A Concept Explication and Taxonomy of Online Content. American Behavioral Scientist, 65(2), 180-212. https://doi.org/10.1177/0002764219878224

Morcellini, M. (Ed.) (2011). Neogiornalismi. Tra crisi e rete. Come cambia il sistema dell'informazione. Milano, Mondadori.

Murialdi, P. (2014). Storia del giornalismo italiano. Dalle gazzette a Internet. Bologna, Il Mulino.

Nazem, S. N., Abid, M. M., Gdheeb, S. H., Altememy, H. A., Al Jouani, I. S. H., Mohsen, K. S., et al. (2023). Cybersecurity Determinants in Iraq's Digital Workplace: Attitude, Policy, and Compliance Roles. International Journal of Cyber Criminology, 17(2), 1-19. doi: 10.5281/zenodo.4766701

Nera, K., Bertin, P., Biddlestone, M., Tagand, M., & Klein, O. (2024). Are conspiracy theory believers drawn to conspiratorial explanations, alternatives explanations, or both? Journal of Experimental Social Psychology, 115(104640), 104640. doi: 10.1016/j.jesp.2024.104640

Nocentini A. Calmaestra J. Schultze-Krumbholz A, et al. (2010). Cyberbullying: labels, behaviours and definition in three European countries. Australian Journal of Guidance and Counselling,20, 129–142. doi: 10.1375/ajgc.20.2.129

Organisation for Economic Co-operation and Development (OECD). (2019). Artificial Intelligence in Society: Understanding the Implications and Policy Responses. Organisation for Economic Co-operation and Development.

Paasch-Colberg, S., Strippel, C., Trebbe, J., & Emmer, M. (2021). From Insult to Hate Speech: Mapping Offensive Language in German User Comments on Immigration. Media and Communication, 9(1), 171-180. doi: 10.17645/mac. v9i1.3399

Palmieri, L., Cadeddu, C., Rosano, A., Donfrancesco, C., Elia, D., Mastrilli, R., de Castro, V., Lo Noce, P., Acampora, C., & Giampaoli, A. (2020). The Italian Pilot of the Health Literacy Survey 2019 in the Cuore Project for the Who Action Network on Measuring Population and Organizational Health Literacy (m-pohl). Circulation. 141 (Suppl_1).

Paris-Albert, S. (2021). La fuerza de los sentimientos y el reclamo de una educación sentimental creativa. Autónoma de Madrid. doi:10.15366/bp2022.30.014

Parsloe, S.M., & Holton, A.E. (2018). Boycottautismspeaks: Communicating a counternarrative through cyberactivism and connective action. Information, Communication & Society 21(8):1116–1133. Doi: 10.1080/1369118X.2017.1301514

Pasquetto I. V.. Olivieri A. F., Tacchetti L., Riotta G. & A. Spada (2022). Disinformation as Infrastructure: Making and Maintaining the QAnon Conspiracy on Italian Digital Media. Proc. ACM Hum. -Comput. Interact, 6, CSCW1

Pasta, S. (2021). Ostilità: vecchi e nuovi bersagli, vecchi e nuovi virus. Scholé: revista di educazione e studi cultural, LIX(2), 89-102. https://bit.ly/3Sc0TGu

Polesana, M.A. (2010). Criminality show. La costruzione mediatica del colpevole. Roma: Carocci.

Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., & Stein, B. (2017). A stylometric inquiry into hyperpartisan and fake news. doi:10.48550/arXiv.1702.05638

Qozmar, N., Flayyih, N., Al-Makhzoumi, O., Abughazleh, A. M., Odeh, M., & Abdulkadir, A. (2023). An Applicable Law on Custody: A Comparative Study. Croatian International Relations Review, 29(94), 170-187.

Quattrociocci W. & A. Vicini (2016). Misinformation. Guida alla società dell'informazione e della credulità, Milano: FrancoAngeli.

Rawat, A., Kumar, S., & Samant, S. S. (2024). Hate speech detection in social media: Techniques, recent trends, and future challenges. Wiley Interdisciplinary Reviews. Computational Statistics, 16(2). doi:10.1002/wics.1648

Richardson-Self, L. (2018). Woman-Hating: On Misogyny, Sexism, and Hate Speech. Hypatia, 33(2), 256-272. doi:10.1111/hypa.12398

Riva, G. (2018). Fake news. Bologna, Il Mulino.

Rizzuto, F. (2009). Giornalismo e democrazia. L'informazione politica in Italia. Palermo, Palumbo.

