

Guidance Note on countering the spread of online mis- and disinformation through fact-checking and platform design solutions in a human rights compliant manner



Adopted by the Steering Committee for Media and Information Society (CDMSI) at its 24th meeting, 29 November-1st December 2023

COUNCIL OF EUROPE



CONSEIL DE L'EUROPE

Guidance Note on countering the spread of online mis- and disinformation through fact-checking and platform design solutions in a human rights compliant manner

adopted by the Steering Committee for Media
and Information Society (CDMSI) at its 24th meeting,
29 November-1st December 2023

French edition:

*Note d'orientation sur la lutte contre
la propagation de la désinformation
et de la désinformation en ligne par
le biais de la vérification des faits et
de la conception de plateformes dans
le respect des droits de l'homme*

Reproduction of the texts in this
publication is authorised provided
the full title and the source, namely
the Council of Europe, are cited. If
they are intended to be used for
commercial purposes or translated
into one of the non-official
languages of the Council of Europe,
please contact publishing@coe.int.

Cover design and layout:
Documents and Publications
Production Department
(SPDP), Council of Europe

Council of Europe
F-67075 Strasbourg Cedex
© Council of Europe, March 2024
Printed at the Council of Europe

Contents

GUIDANCE NOTE	5
RECOMMENDATIONS ON FACT-CHECKING	10
RECOMMENDATIONS ON PLATFORM-DESIGN SOLUTIONS	13
RECOMMENDATIONS ON EMPOWERMENT OF USERS	18
EXPLANATORY MEMORANDUM TO THE GUIDANCE NOTE	23

Guidance Note

on countering the spread of online mis- and disinformation through fact-checking and platform design solutions in a human rights compliant manner

Preamble

1. As recognised by the Committee of Ministers, in Recommendation [CM/Rec\(2022\)4](#) on promoting a favourable environment for quality journalism in the digital age, disinformation undermines trust in the media and threatens the reliability of information that feeds public debate and enables democracy. As noted in Recommendation [CM/Rec\(2022\)12](#) on electoral communication and media coverage of election campaigns, individuals are exposed to various kinds of disinformation online at all levels, making it more challenging to maintain the integrity of elections, ensure healthy pluralism, and protect the democratic process from manipulation. Malicious actors, including some foreign governments, often spread disinformation online to disrupt free and fair elections and undermine the very notion that facts matter to democracy and can be meaningfully identified and discussed. Also, there is a growing amount of disinformation generated and spread with Artificial Intelligence tools that poses distinctive threats to democratic dialogue. False content often spreads at such a large scale that it cannot be tackled solely by human intervention. The quality of public debate is also threatened by the propagation of false information by individuals who consider it as true and share it in good faith. Importantly, however, measures such as support for fact-checking, platform-design solutions and user empowerment may all contribute, at different levels and in different ways, to reducing the spread and negative impacts of online mis- and disinformation. By implementing all these measures coherently and continuously adverse implications for democracy and individuals' safety and well-being may be prevented.

2. Any State regulatory frameworks, including co-regulation, aiming to reduce the spread of false information must comply with the Convention for the Protection of Human Rights and Fundamental Freedoms (ETS No. 5, “the Convention”), particularly the right to freedom of expression under Article 10 of the Convention. Freedom of expression should be seen as the overarching principle when developing policies that may affect a natural or legal person’s ability to contribute to public debate, even though it may need to be balanced against other human rights enshrined in the Convention, such as for instance, the right to respect for private and family life (Article 8), and prohibition of discrimination (Article 14).
3. For the purposes of this Guidance Note:
 - a. “disinformation” refers to verifiably false, inaccurate or misleading information deliberately created and disseminated to cause harm or pursue economic or political gain by deceiving the public, as defined in Recommendation [CM/Rec\(2022\)12](#) on electoral communication and media coverage of election campaigns, and Recommendation [CM/Rec\(2022\)11](#) on principles for media and communication governance;
 - b. “misinformation” refers to verifiably false, inaccurate or misleading information disseminated without an intention to mislead, cause harm, or pursue economic or political gain; users who share misinformation generally believe it to be true;
 - c. “clickbait” refers to the use of sensationalist language that establishes a false connection between some promotional material (such as a headline or social media post) and the content that is being promoted (such as a news article, a web page, or social media profile) with the purpose of attracting attention and encouraging visits;
 - d. “platforms” refer to those providers of digital services that connect users, set the rules for their interactions and make use of algorithmic systems to collect and analyse data and personalise their services, as defined in Recommendation [CM/Rec\(2022\)11](#) on principles for media and communication governance; in the field of communications, such platforms include search engines, news aggregators, video-sharing services, and social networks;
 - e. “platform design” refers to all the key decisions that shape the functioning of a digital platform and the ways in which users experience it; the definition also encompasses the technical means by which platforms implement, maintain, and update their architectures and interfaces;
 - f. “algorithms”, as related to digital platforms, refers to tools that select what content is displayed for users based on relevance and preferences,

which are inferred based on data on users' behaviours collected as a result of their online interactions;

- g. "algorithmic systems" refers to applications that, often using mathematical optimisation techniques, perform one or more tasks such as gathering, combining, cleaning, sorting, classifying and inferring data, as well as selection, prioritisation, the making of recommendations and decision making, as defined in Recommendation [CM/Rec\(2020\)1](#) on the human rights impacts of algorithmic systems, Appendix, paragraph 2;
- h. "artificial intelligence system" refers to any algorithmic system or a combination of such systems that uses computational methods derived from statistics or other mathematical techniques and that generates text, sound, image or other content, or either assists or replaces human decision-making, as defined by the Committee on Artificial Intelligence in [CAI\(2023\)18](#);
- i. "news organisations" refers to organisations that create and distribute news about current affairs and events to their audiences; such organisations can operate under different business models and rely on different combinations of funding sources, including advertising, subscriptions, sales, philanthropic donations, licence fees, and direct and indirect State subsidies;
- j. "fact-checking" refers to the process and techniques employed to verify the accuracy of information available in the public domain, as well as the identities, strategies and practices of subjects producing or disseminating disinformation; fact-checking can be performed by media organisations, professional journalists, specialised fact-checkers, and non-professional users;
- k. "user empowerment" refers to measures aimed at expanding users' understanding, informed choice and control of the impact of digital technologies on their rights, including by promoting media and information literacy, opportunities for exercising user rights, and avenues to collective action;
- l. "media and information literacy" refers to the development of individuals' necessary knowledge and skills to critically engage with media content, to navigate the complex media and information ecosystem and, ultimately, to make their political and other choices in an autonomous manner, as defined in [Resolution on the changing media and information environment - Conference of Ministers responsible for Media and Information Society](#);
- m. "regulation" refers to a set of binding rules developed and enforced by the State to achieve a public policy objective;

- n. “co-regulation” refers to co-operative forms of regulation, jointly developed by the State and private actors to achieve a public policy objective;
 - o. “self-regulation” refers to the process whereby a private actor or sector develop and enforce rules for themselves to achieve an industry or public policy objective.
4. The core purpose and scope of this *Guidance Note* is to provide practical guidance and recommendations to policymakers and stakeholders on countering the dissemination of online mis- and disinformation through fact-checking and platform-design solutions, all in a human-rights compliant manner, and with due regard to user empowerment as an important precondition for addressing the negative impacts of online mis- and disinformation.
 5. To reduce the risks mis- and disinformation pose to the public, this *Guidance Note* focuses in particular on tackling the mechanisms through which its scale and scope can rapidly increase and thus reach vast audiences before they can be countered, as well as on enhancing the prominence and discoverability of accurate and reliable information provided by trustworthy sources.
 6. This *Guidance Note* contains recommendations for member States of the Council of Europe, platforms which have human rights responsibilities of their own (see Committee of Ministers’ Recommendation [CM/Rec\(2016\)3](#) on human rights and business; Recommendation [CM/Rec\(2018\)2](#) on the roles and responsibilities of internet intermediaries ; the United Nations “Guiding Principles on Business and Human Rights: Implementing the United Nations ‘Protect, Respect, and Remedy’ Framework”; and the European Union Strengthened Code of Practice on Disinformation), and other stakeholders engaged in addressing the spread of mis- and disinformation online, in particular news organisations and fact-checkers.
 7. This *Guidance Note* is accompanied by an *Explanatory Memorandum*, which is designed to explain the context, reasoning, and nuances behind the recommendations.
 8. The production, consumption, spread, and effects of false and misleading information in evolving media ecosystems and turbulent societal and political conditions are complex phenomena that cannot be explained, let alone solved, by a single set of factors. This *Guidance Note* is organised around three sets of recommendations (on fact-checking, platform-design solutions, and empowerment of users) that aim to comprehensively address the main causes of these problems. These three sets of recommendations

have been designed to support and encourage cross-border cooperation among all relevant stakeholders (including governments, regulators, industry, journalists, civil society, researchers, and users) and need to be undertaken together to achieve the desired outcomes. Only by pulling all these levers together will democratic societies make concrete progress on tackling these challenges.

Recommendations on fact-checking

9. **Role and centrality:** In developing policies towards countering the spread of online mis- and disinformation, States and platforms should recognise the centrality of fact-checking as a key practice for the health of public debate. States should strengthen fact-checking by creating and supporting conditions for their financial independence, transparent governance, and public service orientation, as well as strengthening the broader ecosystem of organisations and actors that can help counter mis- and disinformation in democratic societies. Platforms should collaborate with fact-checking organisations to ensure that their users are exposed and encouraged to share high-quality information on matters of public debate, and to challenge and debunk mis- and disinformation when they encounter it.
10. **Transparency:** Any measures entailing regulation, co-regulation, and self-regulation should ensure the transparency of the fact-checking process as an essential tool for fact-checkers' independence. Measures enhancing transparency of fact-checking should focus on its funding, internal organisation, professional processes and standards, accountability mechanisms, and quality control measures. To enhance the quality of public debate and policymaking, fact-checking organisations should also develop mechanisms to disseminate publicly relevant aggregate information deriving from their work, for instance on the topics about which mis- and disinformation is spreading most intensely at a given moment or the actors that produce and spread the highest amounts of mis- and disinformation in a certain context.
11. **Independence:** Any form of regulation, co-regulation or self-regulation should guarantee the editorial freedom and independence of news organisations, fact-checking organisations and fact-checkers working in any organisation, for example regarding choice and prioritisation of items selected for fact-checking. Any initiative by public or private actors that involves funding of fact-checking organisations should preserve their independence from any State and government interference, as well as from powerful economic interests. In particular, when platforms provide funding in support of fact-checking initiatives, institutional and procedural safeguards need to be in place that protect fact-checkers' independence and to put them at arm's length from the funders. To protect fact-checking

from funders' interference, public and private actors should facilitate the creation of independent bodies that autonomously allocate funds provided by public and private stakeholders for fact-checking initiatives and organisations. Such allocation should be based on transparent criteria established *ex ante* and measured rigorously according to state-of-the-art academic knowledge and international standards.

12. **Organisational challenges:** Platforms should enhance the integration of (external) fact-checking organisations in their content curation systems, overcoming challenges such as lack of data access for fact-checkers and the changing role of human agency in using Artificial Intelligence systems that flag problematic content. Platforms should develop adequate processes to empower fact-checking by both professional organisations and individuals.
13. **Linguistic challenges:** In their initiatives to support fact-checking, States and platforms should work to remove language-related barriers to the viability and positive impact of fact-checking organisations, such as lack of support for non-English language fact-checking efforts and uncertainties as to the effectiveness for non-English language content of Artificial Intelligence systems trained mainly or solely with English-language datasets.
14. **Sustainable working conditions:** As fact-checking cannot at the moment be reliably automated besides basic tasks, and as fact-checkers perform important journalistic work, States should grant them appropriate level of support and legal guarantees, particularly in terms of contractual protections, public recognition of their professionalism, support for skills training initiatives, and access to data held by public institutions. When platforms rely on fact-checkers, they should guarantee them working conditions and compensations commensurate with international professional standards.
15. **Funding:** As part of their broader efforts to ensure a healthy, well-funded and safe environment for media, news organisations and media professionals, States should promote adequate conditions that ensure national and international fact-checking organisations can access adequate, stable, and sustainable funding from diverse and reliable sources, following the principles outlined in Recommendation [CM/Rec\(2022\)4](#) on promoting a favourable environment for quality journalism in the digital age. Platforms should identify means by which they can transparently support independent fact-checking, for instance through the establishment and funding of third-party national and international fact-checking networks

or the promotion of transparency in internal verification responses. Irrespective of its source (public or private) and nature (direct or indirect), funding to fact-checking organisations should be provided and distributed based on institutional and procedural safeguards that guarantee their independence, following the provisions in paragraph 11 above.

16. **Pluralism:** In accordance with the principles for media and communication governance outlined in Recommendation [CM/Rec\(2022\)11](#) on principles for media and communication governance, any direct or indirect State funding for fact-checking, including subsidies, should be geared towards enhancing media pluralism and the promotion of media content reflecting societal diversity, including gender and ethnic diversity, while protecting media freedom and ensuring functioning markets.
17. **User trust:** Users' trust in fact-checking as performed by different types of organisations and actors is fundamental to ensure the societal effectiveness of these endeavours. Thus, any regulation, co-regulation, or self-regulation initiative should enshrine the principles of independence, and transparency, as outlined above as well as the principle of quality control as described below.
18. **Quality control:** Fact-checkers should develop and implement robust and transparent quality control processes, based on international professional standards and best practices developed in collaboration with industry representatives, civil society organisations, and researchers. Fact-checkers should publicly disclose their quality control procedures and publish periodical reports on the outcome of their quality control assessments.

Recommendations on platform-design solutions

19. **Designing healthy information environments:** States and platforms should work collaboratively to remove or reduce the obstacles to well-informed public debate that derive from platforms' design choices. Platforms should recognise their potential impact on the quality of public debate exceeds that of most other actors, and promote the conditions for healthy and constructive public debate. In particular platforms should create adequate processes to assess and mitigate the risks that their content prioritisation algorithms promote mis- and disinformation through clickbait and other mechanisms.
20. **Human-rights-by-design:** Online platforms in democratic societies should be designed to promote and protect human rights and fundamental freedoms, as guaranteed under the Convention. When States consider regulation or co-regulation that may impact platform design in ways that may affect human rights, they should first conduct and publish human rights impact assessments. If such impact assessments conclude that the proposed interventions pose risks to human rights, including those of vulnerable groups, they should also include concrete measures to prevent or mitigate such risks, which States should then be bound to implement before introducing regulation or co-regulation. By the same token, when platforms consider design changes and self-regulation that may affect human rights, including those of vulnerable groups, they should first conduct and publish human rights impact assessments. If such impact assessments conclude that the proposed interventions pose risks to human rights, they should also include concrete measures to prevent or mitigate such risks, which platforms should be bound to implement before introducing design changes and self-regulation.
21. **Safety-by-design:** Platforms should incorporate safety in all their key design decisions, especially where they relate to Artificial Intelligence systems. This includes creating a safe, inclusive and favourable online environment for participation in public debate that enables all persons concerned, especially women and girls and other vulnerable groups as listed in paragraph 36 below, to enjoy freedom of expression and the right to access to information. Platforms' policies should be particularly

responsive to the risks entailed when Artificial Intelligence systems are used to produce and distribute false information at a scale that surpasses the capacity of any individuals or groups to defend themselves against its potential harms.

22. **Investment in non-English translation and content moderation tools:** Platforms should invest in non-English moderators with in-depth understanding of different cultural contexts and enhance their automated content moderation tools so that they function at similar levels of efficacy across different languages. Moderators play a critical role in ensuring that content in languages other than English is adequately assessed and moderated. Likewise, automated systems should be designed to comprehend local contexts, discern delicate nuances, and differentiate between various dialects.
23. **Proportionality:** Any State regulatory frameworks, including co-regulation, targeting platform design in relation to mis- and disinformation, should contain measures and requirements that are proportional to the risk level that platforms' functioning involves (e.g., the adverse consequences for human rights that some of their users are likely to entail). The criteria based upon which the risk levels of different platforms are assessed may include factors such as the size (e.g., number of users and capitalisation), resources (e.g., technical and economic means), and reach (e.g., potential impact on audiences). Such criteria should be specified clearly, reviewed periodically, measured precisely in collaboration with independent researchers, and communicated transparently by an independent regulator with sufficient expertise and adequate resources.
24. **Focus on processes:** To protect freedom of expression, co-regulation and regulation targeting the design of platforms to reduce mis- and disinformation should not focus directly on specific pieces of content. Any co-regulation or regulation should primarily address the processes through which platforms limit the activities of accounts (including suspension and removal) and the spread of content deemed to contain mis- and disinformation, as well as the processes through which platforms rank, moderate, and remove content.
25. **Consideration of tactics of actors:** Any form of regulation, co-regulation or self-regulation dealing with mis- and disinformation should take into consideration the fact that actors that deliberately disseminate disinformation, such as covert foreign operations and individuals or groups that disguise their identity to deceive the public, employ a variety of complex

tactics to mislead the public that often entail combinations of accurate and inaccurate content.