Rizzuto, F. (2012). Lo spettacolo delle notizie. Roma, Aracne.

Robertson C. T., Mourão R. R., & E. Thorson (2020). Who Uses Fact-Checking Sites? The Impact of Demographics, Political Antecedents, and Media Use on Fact-Checking Site Awareness, Attitudes, and Behaviour. The International Journal of Press/Politics, 25(2), 217-237. doi: 10.1177/1940161219898055

Rubin, V. L., Chen, Y., & Conroy, N. K. (2015). Deception detection for news: three types of fakes. Proceedings of the Association for Information Science and Technology. 52(1), 1-4. doi: 10.1002/pra2.2015.145052010083

Sandvig, C., Hamilton, K., Defilippis, E., Radanovic, G., Parkes, D., Liu Y., (2016). When the algorithm itself is a racist: Diagnosis ethical harm in the basic components of software. International Journal of Communication, 10, 4972-4990.

Schudson, M., & Zelizer, B. (2017). Fake News in context. In AA. VV., Understanding and Addressing the Disinformation Ecosystem. Annenberg, University of Pennsylvania.

Seargeant P. & C. Tagg (Eds) (2014). The Language of Social Media. Identity and Community on the Internet. London, Palgrave Macmillan.

Sestino, A., Simonetti, A., & Gastaldi, L. (2021). COVID-19 and digital acceleration: An investigation about citizens' perception of digitalized public services in Italy. Proceedings 4th UNICART Interdisciplinary International Conference, La Valletta, Malta, 15-17 June

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD explorations newsletter, 19(1), 22-36.

Sørensen, K., Pelikan, J.M., Röthlin, F., Ganahl, K., Slonska, Z., Doyle, G., Fullam, J., Kondilis, B., Agrafiotis, D., & Uiters, E. (2015). Health literacy in Europe: Comparative results of the European health literacy survey (HLS-EU). Eur. J. Public Health. 25 (6): 1053-1058. doi: 10.1093/eurpub/ckv043

Sternberg, R. R. (2007). Capire e combattere l'odio. In R. R. Sternberg (Ed.), Psicologia dell'odio. Conoscerlo per superarlo (pp. 45-58). Erickson.

Tabasco M., Ruffo G., Flammini A., & F. Menczer (2015). Fact-checking effect on viral hoaxes: A model of misinformation spread in social networks. Proceedings of the 24th ACM International Conference on World Wide Web, of WWW '15 Companion, 977–982. doi: 10.1145/2740908.27425

Tandoc, E.C., Jenkins, J., & Craft, S. (2018). Fake News as a Critical Incident in Journalism. Journalism Practice, 13(6), 673–689. doi: 10.1080/17512786.2018.1562958

Tandoc, E.C., Lim, W., & Ling, R. (2017). Digital Journalism Defining "Fake News". A typology of scholarly definitions. Digital Journalism, 6(2), 137–153. doi: 10.1080/21670811.2017.1360143

Tangi, L., Benedetti, M., Gastaldi, L., Noci, G., & Russo, C. (2021). Mandatory provisioning of digital public services as a feasible service delivery strategy: Evidence from Italian local governments. Government Information Quarterly, 38, 101543. doi: 10.1016/j.giq.2020.101543

Trevisan, F. (2020). 'Do You Want to Be a Well-Informed Citizen, or Do You Want to Be Sane?' Social Media, Disability, Mental Health, and Political Marginality. Social Media + Society 6 (1). Doi: 10.1177/2056305120913909

Turner Lee, N., (2018). Detecting racial bias in algorithms and machine learning. Journal of Information, Communication and Ethics in Society, 16 (3), 252-260. doi: 10.1108/JICES-06-2018-0056

UNESCO. (2015). Countering online hate speech. UNESCO series on internet freedom. https://bit.ly/3y1PeDu

UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. United Nations Educational, Scientific and Cultural Organization.

Uuk, R. (2024). A Taxonomy of Systemic Risks from General-Purpose AI. arXiv preprint.

Vaccaro, S., Rizzuto, F, & Brancato, G. (Eds). (2019). La Comunicazione alla sfida della (dis)informazione. Roma, Aracne.

Valentini, E. (2012). Dalle gazzette all'iPad. Milano, Mondadori.

Vargo C. J., Guo L., & M. A. Amazeen (2018). The agenda-setting power of fake news: A big data analysis of the online media landscape from 2014 to 2016. New Media & Society, 20(5), 2028–2049. doi: 10.1177/1461444817712086

Vicari, S., & Cappai, F. (2016). Health activism and the logic of connective action. A case study of rare disease patient organisations. Information, Communication & Society, 19(11), 1653–1671. doi: 10.1080/1369118X.2016.1154587

von Gillern, S., Rose, C., & Hutchison, A. (2024). How students can be effective citizens in the digital age: Establishing the Teachers' Perceptions on Digital Citizenship Scale. British Journal of Educational Technology: Journal of the Council for Educational Technology, 55(5), 2093–2109. doi:10.1111/bjet.13434

Waldron, J. (2012). The Harm in Hate Speech. Harvard University Press. https://doi.org/10.4159/harvard.9780674065086

Wardle C., & H. Derakhshan (2017). Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking. Council of Europe.

Wardle, C., & Derakhshan, H. (2018). Thinking about 'information disorder': formats of

Webb, H., Patel, M., Rovatos, M., Davoust, A., Ceppi, S., Koene, A., Downthwaite, L., Portillo, V., Jirotka, M., Cano, M., (2019). 'It would be pretty immoral to choose a random algorithm': Opening up altorithmic interpretability and transparency. Journal of Information, Communication and Ethics in Society, 17 (2), 210-228. doi: 10.1108/JICES-11-2018-0092

WHO. (2023). Ethics and Governance of Artificial Intelligence for Health. World Health Organization.

Zaff, J. F., Kawashima-Ginsberg, K., & Lin, E. S. (2011). Advances in civic engagement research. In Advances in Child Development and Behavior. Positive Youth Development (pp. 273–308). doi:10.1016/b978-0-12-386492-5.00011-7

Zarocostas, J. (2020). How to fight an infodemic. Lancet, 395 (10225), 676. doi:10.1016/S0140-6736(20)30461-X

Zelizer, B. (2004). Taking journalism seriously. UK, Sage.

Zerilli, J., Knott, A., Maclaurin, J., Gavaghan, C., (2019). Transparency in algorithmic and human decision-making: Is there a double standard? Philosophy & Technology, 32 (4), 661-683. doi: 10.1007/s13347-018-0330-6

Turkey:

Akser, M. (2020). Digital Populism in Turkey: Trolls and algorithmic control of online discourse. International Journal of Communication, 14, 1080-1099.

Benesch, S. (2014). Countering Dangerous Speech: New ideas for genocide prevention. Genocide Studies and Prevention, 7(1), 47–53.

boyd, d. (2010). Social network sites as networked publics. Networked Self, 39-58.

BTK (2022). Cybersecurity and Data Privacy Regulations in Türkiye.

Bulger, M., & Davison, P. (2018). The Promises, Challenges, and Futures of Media Literacy. Data & Society Research Institute.

Cinelli, M., Morales, G. D. F., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). "The echo chamber effect on social media." Proceedings of the National Academy of Sciences, 118(9).

Dahlgren, P. (2021). Media and Political Engagement: Citizens, Communication, and Democracy. Cambridge University Press.

Douglas, K. M., Uscinski, J. E., Sutton, R. M., Cichocka, A., Nefes, T., Ang, C. S., & Deravi, F. (2019). Understanding conspiracy theories. Political Psychology, 40(S1), 3–35.

Ellcessor, E. (2016). "Digital inclusion." In G. Goggin & K. Ellis (Eds.), The Routledge Companion to Disability and Media (pp. 171–181). Routledge.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. Journal of Communication, 43(4), 51-58.

European Commission. (2021). Media literacy and online disinformation. Publications Office of the EU.

Gillespie, T. (2018). Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions that Shape social media. Yale University Press.

Gillespie, T. (2018). Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. Yale University Press.

Gurin, P., Nagda, B. R., & Zúñiga, X. (2013). Dialogue Across Difference: Practice, Theory, and Research on Intergroup Dialogue. Russell Sage Foundation.

Habitat Association (2022). Digital Literacy and Inclusion Programs.

Habitat Derneği. (2022). Dijital Dönüşüm Programları.

Hobbs, R., & Frost, R. (2003). Measuring the acquisition of media-literacy skills. Reading Research Quarterly, 38(3), 330–355.

Huguet, A., Kavanagh, J., Baker, G., & Blumenthal, M. S. (2019). Exploring Media Literacy Education as a Tool for Mitigating Truth Decay. RAND Corporation.

Istanbul Bilgi Üniversitesi. (2022). Digital Media and Information Literacy Programs.

KVKK (2023). Personal Data Protection Law No. 6698.