26. **Granular and differentiated response:** Any co-regulation or regulation targeting the design of platforms in relation to mis- and disinformation should treat content removal as a last resort, as platforms can (and should) employ various content moderation techniques that have less severe effects on freedom of expression, such as for instance prioritisation and de-prioritisation, promotion and demotion, monetisation and demonetisation, and the provision of supplementary information to users, including debunking, age-related alerts, “trigger warnings”, and additional content from official and independent authoritative sources such as professional news organisations and Public Service Media. Following the principle highlighted in paragraph 23, any regulation or co-regulation should avoid dictating what specific responses platforms should adopt to combat different forms of mis- and disinformation and focus primarily on the processes through which platforms address these challenges.
27. **Prioritisation of professional news sources and public interest content:** As part of their policies on countering the spread of online mis- and disinformation, States and platforms may, where necessary, and consistent with the Convention, introduce appropriate and proportionate measures to ensure prominence of public interest content online, particularly content produced by reliable and professional news organisations, in line with the principles set out in the Guidance Note on the Prioritisation of Public Interest Content Online ([CDMSI\(2021\)009](#)). Importantly, States and platforms should establish clear, transparent, and independently validated criteria and processes for defining what types of sources can be considered reliable professional news organisations and what types of content are of public interest. In setting these standards, States and platforms should follow and build on existing best practices. States should refrain from obliging platforms to carry or prioritise any specific pieces of content or information.
28. **Transparency:** Platforms should provide the highest possible levels of transparency on the design of their services, implementation of their terms of service, and their key policies relating to mis- and disinformation, such as information regarding identification and removal of inauthentic accounts and bots, content removal, recommendation, amplification, promotion, downranking, monetisation, and distribution. States should create the most favourable conditions and incentives for platforms to generate and publish all data required to conduct any analyses necessary to guarantee meaningful transparency on how their policies, including

their implementation, enable or constrain the dissemination of mis- and disinformation. These transparency requirements should be proportionate to the risk level of different platforms, as set out in paragraph 23 above. Platforms should publish the necessary information in machine-readable format to ensure transparency of their policies at different levels and in pursuit of different goals, including empowering users, enabling third-party auditing and oversight, and informing independent researchers, news organisations and other actors involved in efforts to counter disinformation.

29. **Accountability:** As discussed in Section 4 of Recommendation [CM/Rec\(2022\)13](#) on the impacts of digital technologies on freedom of expression, States should ensure that any user whose freedom of expression is limited as a result of regulation, co-regulation, or self-regulation aimed at reducing the spread of mis- and disinformation can access effective redress mechanisms against these restrictions in a simple, accessible and affordable way. When platforms enforce any restrictions on freedom of expression aimed at reducing the spread of mis- and disinformation, they should provide users directly or indirectly affected by such restrictions with clear information on the regulations under which their rights have been limited. Platforms should also provide timely and effective redress mechanisms that allow the affected individuals to submit an appeal without undue costs, delays, or difficulties (see Recommendation [CM/Rec\(2018\)2](#) on the roles and responsibilities of internet intermediaries).
30. **Support for independent research:** To reliably assess the effectiveness of platform-design solutions to mis- and disinformation, States should develop regulation or co-regulation to ensure that independent researchers and investigative journalists can access adequate individual-level data on users' online behaviours generated and collected by platforms, as well as data concerning advertisements and search results from search engines and app stores. All these initiatives should ensure researchers can access data in ways that are secure, legal, and compliant with all regulation safeguarding privacy and data protection in the light of the [Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data \(CETS No. 108\)](#). States and private actors should follow the provisions contained in Section 6 of Recommendation [CM/Rec\(2022\)13](#) on the impacts of digital technologies on freedom of expression, which provides detailed indications on how to support independent research on digital technologies and freedom of expression. States should also create suitable conditions so that independent researchers on mis- and disinformation can rely on adequate, stable, and sustainable funding

from both public and private sources, and clarify that researchers who have received approval from an ethical review board or equivalent body cannot be held liable on the grounds that they breached the internet intermediaries' terms of service while scraping publicly available data from platforms.

Recommendations on empowerment of users

31. **Building a healthy media ecosystem:** Any initiatives aimed at empowering users to protect themselves and others against the spread of mis- and disinformation require a steady and abundant supply of quality information by recognised trustworthy sources to counterbalance and verify false and misleading information. As achieving this outcome first and foremost requires supporting and enhancing quality journalism, States should promote regulation and co-regulation, as well as support self-regulation, to implement the principles outlined in Recommendation [CM/Rec\(2022\)4](#) on promoting a favourable environment for quality journalism in the digital age. In addition, States should also ensure the accessibility of information held by their public authorities in line with the requirements of Recommendation Rec(2002)2 on access to official documents and the [Council of Europe Convention on Access to Official Documents \(CETS No.205 – the Tromsø Convention\)](#).
32. **Building resilience:** In developing their policies and/or regulatory frameworks on countering the spread of online mis- and disinformation, States and platforms should aim to enhance users' ability to recognize reliable sources and resilience to online mis- and disinformation, i.e., to accurately identify, critically evaluate, and challenge mis- and disinformation, for instance by fact-checking, reporting, and correcting it on social media platforms, as well as strengthen users' capacity for actions in support of healthy media ecosystems and against online harms.
33. **Building on existing frameworks:** Any regulatory, co-regulatory, or self-regulatory initiatives should start from, and seek to further develop, existing tools for user empowerment, such as ongoing transparency initiatives and user-centric content moderation models on social media platforms, ongoing empowerment programmes by States to support users in exercising their rights online, community rules adopted by platforms, open-source-based software tools, Artificial Intelligence systems to identify trustworthy sources, public value recommender systems, and published research on how to enhance users' media and information literacy.

34. **Building bridges with communities:** Empowering users requires understanding different communities' needs, challenges, and aspirations when it comes to encountering, producing, and disseminating information. States and private stakeholders should promote policies and initiatives that enhance capacity for collective action so that users and communities can balance the power of well-resourced organised actors and the large global corporations that own digital platforms to promote reforms of the media ecosystem so that it is more conducive to healthy public debate based on trustworthy information.
35. **Promotion of user rights:** Platforms should develop user-friendly interfaces and intuitive ways to provide users with accessible information about their rights and easily actionable mechanisms to protect themselves against threats to such rights. Any users whose human rights may be harmed by encountering or being accused to spread mis- or disinformation should have ample and effective opportunities to report complaints and obtain redress. Platforms should ensure that any platform interventions devised to limit the spread of false information do not limit the rights of independent professional news organisations to distribute reliable content, as well as provide such news organisations with rapid and effective mechanisms to obtain redress.
36. **Protection of vulnerable groups:** States, civil society organisations and platforms should develop initiatives to protect the rights of vulnerable individuals and groups targeted by particularly harmful types of mis- and disinformation, for instance inaccurate content that discriminates or incites hatred based on gender, sexual orientation, ethnicity, religion, disabilities, and other protected characteristics. Special attention should be paid to the protection of minors. Platforms should provide access to fast-tracked mechanisms to allow representatives from these groups to register their complaints and obtain redress. Given the high risks they incur in performing their roles, journalists and human rights advocates, particularly those coming from vulnerable groups and those involved in sensitive investigations, should also benefit from enhanced protections and expedite means of redress.
37. **Trusted flaggers:** Platforms should collaborate with independent expert organisations specialised in protecting and representing vulnerable groups to develop transparent mechanisms to reliably identify, recruit, and empower "trusted flaggers" from vulnerable users and human rights advocates, so their warnings and complaints are duly prioritised. Platforms should develop inclusive mechanisms to recruit trusted flaggers from a variety of backgrounds and constituencies, and to effectively onboard

and continuously support them so that they reliably and consistently apply platform policies. Platforms should also develop processes that empower their trusted flaggers to provide feedback and suggest reforms to existing policies to better incorporate the insights and experiences of specific vulnerable groups. Platforms should also work with independent professional news organisations and develop processes through which they can be integrated as trusted flaggers while safeguarding their independence from the platforms.