Livingstone, S., & Helsper, E. J. (2007). Gradations in digital inclusion: Children, young people, and the digital divide. New Media & Society, 9(4), 671-696.

Livingstone, S., & Helsper, E. J. (2007). Gradations in digital inclusion: Children, young people, and the digital divide. New Media & Society, 9(4), 671–696.

Massanari, A. (2017). "#Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures." New Media & Society, 19(3), 329-346.

Millî Eğitim Bakanlığı [MEB]. (2023). Dijital Okuryazarlık Eğitimi Programı. Ankara: MEB Yayınları.

O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group.

Özdemir, E., & Şeker, A. (2023). AI and Disinformation Analysis in Digital Diplomacy. Journal of Security Studies, 5(1), 33-45.

Özkırımlı, U. (2021). Social media, hate speech and content moderation in Turkey. Social media + Society, 7(1), 1-15.

Patchin, J. W., & Hinduja, S. (2012). Cyberbullying: An update and synthesis of the research. Cyberpsychology, Behavior, and Social Networking, 14(2), 104-108.

Putnam, R. D. (2000). Bowling Alone: The Collapse and Revival of American Community. Simon & Schuster.

Resende, G., Melo, P., Sousa, H., Messias, J., Vasconcelos, M., Almeida, J., & Benevenuto, F. (2019). "(Mis)Information Dissemination in WhatsApp: Gathering, Analyzing and Countermeasures." The World Wide Web Conference, 818–828.

Ribeiro, M. H., Ottoni, R., West, R., Almeida, V. A., & Meira Jr., W. (2020). "Auditing radicalization pathways on YouTube." Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, 131–141.

Saka, E. (2018). Social media in Turkey as a space for political battles. Middle East Critique, 27(2), 161-177.

Skafle, I., Nordahl-Hansen, A., Quintana, D. S., Wynn, R., & Gabarron, E. (2022). Misinformation about COVID-19 Vaccines on social media: Rapid Review. Journal of medical Internet research, 24(8), e37367. pGencel Bek, M. (2019). Media literacy in Türkiye: A critical approach. Turkish Journal of Media Studies, 4(1), 34-56.

Suler, J. (2004). The online disinhibition effect. CyberPsychology & Behavior, 7(3), 321-326.

Sunstein, C. R. (2017). Republic: Divided Democracy in the Age of Social Media. Princeton University Press.

Tajfel, H., & Turner, J. (1986). The social identity theory of intergroup behavior. Psychology of Intergroup Relations, 7-24.

Teyit (2023). "Methodology."

Teyit.org. (2023).

These experiences show how Türkiye effectively promotes digital democracy by combining educational policies, civil initiatives, and community engagement.

These initiatives demonstrate Türkiye's commitment and practical experience in fostering digital citizenship and democracy.

Wardle, C., & Derakhshan, H. (2017). Information Disorder: Toward an interdisciplinary framework for research and policymaking. Council of Europe.

Yesil, B., & Sözeri, E. K. (2018). Online surveillance and censorship in Turkey: Implications for freedom of expression. Surveillance & Society, 16(2), 222-232.

Portugal:

Agência Lusa. (2024 October 19). Um em cada 10 jovens foi vítima de cyberbullying muito frequentemente. Retrieved on 3 March 2025, https://cnnportugal.iol.pt/cyberbullying/estudo/um-em-cada-10-jovens-foi-vitima-de-cyberbullying-muito-frequentemente/20241019/671370f6d34ea1acf26fa5b4

APAV. (2018). Campanha contra o Discurso de Ódio – Jovens pelos Direitos Humanos online. https://apav.pt/campanha-contra-o-discurso-de-odio-jovens-pelos-direitos-humanos-online/

Ashikuzzaman, M. (2024 July 11). The Role of Media Literacy in Combating Misinformation. Retrieved on 3 March 2025, de https://www.lisedunetwork.com/the-role-of-media-literacy-in-combating-misinformation/

Asteriou, M. G., & Konstantina. (2023). Digital literacy of older people. Digital Life Learning.

Ausat, A. M. (2023). The role of social media in Shaping Public Opinion and Its Influence on Economic Decisions. Technology and Society Perspectives, 35-44.

Awad, N. (2023 September 22). Common Digital Inclusion Strategies. Retrieved from https://www.accessibility.com/blog/common-digital-inclusion-strategies

Bader, L., & Bender, J. (2022). What is "fake news" and "hate speech" and how do they work in practice? Em S. Brüggemann, N. Kutlu, R. Müller-Török, S. Rucinska, T. Szadeczky, & C. Vrabie, COUNTERFAKE: A scientific basis for a policy gighting fake news and hate speech (pp. 17-36). WU Universitätsverlag/Facultas. d

Barata, J. (2024). THE FUTURE OF FREE SPEECH: OLD THREATS AND NEW CHALLENGES. Em E. Heinze, N. Alkiviadou, T. Herrenberg, S. Parmar, & I. Tourkochoriti, OXFORD HANDBOOK OF HATE SPEECH.

Barreto, A. (2002). A participação cívica e política e a evolução da sociedade portuguesa. O estado e o cidadão, Actas dos VIII Cursos Internacionais de Verão de Cascais, 45-60.

Bi, R. (2024). Digital Democracy: Enhancing Democratic Participation aand Political Transparency Through Digital Techonologies. Lecture Notes in Education Psychology and Public Media, 285-291.

Botelho, F. H. (2021). Accessibility to digital technology: Virtual barriers, real opportunities. Assistive Techonology , 527-534.

BPI. (September 2023). Compreender o crescimento do discurso de ódio online, em Portugal e Espanha: um hiato entre a ocorrência e a denúncia. Retrieved on 11 March 2025, https://oobservatoriosocial.fundacaolacaixa.pt/pt/-/compreender-o-crescimento-do-discurso-de-odio-online-em-portugal-e-espanha-um-hiato-entre-a-ocorrencia-e-a-denuncia

Candeias, P. (2023). Teorias da conspiração e a sua relação com comportamentos de saúde. Em R. R. Santos, O. Santos, & A. Abreu, Saúde Ambiental - caderno de notas soltas III (pp. 111-120).

Casa do Brasil em Portugal. (2025). Discurso de Ódio e Imigração em Portugal. Retrieved from https://casadobrasildelisboa.pt/wp-content/uploads/2024/MM_relatorio2024_digital_v7.pdf

Centro Internet Segura. (n.d.). Desinformação e fake news. Retrieved on 5 March 2025, https://www.internetsegura.pt/FakeNews

Chen, S., Xiao, L., & Kumar, A. (2023). Spread of misinformation on social media: What contributes to it and how to combat it. Computers in Human Behavior.

Coelho, V. A., Sousa, V., Marchante, M., Brás, P., & Romão, A. M. (2016). Bullying and cyberbullying in Portugal: Validation of a questionnaire and analysis of prevalence. School Psychology International, 1-17.

CoExist. (2025). How Social Media Can Promote Dialogue and Understanding. Retrieved from https://www.coexistwithpeace.com/social-media-and-dialogue/#:~:text=Here%20are%20key%20strategies%20to%20enhance%20the%20quality,and%20express%20their%20views%20in%20a%20respectful%20manner.

Comissão de Coordenação e Desenvolvimento Regional do Alentejo. (n.d). Orçamento Participativo de Portugal. Retrieved on 25 February 2025, https://www.ccdr-a.gov.pt/orcamento-participativo-de-portugal/

Comissão Nacional de Proteção de Dados. (n.d). O que somos e quem somos. Retrieved from https://www.cnpd.pt/cnpd/o-que-somos-e-quem-somos/

Costa, F. (2024 May 9). Desinformação e fake news são a mesma coisa? Um guia para não te perderes. Retrieved on 14 March 2025, https://www.publico.pt/publico-na-escola/artigo/desinformacao-fake-news-sao-guia-nao-perderes-2089695

DECO. (n.d.). PROGRAMA EUSOUDIGITAL. Retrieved on 27 February 2025, https://deco.pt/eventos/lancamento-do-eusoudigital-programa-de-capacitacao-digital-de-adultos/?utm_source=chatgpt.com

Diário da República. (2022 December 28). Resolução do Conselho de Ministros n.º 139/2022, de 28 de dezembro. Retrieved on 4 March 2025, https://diariodarepublica.pt/dr/detalhe/resolucao-conselho-ministros/139-2022-205378564

Douglas, K. M., Uscinski, J. E., Sutton, R. M., Nefes, T., Ang, C. S., & Deravi, F. (2019). Understanding Conspiracy Theories. Advances in Political Psychology, 40(1), 4-35.

e-Global. (2024 June 4). Dezenas de milhares de comentários e tweets analisados para compreender o discurso de ódio online em Portugal. Retrieved on 11 March 2025, https://e-global.pt/noticia/vida/dezenas-de-milhares-de-comentarios-e-tweets-analisados-para-compreender-o-discurso-de-odio-online-em-portugal/

Ejaz, W., Fletcher, R., Nielsen, R. K., & McGregor, S. C. (2024). What do People Want? Views on Platforms and the Digital Public Sphere in Eight Countries. Reuters Institute of the Study of Journalism at the University of Oxford.