38. **Collaboration:** States, civil society, platforms, public service media, news organisations, fact-checkers, civil society organisations, user communities, and researchers should collaborate to develop and implement wide-ranging measures to enhance user empowerment. To ensure different actors can be fully engaged in co-designing them, these collaborative initiatives should be adequately funded, organised according to the needs and priorities of the actors involved, and informed by the principles of transparency, accountability, clarity of roles, and realistic goals. Public institutions should refrain from dictating the terms or goals of these collaborations, but should employ their convening power to bring stakeholders together, support collaborative research, and guarantee fairness, accountability, and transparency in the process. States, civil society actors, media organisations and educational institutions should also collaborate in creating a public informational resource where extensive, standardised, and up-to-date data on any initiatives aiming to promote user empowerment are available to the public and to the organisations aiming to contribute to these goals.
39. **Digital tools for user empowerment:** States and private stakeholders should work independently and collaboratively to develop digital tools that empower users to protect themselves against mis- and disinformation. These initiatives may entail the provision of digital tools, including public value recommender systems, educational games, and Artificial Intelligence systems for information verification, and adequate training for both educators and end users. In promoting these tools among the public, developers should ensure that their purposes and modes of usage are presented in an accessible language so they can be adopted as widely as possible, particularly among vulnerable groups. Platforms should ensure interoperability and adequately inform users of the characteristics and benefits of different tools so users have freedom to choose among them. Platforms should also collaborate with the developers of user empowerment tools so they can access relevant platform data to assess their effectiveness and enhance their functioning.

40. **Comprehensive media and information literacy strategies:** States, civil society, Public Service Media, and platforms should support initiatives to develop media and information literacy among all individuals, with a particular emphasis on economically, socially, and technically disadvantaged groups, individuals with special educational needs, and individuals with limited access to quality information, as discussed in Section 5 of the Recommendation [CM/Rec\(2022\)13](#) on the impacts of digital technologies on freedom of expression. Media and information literacy measures should consider that different population groups have varying degrees of vulnerability to different types of mis- and disinformation, and tailor empowerment initiatives to address the specific needs of different vulnerable groups. States should directly and indirectly support charities and civil society groups that assist in these endeavours, including by providing and equipping public spaces such as libraries, schools, and community centres to support vulnerable groups that may be harder to reach with online-only initiatives.
41. **Education reform:** To enhance and consolidate media and information literacy in the long run, States should reform their educational curricula to include key skills that enable users to protect themselves against mis- and disinformation, as well as to strengthen awareness of their rights. States should ensure that schools are adequately equipped for teaching these subjects and that educators are continuously trained so their knowledge of the latest technological developments and the scientific evidence about them is up to date. In reforming educational curricula and training educators, special attention should be paid to individuals with special learning needs, who are particularly vulnerable to the threats posed by the spread of mis- and disinformation.
42. **Forward thinking:** In developing their policies on empowering users against online mis- and disinformation, States should envisage a mix of shorter-term and longer-term measures, taking into account that user resilience entails both day-to-day behaviours that respond to contextual nudges and incentives that may be facilitated by digital tools (see paragraph 38 above), and a set of abilities and predispositions, such as discerning scepticism and critical thinking, that need to be developed and reinforced during the lifelong educational trajectory of a person through media and information literacy (see paragraph 39 above).
43. **Diversification and evidence-based evaluation of empowerment initiatives:** Any measures aimed at empowering users should be based on a multi-dimensional approach that goes beyond traditional awareness-raising campaigns. States should enhance individuals' knowledge about

how to verify information and support the development of specific skills and abilities they can use when encountering new information. Advice and support on how to counter mis- and disinformation should differentiate between public or semi-public platforms, where users interact with more distant connections, and private or semi-private platforms, where users are connected by kinship and friendship and challenging others for sharing inaccurate information may jeopardise intimate personal relationships. States should also promote, fund, and provide guidance on how to conduct and evaluate more stringent, evidence-based, and continuous quality assessments on existing media and information literacy programmes, and share the results of these evaluations with other States and private stakeholders to ensure that similar initiatives benefit from the knowledge acquired in evaluating existing programmes.

Explanatory Memorandum to the Guidance Note

on countering the spread of online
mis- and disinformation through fact-
checking and platform design solutions
in a human rights compliant manner

1. Introduction

Mis- and disinformation has emerged as an important concern in democracies worldwide, becoming an integral part of the digital era. Individuals now navigate a fragmented, interconnected, and complex media landscape that extends beyond traditional mass media to include numerous social media platforms and information channels. Understanding the prevalence and impact of disinformation on democracies has become increasingly crucial for observers of this intricate information environment.

The widely-held assumption about disinformation is that it exists at alarming levels, permeating all layers of society and democratic governance. However, comprehensive empirical data on its prevalence is scarce. The UN Special Rapporteur on freedom of expression recently highlighted that “[a]lthough empirical research suggests that only a small proportion of people are exposed to disinformation, the impacts on institutions, communities and individuals are real, broad and legitimate”,¹ drawing upon submissions from academic institutions on empirical research.² While specialised reports continue to emphasise the threats posed by disinformation in specific areas, such as Russian interference in information dissemination regarding the

-
1. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, “[Disinformation and freedom of opinion and expression](#)”, (A/HRC/47/25), 13 April 2021, paragraph 22.
 2. See, for example, NYU Center for Social Media and Politics [submission](#); and Reuters Institute for the Study of Journalism, University of Oxford [submission](#).

Ukraine invasion since 2022,³ or mis- and disinformation levels in media reports about the climate crisis and the COVID-19 pandemic,⁴ the overall empirical landscape remains complex. Academic research has also explored the secondary effects of mis- and disinformation, uncovering a discrepancy between its perceived prevalence and the severity of the problems it creates globally. It reveals that individuals' fear of exposure to mis- and disinformation can be just as powerful as false information itself, fuelling apprehension, anger, and polarisation of opinions.⁵

Another crucial finding in mis- and disinformation research is that it is a moving target. The technology enabling the creation and dissemination of false information is evolving rapidly, necessitating continuous attention to recent developments such as the role of generative AI in producing text, images, and videos. Key technologies like deepfakes and ChatGPT highlight the need for constant updates and vigilance in understanding disinformation. Moreover, the weaponisation of disinformation by malicious actors is a global process that demands attention because of its risks for democracy.

The work of fact-checkers plays a pivotal role in the global fight against mis- and disinformation, both through the verification of information and the emergence of a professional community dedicated to this endeavour. Fact-checking is both a process and a profession, and it has been studied as such extensively during recent years. Further, in addition to fact-checking, platform-design solutions can also contribute to reducing the spread and negative impacts of online mis- and disinformation. Indeed, platform design solutions can promote the conditions for a healthy and constructive public debate, and ensure a favourable and enabling environment for freedom of expression, which are essential to tackling disinformation. Users are a third key component in building resilience against the spread of mis- and disinformation. Empowering them is critical: while fact-checking can contribute to debunking and platform design solutions can help to reduce the reach of mis- and disinformation, users will always be exposed to some kind of false content and must be equipped to meet this challenge.

-
3. European Union External Action Service (2023), 1st EEAS report on foreign information manipulation and interference threats: Towards a framework for networked defence. Available at: <https://euvsdisinfo.eu/uploads/2023/02/EEAS-ThreatReport-February2023-02.pdf>, accessed 26 September 2023.
 4. First Draft research; available at: <https://firstdraftnews.org/research/>, accessed 26 September 2023.
 5. Lecheler, S. & Egelhofer, J. L. (2022), "Disinformation, misinformation, and fake news: understanding the supply side", in Strömbäck J. et al. (Eds.), *Knowledge resistance in high-choice information environments*, Routledge, pp. 69-87.