Europe, Council of. (n.d.). Dealing with propaganda, misinformation and fake news. Retrieved on 2 Março 2025, de https://www.coe.int/en/web/campaign-free-to-speak-safe-to-learn/dealing-with-propaganda-misinformation-and-fake-news

European Association for the Education of Adults. (2021). Digitalisation & Democracy. Retrieved from https://eaea.org/wp-content/uploads/2021/11/Grundtvig_Award_2021_B5_lowres.pdf?utm_source=chatgpt.com

European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). Official Journal of the European Union, Retrieved from https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:32016R0679

Executive. (17 de Feebruary de 2024). Lei da UE contra conteúdos ilegais na internet entra em vigor este sábado. O que muda? Retrieved on 10 March 2025, https://executivedigest.sapo.pt/noticias/lei-da-ue-contra-conteudos-ilegais-na-internet-entra-em-vigor-este-sabado-o-que-muda/

Feio, C., Oliveira, L., & Martins, J. M. (2023). Social media as a way of young people's political expression in Portugal. Observatorio Journal, 205-223.

Fischli, R., & Muldoon, J. (2024). Empowering Digital Democracy. Perspectives on Politics, 22(3), 819-835.

GDPR. (n.d). Cookies, the GDPR, and the ePrivacy Directive. Retrieved from https://gdpr.eu/cookies/

Gonçalves, J. (2024 July 10). Investigadores portugueses criam ferramenta para detectar discurso de ódio online. Retrieved on 13 March 2025, https://www.publico.pt/2024/07/10/sociedade/noticia/investigadores-portugueses-criam-ferramenta-detectar-discurso-odio-online-2096983

Governo da República Portuguesa. (2025 February 24). Governo vai criar linha de denúncia e apoio a alunos vítimas de bullying. Retrieved on 4 March 2025, https://www.portugal.gov.pt/pt/gc24/comunicacao/noticia?i=governo-vai-criar-linha-de-denuncia-e-apoio-a-alunos-vitimas-de-bullying

Governo da República Portuguesa. (2025). Plano Nacional de Literacia Mediática - Estratégia 2025-2029. Retrieved on 25 February 2025, https://www.portugal.gov.pt/download-ficheiros/ficheiro.aspx?v=%3d%3dBQAAAB%2bLCAAAAAAABAAzNDExNgAADRH61AUAAAA%3d

Group, FI. (2025 February 4). Proteção de dados: quais os desafios para as empresas? Retrieved on 11 March 2025, https://pt.fi-group.com/protecao-de-dados-quais-os-desafios-para-as-empresas/

Hatano, A. (2023). Regulating Online Hate Speech through the Prism of Human Rights Law: The Potential of Localised Content Moderation. The Australian Year Book of International Law Online, 127-156.

Hive Mind. (2023 February 17). Empowering Europeans Against Disinformation (EEAD) - Project Description, Scope, and Activity Updates. Retrieved on 6 March 2025, de https://en.hive-mind.community/blog/297%2Cempowering-europeans-against-disinformation-eead-project-description-scope-and-activity-updates?utm_source=chatgpt.com

Jākobsone, M. (2021). Portugal - InCode2030 Programme. Retrieved on 26 February 2025, https://digital-skills-jobs.europa.eu/en/actions/national-initiatives/national-strategies/portugal-incode2030-programme

Keen, E. (2016). Manual para o combate do discurso de ódio online através da educação para os direitos humanos. Fundação Calouste Gulbenkian.

Livingstone, S. (5 de Juy de 2022). The vital role of measuring impact in media literacy initiatives. Retrieved on 2025 March, https://edmo.eu/edmo-news/the-vital-role-of-measuring-impact-in-media-literacy-initiatives/

Lusa. (20 de February de 2025). Centenas de pessoas assinam carta a pedir sanções por discurso de ódio no Parlamento. Retrieved on 6 March 2025, https://www.publico.pt/2025/02/20/politica/noticia/centenas-pessoas-assinam-carta-pedir-sancoes-discurso-odio-parlamento-2123220

Majka, M. (2024). Understanding the Importance of Pre and Post -Testing in Research and Evaluation.