Finally, it should be noted that public, academic, and regulatory debates often rely on the distinction between misinformation and disinformation. “Misinformation” can refer to false, inaccurate or misleading information disseminated without an intention to mislead, cause harm, or pursue economic or political gain; and users who share misinformation generally believe it to be true.⁶ “Disinformation” refers to verifiably false, inaccurate or misleading information deliberately created and disseminated to cause harm or pursue economic or political gain by deceiving the public, as defined by the Council of Europe’s Committee of Ministers.⁷ However, the distinction between misinformation and disinformation is difficult to apply in practice, because it is often impossible to ascertain the intention of those who created and disseminated some content and the same content may be spread with different intentions by different individuals and groups.⁸ Ultimately, both forms contribute to the distribution of false information, which weakens the information ecosystem and endangers fundamental values. It is therefore crucial to develop mechanisms to counter both the intentional and unintentional dissemination of false information more effectively. For these reasons, it is suggested that further reflection, research and guidance would be useful from the Council of Europe on this matter.

This Explanatory Memorandum provides a concise insight into the principles underlying the Guidance Note, addressing its three central pillars: the role of fact-checkers, platform design, and empowerment of users. By drawing

-
6. See also Wardle, C., Derakhshan, H. (2017), [Information Disorder: Toward an interdisciplinary framework for research and policymaking](#), Council of Europe report DGI(2017)09, p. 20 (which defines misinformation as “[i]nformation that is false, but not created with the intention of causing harm”).
 7. See [Recommendation CM/Rec\(2022\)11 of the Committee of Ministers to member States on principles for media and communication governance](#), appendix, paragraph 4; and [Recommendation CM/Rec\(2022\)12 of the Committee of Ministers to member States on electoral communication and media coverage of election campaigns](#) appendix, paragraph 4.
 8. For critical discussions on the definitions of mis- and disinformation, see, for example, Ó Fathaigh, R., Helberger, N., & Appleman N. (2021), “The perils of legally defining disinformation”, 10(4) *Internet Policy Review* 1; Lecheler, S., & Egelhofer, J. L. (2022), “Disinformation, misinformation, and fake news: understanding the supply side”, in Strömbäck J. et al. (Eds.), *Knowledge resistance in high-choice information environments*, Routledge, pp. 69-87; Pérez-Escobar, M., Lilleker, D., & Tapia-Frade, A. (2023), “A systematic literature review of the phenomenon of disinformation and misinformation”, *Media and Communication*, 11(2), pp. 76-87; Vraga, E. K., & Bode, L. (2020), “Defining misinformation and understanding its bounded nature: Using expertise and evidence for describing misinformation”, *Political Communication*, 37(1), pp. 136-144; Freelon, D., & Wells, C. (2020), “Disinformation as Political Communication”, *Political Communication*, 37(2); Hameleers, M., Brosius, A., Marquart, F., Goldberg, A. C., van Elsas, E., & de Vreese, C. H. (2022), “Mistake or Manipulation? Conceptualizing Perceived Mis- and Disinformation among News Consumers in 10 European Countries”, *Communication Research*, 49(7), pp. 919–941.

on scientific research, human rights standards, and other relevant sources, it further elaborates on these pillars, offering valuable standards and evidence in the fight against disinformation.

2. Fact-checking

This first section of the Explanatory Memorandum aims to provide an understanding of the emergence of fact-checking and its significance within democratic media ecosystems. It seeks to contextualise the impact that fact-checking can have on these ecosystems, highlighting its role and limits in promoting accuracy, accountability, and informed public discourse. Furthermore, the section aims to present examples of best practices and common standards that are currently being developed within fact-checking networks. These practices and standards serve as guidelines for ensuring the credibility and effectiveness of fact-checking initiatives, fostering transparency, and enhancing public trust in the information ecosystem.

a. Current situation: the practice and the profession of fact-checking

In an era characterised by the rapid dissemination of information across various platforms and channels, the importance of accurate and reliable information has become increasingly crucial. Fact-checking has emerged as a prominent (a) practice and (b) profession in this era, assuming an important role in the modern information landscape. However, it is important to recognize that both aspects of fact-checking are not solely linked to the “disinformation age” but also draw upon decades of journalistic culture and the application of scientific methods of verification during the news production process. This means that any efforts in addressing current and future challenges related to mis- and disinformation can build on existing good practices in this area, apply it in more comprehensive ways, and develop it further in light of developing technological realities.

First, fact-checking can be understood as a combination of norms, routines, and behaviours aimed at verifying the accuracy of information in the digital era.⁹ This process involves not only professional fact-checkers, but also other communities engaged with these goals and other technological means, such

9. Graves, L. (2016), *Deciding what's true: The rise of political fact-checking in American journalism*, Columbia University Press. See also, Silverman, C. (2015), *Lies, Damn Lies, and Viral Content: How News Websites Spread (and Debunk) Online Rumors, unverified, Claims, and Misinformation*, Tow Center for Digital Journalism; and Guerrini, F. (2013), “From traditional to online fact-checking”, Oxford Magazine, Eighth Week, Trinity Term.

as fact-checking tools that utilise natural language processing,¹⁰ all with the goal of verifying information. Fact-checking serves as a powerful toolbox in combating the dissemination of mis- and disinformation and aims to enhance trust in news and public communication. Conceptually, fact-checking encompasses the utilisation of journalistic routines and norms related to verification, as well as scientific methods to conduct its analysis.¹¹ Going beyond the abstract objective of correcting false information, the practice of fact-checking today involves the “regular assessments about the validity of questionable claims made by a wide array of sources, including governments, politicians, institutions, news organisations, and social media users”.¹² It encompasses not only the verification process but also the publication of corrected information and the identification of actors responsible for disinformation, reaching a broader public audience.

Importantly, fact-checking today encompasses a broad range of techniques employed to verify the accuracy of information in the public domain, as well as the identities, strategies and practices of those responsible for producing or disseminating disinformation. It has evolved into a constantly developing practice and profession that reflects how information is managed within societies, shedding light on the individuals or entities responsible for safeguarding information throughout the political process. Fact-checking is not limited to the news production process alone but extends to information already present in public discourse. Fact-checkers serve as guardians of information integrity, both before and after its dissemination. This form of fact-checking involves collaboration among journalists, communication specialists, researchers and fact-checking organisations to combat mis- and disinformation. Furthermore, professional fact-checking is increasingly undertaken by governmental institutions, as well as in collaborative efforts between fact-checking organisations and governments. This state-driven fact-checking, however, must be evaluated with caution as it brings with it potential threats to the independence of media actors in democracies.

Secondly, fact-checking is a journalistic profession. Dedicated fact-checkers have been working in newsrooms worldwide for many years, particularly in the US. These fact-checkers are responsible for verifying the accuracy of

-
10. Zeng, X., Abumansour, A. S., & Zubiaga, A. (2021), “Automated fact-checking: A survey”, *Language and Linguistics Compass*, 15(10), available at : <https://compass.onlinelibrary.wiley.com/doi/10.1111/lnc3.12438>, accessed 28 September 2023.
 11. Mena, P. (2019), “Principles and boundaries of fact-checking: Journalists’ perceptions”, *Journalism Practice*, 13(6), pp. 657-672.
 12. Vinhas, O., & Bastos, M. (2022), “Fact-Checking Misinformation: Eight Notes on Consensus Reality”, *Journalism Studies*, 23(4), pp. 448-468, particularly p. 449.

journalists' work, double-checking important details, while leaving the reporting to others. Their primary focus is to prevent the publication of inaccurate information within news outlets. The increasing spread of and concern for mis- and disinformation resulted in a substantial growth of the fact-checking profession and an expansion of its focus. Large news organisations and public broadcasters are increasingly establishing specialised disinformation and fact-checking news desks that utilise digital tools to verify information – sometimes through cooperative trans-border efforts. These desks conduct fact-checking efforts focusing on “ex post” verification of information and reporting that is already present in the public domain.

The emergence of fact-checking as a profession, however, is most evident in the emergence of fact-checking organisations or independent fact-checking projects, meaning organisations that have fact-checking as their main scope but are not linked to any traditional media organisation. This type of organisation has developed into its own professional ecosystem. Fact-checking organisations and their websites and/or tools dedicated to fact-checking often also employ journalists but have different values and norms than traditional news outlets. Some fact-checking organisations are project-based, emerging from universities, think tanks, governmental organisations, and other stakeholders to address specific disinformation challenges.¹³ Graves and Cherubini (2016) summarise the mission of professional fact-checkers into three role identities. First, there is a prominent group of “reporters,” those who self-identify as journalists and see fact-checking as part of information services. “Reformers” conceptualise fact-checking as activism and connect their activities to political and policy change. “Experts” are those that “place a particular emphasis on their own domain expertise or distinctive methodology, positioning themselves as something like a think tank rather than as journalists or campaigners.”¹⁴

b. Key challenges to fact-checking in the new digital information environment

Fact-checking can thus be conceived as both a practice and a profession, and both aspects have an impact on the information environment in which they operate. However, the exact nature of this impact is still debated in the current research literature, highlighting a number of key challenges.

13. Graves, L. (2018), “Boundaries Not Drawn”, *Journalism Studies*, 19:5, pp. 613-631.

14. Graves, L., & Cherubini, F. (2016), “The Rise of Fact-Checking Sites in Europe”, in *Digital News Project Report (Reuters Institute Digital News Report)*, Reuters Institute for the Study of Journalism, p. 17.

The practice of fact-checking has long been predominantly associated with work carried out in newsrooms. Verification techniques based on journalistic professionalism primarily focus on investigating the sourcing of information and cross-validating sources across different information channels. Traditionally, these channels involved verifying information through personal networks and investigating other sources within the media elite. For instance, a political journalist who receives new information about current political developments in a specific country would typically attempt to verify this information through a network of sources within government institutions and affiliated organisations. Another step would involve checking whether and how other media outlets or news agencies, both domestically and internationally, have reported on this information. Furthermore, when publishing a new story, it would often undergo fact-checking by newsroom editors and fact-checkers to ensure its accuracy.¹⁵ Today, these traditional techniques and processes face challenges due to several changes in the journalistic ecosystem.

First, there is a shift in journalism towards generalism, which assumes that journalists themselves possess the skills to handle the entire production process, from research to publication. As a result, much of the responsibility for fact-checking within newsrooms falls on individual journalists rather than dedicated fact-checking departments. This increased responsibility and pressure limits the time an individual journalists can spend on fact-checking practices such as verifying their sources for a news item - putting them at risk of overlooking potentially false content. All these challenges come at a time when the working conditions of journalists are steadily and substantially deteriorating, due to the crisis of news organisations' established business models,¹⁶ which affect fact-checkers as well.

The increased range of activities individual journalists are responsible for during news production is exacerbated by an increased digitalisation of both news production and fact-checking. In a digital age, sources are not anymore only the traditional news agencies or institutional press offices, but contain a myriad of websites, personal social media pages and complex data sources. This means that, nowadays, fact-checking must involve verification through complex digital technologies, including search engines, AI tools, and dedicated fact-checking applications. The integration of digital information into the journalistic research process has presented new challenges to traditional verification skills that are still taught to aspiring journalists worldwide. The complexity and algorithmic

15. Van Leuven et al. (2018), "Online and newsworthy: Have online sources changed journalism?", *Digital Journalism*, 6(7), pp. 798-806.

16. Recommendation CM/Rec(2022)4 on promoting a favourable environment for quality journalism in the digital age.

nature of these technologies have posed difficulties for journalists. Experimental studies indicate that fact-checking in newsrooms continues to heavily rely on traditional methods, such as telephone conversations, with only specialised journalists like data journalists and investigative journalists consistently and proficiently utilising advanced technological tools for verification processes.¹⁷ Consequently, fact-checking in newsroom settings may need innovative forms of collaboration. Journalists and newsrooms are increasingly partnering with other news organisations, NGO-based fact-checkers, data scientists, and social media platforms during the research process.

As mentioned above, the practice of fact-checking is increasingly evolving, now including a range of new non-profit organisations aimed at verifying information available in the public domain. Undoubtedly, these organisations have led to a further professionalisation of fact-checking techniques within a growing sector. Just like their colleagues in journalistic organisations, professional fact-checkers working in fact-checking organisations employ systematic and rigorous methods of verification while also possessing skills in effectively communicating their findings to the public. Although some of these organisations have experimented with innovative digital tools, the evidence suggests that most still rely on traditional verification techniques. Some studies indicate that fact-checking organisations are based on manual verification methods, especially in cases where larger tools or automated approaches are not available due to language limitations or financial constraints¹⁸. This means that fact-checking organisations may be vulnerable to the same challenges described above in the context of journalism, including working conditions.

Advancing fact-checking organisations is, however, crucial: fact-checking organisations are an alternative to traditional journalism in an era where citizens have direct access to information and political communication without relying solely on traditional media filters. This is particularly the case in Europe, whereas newsroom-based fact-checking is still more commonly practised in the United States.¹⁹ In many countries, particularly in Eastern Europe, fact-

-
17. Lecheler, S., & Kruike-meier, S. (2016), "Re-evaluating journalistic routines in a digital age: A review of research on the use of online sources", *New media & society*, 18(1), pp. 156-171; de Haan, Y., van den Berg, E., Goutier, N., Kruike-meier, S., & Lecheler, S. (2022), "Invisible Friend or Foe? How Journalists Use and Perceive Algorithmic-Driven Tools in Their Research Process", *Digital Journalism*, 10(10), pp. 1775-1793.
 18. Hrcckova, A., Moro, R., Srba, I., Simko, J., & Bielikova, M. (2022), "Automated, not Automatic: Needs and Practices in European Fact-checking Organizations as a basis for Designing Human-centered AI Systems", *arXiv preprint arXiv:2211.12143*.
 19. Vinhas, O., & Bastos, M. (2022), "Fact-Checking Misinformation: Eight Notes on Consensus Reality", *Journalism Studies*, 23(4), pp. 448-468.

checking organisations, aim to enhance the accountability of politicians and media entities external to the traditional media system.²⁰

As mentioned previously, both newsroom-based and other fact-checkers are increasingly engaging with a third group of actors, which includes social media platforms and other technology companies. Since a significant amount of information and discussions occur on social media platforms worldwide, fact-checking has become a prevalent practice within these spaces. Some platforms have increased their content moderation efforts or established partnerships with external fact-checking organisations. In the realm of platforms, fact-checking is extensively discussed in the growing literature on platform content moderation, specifically regarding the effectiveness and ethical considerations of human-led interventions versus machine-led interventions on social media platforms. While platforms have tended to leave fact-checking to individual users, they have mostly neglected to provide the infrastructure and tools that would make these individual activities easier and more scalable.²¹ Moreover, this means that boundaries between users and professional fact-checkers become blurred, which poses its own set of challenges.

Indeed, there is a growing interest in generating data to assess the effectiveness of professionalised fact-checking methods in combating mis- and disinformation. As mentioned earlier, research indicates that traditional journalistic verification techniques face challenges in the complex digital information environments. When it comes to fact-checking public information, meta-analytical evidence suggests that fact-checking methods can be effective, but their success heavily relies on individual and contextual differences.²² For example, the acceptance of fact-checked results varies depending on individuals' ideological backgrounds. Individuals may be unwilling to accept that a particular piece of information is false, if it contradicts their own political ideology.²³ In addition to individual factors such as ideology and personality traits, the acceptance of corrected mis- and disinformation also depends on

20. See Graves, L. & Cherubini, F. (2016), *The Rise of Fact-Checking Sites in Europe*, Reuters Institute for the Study of Journalism, p.10.

21. One example of a major platform providing such infrastructure are the X (formerly Twitter) Community Notes, which provide users with the opportunity to add contextual information to content posted by others. Available at: <https://communitynotes.twitter.com/guide/it/about/introduction>, accessed 28 September 2023.

22. Oeldorf-Hirsch, A., Schmierbach, M., Appelman, A., & Boyle, M. P. (2023), "The influence of fact-checking is disputed! The role of party identification in processing and sharing fact-checked social media posts", *American Behavioral Scientist*, 00027642231174335; Nieminen, S., & Rapeli, L. (2019), "Fighting misperceptions and doubting journalists' objectivity: A review of fact-checking literature", *Political Studies Review*, 17(3), pp. 296-309.