Mishova, A. (August de 2024). Data Protection Laws Around the World: A Global Perspective. Retrieved from https://gdprlocal.com/data-protection-laws-around-the-world-a-global-

perspective/#:~:text=Data%20protection%20laws%20vary%20globally%2C%20with%20each%20region,tackle%20the%20unique%20challenges%20of%20the%20digital%20age.

Monteiro, R. (2025, 25 February). Jovens portuguses são dos que mais votam e vêem notícias na TV (uma surpresa). Público. https://www.publico.pt/2025/02/17/p3/noticia/jovens-portugueses-sao-votam-veem-noticias-tv-surpresa-2122810

MUDA. (n.d.). Quem somos. https://www.muda.pt/quem-somos/

Nguyen, A. (2021). Digital Inclusion: Social Inclusion in the Digital Age. Em P. LiamPuttong, Handbook of Social Inclusion, Research & Practices in health and social care.

Observátorio Português da Acessibilidade Web. (n.d). Observátorio. Retrieved on 26 February 2025, https://observatorio.acessibilidade.gov.pt/

Oliveira, A. F., Maneta, M., Brites, M. J., & Ribeiro-Rodrigues, V. (2024). How is Portugal addressing disinformation? Results of a mapping of initiatives (2010-2023). Observatório Journal, 158-174.

Paraense, A. C. (2022). O debate político e os desafios da democracia nas redes sociais. Universidade Portucalense.

Pardar, F. (19 September 2023). The disinformation landscape in Portugal. Retrieved on 10 March 2025, https://www.disinfo.eu/publications/disinformation-landscape-in-portugal/

Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. Proc Natl Acad Sci U S A, 2521-2526. doi:10.1073/pnas.1806781116.

Público. (22 de November de 2022). Conseguir perceber notícias e informação é “tão importante como saber ler e escrever”. Retrieved on 7 March 2025, https://www.publico.pt/2022/11/22/sociedade/noticia/conseguir-perceber-noticias-informacao-tao-importante-saber-ler-escrever-2028567

Ramos, G., Batista, F., Riberiro, R., Fialho, P., Moro, S., Fonseca, A., . . . Silva, C. (2024). Leveraging Transfer Learning for Hate Speech. Portuguese Social Media Posts. IEEE Access., 101374-101389.

RTP. (20 February de 2025). Carta aberta pede "sanção efetiva" para responsáveis por discurso de ódio no Parlamento. Retrieved on 6 March de 2025, de https://www.rtp.pt/noticias/politica/carta-aberta-pede-sancao-efetiva-para-responsaveis-por-discurso-de-odio-no-parlamento_n1635728

Sábado. (n.d.). "Médicos pela Verdade" suspendem a atividade online. Retrieved on 4 de March de 2025, de https://www.sabado.pt/portugal/detalhe/medicos-pela-verdade-suspendem-a-atividade-online

Snow, N., Jowett, G. S., & O'Donnell, V. J. (2005). Propaganda & Persuasion.

Stepney, E. S., & Lally, C. (2024). Disinformation: sources, spread and impact. Parliamentary Office of Science and Technology (POST), 1-31. doi: https://doi.org/10.58248/PN719

Subramanian, M., Sathiskumar, V. E., Deepalakshmi, G., Cho, J., & Manikandan, G. (2023). A survey on hate speech detection and sentiment analysis using machine. Alexandria Engineering Journal, 110-121.

Tanvir, S. (2022). Hate Speech vs Free Speech: The Ongoing Debate and Challenges in the Globalized World. SSRN Electronic Jounal, 1-6.

Tewdwr-Jones, M. (2021). The dangers of digital democracy. Retrieved on 4 March 2025, https://www.ucl.ac.uk/bartlett/bartlett-review/dangers-digital-democracy

Waal, S. V., Stikker, M., Kortlander, M., Eeden, Q. V., Demeyer, T., & Bocconi, S. (2020). European Digital Public Spaces. Waag.

Walker, S. (19 de February de 2019). The Ethics of Content Moderation: Balancing Free Speech and Harm Prevention. Retrieved on 2 March 2025, de https://chekkee.com/the-ethics-of-content-moderation-balancing-free-speech-and-harm-prevention/

Windisch, S., Wiedlitzka, S., & Olaghere, A. (2021). PROTOCOL: Online interventions for reducing hate speech and cyberhate: A systematic review. Campbell Syst Rev., 1-17.

World Wide Web Consortium. (s.d.). Web Content Accessibility Guidelines (WCAG). https://www.w3.org/WAI/standards-guidelines/wcag/

Zacky, E. U., & Zacky-Eze, C. J. (2025). Combatting Disinformation: Analysing Effective Communication Strategies and Tactics for Organizational Resilience. Tripodos, 1-25.