23. See Walter, N., Cohen, J., Holbert, R. L., & Morag, Y. (2020), "Fact-checking: A meta-analysis of what works and for whom", *Political Communication*, 37(3), pp. 350-375.

message and topic characteristics.²⁴ For example, the hypothesis that audio-visual disinformation is more challenging to correct is prevalent in the current literature but has not been fully explored yet.²⁵

c. Addressing the challenges: best practice and standards of fact-checking

With the professionalisation of fact-checking, there arises a need to establish international reference frameworks and standards.²⁶ These principles, today formalised in works such as the International Fact-Checking Network (IFCN)'s "Code of Principles", are primarily rooted in the principle of independence, which emphasises the importance of being free from partisan biases. Transparency in sourcing, funding, methodology, and communication with the public is also essential. It is crucial to emphasise the development of internationally recognized methodologies and best practices, along with the establishment of standards in the field of fact-checking.

A significant portion of the literature focuses on the close relationship between fact-checking and journalistic work. Best practices in fact-checking are intrinsically connected to the standards of "good" journalism.²⁷ Relevance, credibility, and independence are among the standards by which fact-checking can be assessed. However, fact-checking is not merely an extension of journalism but rather its own distinct practice and profession. As the professional identity of fact-checking continues to evolve, standards that are based on journalistic work are also developing.²⁸

For example, until recently (or perhaps still), there has been little consensus on the optimal length of fact-checks, the information they should include, or the specific methods by which veracity should be sourced (e.g., through new sources or through analysis of existing sources within the original text/visual

-
24. See Chung, M., & Kim, N. (2021), "When I learn the news is false: How fact-checking information stems the spread of fake news via third-person perception", *Human Communication Research*, 47(1), pp. 1-24.
 25. See Thomson, T. J., Angus, D., Dootson, P., Hurcombe, E., & Smith, A. (2022), "Visual mis/disinformation in journalism and public communications: Current verification practices, challenges, and future opportunities", *Journalism Practice*, 16(5), pp; 938-962; Weikmann, T., & Lecheler, S. (2022), "Visual disinformation in a digital age: A literature synthesis and research agenda", *new media & society*, available at : <https://journals.sagepub.com/doi/10.1177/14614448221141648>, accessed 28 September 2023.
 26. Humprecht, E. (2020), "How do they debunk "fake news"? A cross-national comparison of transparency in fact checks", *Digital Journalism*, 8(3), pp. 310-327.
 27. Himma-Kadakas, M., & Ojamets, I. (2022), "Debunking false information: investigating journalists' fact-checking skills", *Digital Journalism*, 10(5), pp. 866-887.
 28. Graves, L., & Amazeen, M. A. (2019), "Fact-checking as idea and practice in journalism", in *Oxford research encyclopedia of communication*.

being checked). Moreover, fact-checkers have developed unique systems of measuring veracity, such as rating and ranking systems.²⁹ These systems diverge from traditional journalistic norms of verification and can present challenges, as they apply a quasi-quantitative logic to the process of verifying complex facts in an online context.³⁰

Due to the global nature of fact-checking as a profession, initiatives have emerged to create international standards and best practices. The International Fact-Checking Network (IFCN) first developed a Code of Principles accompanied by a governance structure. Signatories of the code pledge to adhere to five principles: non-partisanship and fairness, transparency of sources, transparency of funding and organisation, transparency of methodology, and an open and honest correction policy.³¹ Also, the European Fact-Checking Standards Network (EFCSN) has published a Code of Professional Integrity specifically for European fact-checkers, developed in collaboration with fact-checking organisations and approved by them. The code defines European projects as those that have a “substantial and demonstrable focus on one (or more) of the countries of the Council of Europe, plus Kosovo or Belarus”.³² It includes guidelines for compliance and governance. The code focuses on important standards of good practice, such as clear and transparent fact-checking methodologies, the establishment of ethical standards, and transparency in financing and organisational practices. Overall, the establishment of these codes and guideline principles signifies the development of structural professional identities within the fact-checking community as it distinguishes itself further from traditional journalistic fact-checking practices.

The emerging codes and guidelines in fact-checking aim to safeguard the independence of fact-checking activity models (both as independent projects and as part of traditional media organizations), especially in an era where verification and truth-telling are increasingly politicised. These bottom-up initiatives are crucial because an increasing number of governmental and international institutions are also invested in combating disinformation campaigns on both national and international scales. The codes and guidelines reflect a bottom-up, field-driven approach that counters platform initiatives

29. Graves, L. (2018), “Boundaries not drawn: Mapping the institutional roots of the global fact-checking movement”, *Journalism studies*, 19(5), pp. 613-631.

30. Nieminen, S., & Sankari, V. (2021), “Checking PolitiFact’s fact-checks”, *Journalism Studies*, 22(3), pp. 358-378.

31. International Fact-Checking Network (IFCN), Code of Principles, available at : <https://ifcncodeofprinciples.poynter.org/know-more/the-commitments-of-the-code-of-principles>, accessed 28 September 2023.

32. European Fact-Checking Standards Network (EFCSN), European Code of Standards for Independent Fact-Checking Organisations, Art. 5.2, C, available at : <https://eufactcheckproject.com/app/uploads/2022/10/EU-CODE-EFCSN-.pdf>; accessed 28 September.

to establish their own guidelines for content moderation and fact-checking. Platform initiatives often focus on content visibility and findability rather than verification itself.³³ Therefore, the development of independent codes and guidelines ensures that fact-checking efforts remain dedicated to verification and maintain their integrity amidst political pressures and platform-driven interests and business models.

Finally, when it comes to addressing challenges to fact-checkers' working conditions, it may be helpful to build upon the Council of Europe's [Recommendation CM/Rec\(2022\)4 on promoting a favourable environment for quality journalism in the digital age](#).

3. Platform-design solutions

This section explains the context, reasoning and principles underpinning the recommendations on platform-design solutions contained in the Guidance Note; and serves as a reference point for relevant human-rights standards and best practices when implementing the recommendations.

a. Key challenges

It is important to begin by first recognising the considerable challenges which the Guidance Note seeks to address around platform-design solutions to mis- and disinformation. In this regard, among the major challenges are the lack of transparency, accountability and explainability of platforms' algorithmic systems which affect the dissemination of disinformation. Indeed, these specific challenges have been emphasised by both European and international human rights bodies, including the Committee of Ministers, and various Council of Europe reports. It has been specifically described how the lack of transparency and access to data are "major failings" of platforms across "almost all the concerns in relation to disinformation and misinformation", which prevents independent scrutiny, and undermines accountability and trust.³⁴ Crucially, the Committee of Ministers has highlighted the need to ensure the transparency, accountability and explainability of platforms' algorithmic systems, and strengthening accountability of those "developing and implementing" these systems.³⁵

33. Cavaliere, P. (2020), "From journalistic ethics to fact-checking practices: defining the standards of content governance in the fight against disinformation", *Journal of media law*, 12(2), pp. 133-165.

34. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, "Disinformation and freedom of opinion and expression", (A/HRC/47/25), 13 April 2021, paragraph 80.

35. [Recommendation CM/Rec\(2022\)11 of the Committee of Ministers to member States on principles for media and communication governance](#), pp. 6-7.

It has also been emphasised that the lack of transparency of platforms' content moderation and algorithmic systems makes it "impossible" to actually assess the "effectiveness" of measures adopted by platforms in addressing mis- and disinformation and these measures "impact on human rights".³⁶ This has been made all the more difficult due to the lack of access to data for independent researchers, civil society, journalists, and independent regulators to ensure independent and objective assessments. It is absolutely essential that platform responses to disinformation must be grounded in transparency and accountability.³⁷ As such, the recommendations contained in the Guidance Note on platform-design solutions seek to specifically address these challenges around a lack of transparency and accountability, and the central concern that transparency is essential to assess the effectiveness of measures adopted by platforms to tackle mis- and disinformation.

b. Current situation: human rights standards on disinformation regulation

A major point that informs the Guidance Note provisions on platform-design solutions to mis- and disinformation is that any regulatory measure must be consistent with the right to freedom of expression, guaranteed under Article 10 of the European Convention on Human Rights and international human rights law. There have been concerns raised by bodies such as the Council of Europe Commissioner for Human Rights on measures being implemented to tackle disinformation, including regulation permitting the removal and blocking of content deemed "false information".³⁸ The Commissioner has recommended that measures to combat disinformation must never lead to content "being unduly blocked" online.³⁹ As such, measures targeting platforms in particular must be fully consistent with freedom of expression standards on disinformation.