Zhang, X., Yadollahi, M. M., Dadkhah, S., Isah, H., Le, D.-P., & Ghorbani, A. A. (2023). DATA BREACH: ANALYSIS, COUNTERMEASURES ANDCHALLENGES. International Journal of Information and Computer Security, 1-27.

Luxembourg:

Bal, E., & Akcil, U. (2024). The implementation of a sustainable online course for the development of digital citizenship skills in higher education. Sustainability (Switzerland), 16(1), 445.

Bengio, Y. et al. (2025): International AI Safety Report. Department for Science, Innovation and Technology.

Boudana, S., & Segev, E. (2024): Fake news makes the news: Definitions and framing of fake news in mainstream media. Journalism Practice. Advance online publication.

Cho, H., Cannon, J., Lopez, R., & Li, W. (2024). Social media literacy: A conceptual framework. New Media & Society, 26(2), 941-960.

Christakis, Dimitri A./Hale, Lauren (eds.) (2025). Handbook of Children and Screens -

Digital Media, Development, and Well-Being from Birth Through Adolescence

Douglas, K. M., & Sutton, R. M. (2023): What are conspiracy theories? A definitional approach to their correlates, consequences, and communication. Annual Review of Psychology, 74, 271–298.

Gerke, A., & Wegner, C. (2022). Digitale Revolution im Unterricht?! Eine Studie zur Erhebung der Medienkompetenz angehender Lehrkräfte. merz| medien+ erziehung, 66(3), 61-68.

Kim, M., & Choi, D. (2018). Development of youth Digital Citizenship Scale and implication for educational setting. Educational Technology and Society, 21(1), 155-171.

Meigs D., O'Neill B., Soriano A., Tomé V. (2017). Digital Citizenship Education – Vol.1 Overview and new perspectives (Council of Europe)

Meyer, F., Becker, J., & Bock, A. (2023). Medienkompetenz aus bildungswissenschaftlicher, bildungspolitischer und schulpraktischer Perspektive. MedienPädagogik: Zeitschrift für Theorie Und Praxis Der Medienbildung, 2023 (Occasional Papers), 118–138.

Obermaier, M. (2024). Youth on standby? Explaining adolescent and young adult bystanders' intervention against online hate speech. New Media & Society, 26(8), 4785-4807.

Obermaier, Magdalena/Schmuck, Desirée, Youths as targets: factors of online hate speech victimization among adolescents and young adults, Journal of Computer-Mediated Communication, Volume 27, Issue 4, July 2022, zmac012

Peissl, Helmut; Sedlaczek, Andrea: Kritische Medienkompetenz vor dem Hintergrund der Digitalisierung. Media and Information Literacy (MIL) und Critical Media Literacy (CML) im Vergleich - In: Magazin erwachsenenbildung.at (2022) 44/45, 11 pp.

Polanco-Levicán, K., & Salvo-Garrido, S. (2022). Understanding Social Media Literacy: A Systematic Review of the Concept and Its Competences. International Journal of Environmental Research and Public Health, 19(14), 8807.

Richardson, J., Milovidov, E. (2019). Digital Citizenship Education Handbook

Schmuck, D., Fawzi, N., Reinemann, C., & Riesmeyer, C. (2021). Social media use and political cynicism among German youth: the role of information-orientation, exposure to extremist content, and online media literacy. Journal of Children and Media, 16(3), 313–331.

Sponholz, L. (2023): Hate speech. In C. Strippel, S. Paasch-Colberg, M. Emmer, & J. Trebbe (Eds.), Challenges and perspectives of hate speech research (p. 143–163). Digital Communication Research, 12.

Switzerland Klopfenstein Frei, N., Wyss, V., Gnach, A., & Weber, W. (2024). "It's a matter of age": Four dimensions of youths' news consumption. Journalism, 25(1), 100-121.

UNESCO (2024, lead author: Richardson. J). Global Citizenship Education in a Digital Age

Wardle, C. (2024): A conceptual analysis of the overlaps and differences between hate speech, misinformation and disinformation. New York, NY: United Nations Department of Peace Operations. (UN-commissioned report)

Wuyckens, G., Landry, N., & Fastrez, P. (2022). Untangling media literacy, information literacy, and digital literacy: A systematic meta-review of core concepts in media education. Journal of Media Literacy Education, 14(1), 168-182.