Of note, the European Court of Human Rights has considered instances where individuals have been prosecuted for "dissemination of untrue information", with the Court establishing the fundamental principle that "Article 10 of the Convention as such does not prohibit discussion or dissemination of information received even if it is strongly suspected that this information might not

36. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, "[Disinformation and freedom of opinion and expression](#)", (A/HRC/47/25), 13 April 2021, paragraph 81.

37. See, for example, [Recommendation CM/Rec\(2022\)13 on the impacts of digital technologies on freedom of expression](#), appendix (for standards on ensuring accountability and transparency).

38. Council of Europe Commissioner for Human Rights, [Press freedom must not be undermined by measures to counter disinformation about COVID-19](#), 3 April 2020.

39. Ibid.

be truthful”.⁴⁰ A similar principle has been established under international human rights law. Crucially, four international special mandates on freedom of expression have issued a Joint Declaration on disinformation, stating that prohibitions on the dissemination of information based on “vague and ambiguous” concepts such as “false news” are “incompatible” with international standards on freedom of expression.⁴¹ Indeed, the UN Special Rapporteur on freedom of expression has stated that “penalization of disinformation is disproportionate”; while the concept of disinformation itself is an “extraordinarily elusive concept” to define in law, and “susceptible to providing executive authorities with excessive discretion to determine what is disinformation, what is a mistake, what is truth.”⁴² As such, any regulatory measures which target platform-design to mis- and disinformation should be consistent with these freedom of expression principles, and the provisions of the Guidance Note are also underpinned by these standards.

c. Addressing the challenges: platform-design solutions grounded in human rights

Crucially, the provisions in the Guidance Note build upon a considerable amount of important Council of Europe standards to ensure that platform-design solutions are fully grounded in human rights law. And in this regard, there are a number of essential principles that underpin these recommendations.

First, a fundamental principle flowing throughout the provisions on platform-design solutions (and the Guidance Note as a whole) is that ensuring a favourable and enabling environment for freedom of expression is essential to tackling mis- and disinformation. Indeed, the European Court of Human Rights has held that States have a positive obligation (i.e. duty) to create such

40. *Salov v. Ukraine*, Application No. 65518/01, judgment of 6 September 2005, paragraph 113. See Ó Fathaigh, R., Helberger, N., & Appleman N. (2021), “The perils of legally defining disinformation”, 10(4) *Internet Policy Review* 1. See also, Nenandic, I. & Verza, S. (2022), “European Policymaking on Disinformation and the Standards of the European Court of Human Rights”, in Psychogiopoulou E. and De la Sierra S. (Eds.), *Digital Media Governance and Supranational Courts, Selected Issues and Insights from the European Judiciary*, Edward Elgar Publishing, pp. 175-198.

41. United Nations Special Rapporteur on Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe Representative on Freedom of the Media, the Organization of American States Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples’ Rights Special Rapporteur on Freedom of Expression and Access to Information, *Joint Declaration on Freedom of Expression and “Fake News”, Disinformation and Propaganda*, (FOM.GAL/3/17), 3 March 2017, paragraph 2(a).

42. *Disease pandemics and the freedom of opinion and expression*, Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, A/HRC/44/49, paragraph 42.

an enabling environment for freedom of expression.⁴³ Importantly, under human rights standards, fostering an enabling environment for free expression, with a diverse communications environment, including media diversity, is a “key means of addressing disinformation”.⁴⁴ It is essential that platforms apply human rights standards to their systems, and the Committee of Ministers’ Recommendation CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries is still very much a gold-standard standard-setting instrument platforms should follow, and has detailed provisions on how to operationalise human rights in all actions by platforms.⁴⁵

Second, safety-by-design is essential for algorithmic design, which has also been highlighted by the Committee of Ministers in Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems.⁴⁶ This principle is crucial to prevent and mitigate the risk of adverse effects of mis- and disinformation on individuals and society, particularly disinformation targeting certain groups and disinformation campaigns impacting the online safety of journalists, which has been documented in important studies for UNESCO and the International Center for Journalists, among others.⁴⁷

Third, the principle of proportionality also underpins the provisions on platform-design solutions, which is a bedrock principle under Article 10 of the Convention and the case law of the European Court of Human Rights. It follows that when platforms apply measures to restrict mis- and disinformation, such measures must respect the rights of users, and be implemented using the “least restrictive means”, done in a “transparent and non-discriminatory manner”, and should be limited in scope and duration to what is strictly necessary to avoid the “collateral unjustified restriction or removal of legal content”.⁴⁸ In this regard, it has been noted that certain platforms’ anti-disinformation measures have resulted

43. See, for example, *OOO Memo v. Russia*, Application No. 2840/10, judgment of 15 March 2022, paragraph 9.

44. *Joint Declaration on Freedom of Expression and “Fake News”; Disinformation and Propaganda*, (FOM.GAL/3/17), 3 March 2017, paragraph 3(a).

45. *Recommendation CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries*.

46. *Recommendation CM/Rec(2020)1 on the human rights impacts of algorithmic systems*, appendix, B3.2.

47. See, for example, Posetti, J., Bontcheva, K., & Shabbir, N. (2022), *The Chilling: Assessing Big Tech’s Response to Online Violence Against Women Journalists*, UNESCO, available at: <https://unesdoc.unesco.org/ark:/48223/pf0000383044.locale=en>, accessed 28 September 2023 ; International Center for Journalists, “How Disinformation Fuels Online Violence Storms Targeting Women Journalists”, 14 February 2023. available at: <https://www.icfj.org/news/how-disinformation-fuels-online-violence-storms-targeting-women-journalists>, accessed 28 September 2023.

48. *Recommendation CM/Rec(2018)2 on the roles and responsibilities of internet intermediaries*, appendix, para. 2.3.1 - 2.3.2.

in restrictions on independent news content,⁴⁹ and in implementing design-solutions for disinformation, specific account should be taken of ensuring no disproportionate interferences with media freedom. Recent European regulatory measures, such as the EU’s Digital Services Act (DSA), also seek to ensure that platforms apply terms and conditions in a “proportionate” manner, and consistent with freedom of expression and media freedom, which is also in line with the recommendations of the Council of Europe.⁵⁰

Of further importance is the focus on guaranteeing procedural safeguards. This focus is consistent with international human rights standards on tackling mis- and disinformation, where focusing on due process obligations, instead of “viewpoint - or content-based” regulation, can make a “positive contribution” to protection of human rights and “greater public accountability of platforms”.⁵¹ This focus on guaranteeing procedural rights is also underpinning recent regulatory measures in the EU, with provisions in the DSA on platforms being required to providing statements of reasons for restrictions on content, and obliging platforms to establish internal complaint-handling systems for users.⁵² Importantly, the European Court of Human Rights has also held that Article 10 incorporates important procedural safeguards, including that users know the grounds for content being blocked, and have a forum to challenge a restriction on expression.⁵³

Finally, the recommendations in the Guidance Note on transparency and accountability are central. As mentioned above, the lack of transparency and access to data have been described as crucial failings on the part of platforms in relation to mis- and disinformation. This includes lack of specific data on user engagement with mis- and disinformation, lack of transparency of automated systems used to identify and remove mis- and disinformation, and an overall lack of transparency regarding platforms’ content moderation systems; making

49. [Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression](#), (A/HRC/38/35), 6 April 2018, paragraph 31.

50. See Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act), Article 14. See also Quintais, J.P, Appelman, N., & Ó Fathaigh, R. (2023), “Using Terms and Conditions to Apply Fundamental Rights to Content Moderation”, *German Law Journal*, 24(5), pp. 881-911.

51. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, “[Disinformation and freedom of opinion and expression](#)”, (A/HRC/47/25), 13 April 2021, paragraph 59.

52. See Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), Articles 17 and 20.

53. See, for example, *Vladimir Kharitonov v. Russia*, Application No. 10795/14, judgment of 23 June 2022, paragraph 36.

it near-impossible to assess the effectiveness or impact of measures taken by platforms in relation to mis- and disinformation.⁵⁴ And again, recent European policymaking, such as the DSA, is firmly rooted in building frameworks for ensuring transparency of platform systems. Moreover, focusing on transparency is also consistent with international human rights standards in the approach to tackling mis- and disinformation.⁵⁵ Crucially, the recommendations on transparency and accountability in the Guidance Note should be read in light of the Guidance Note on Best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation, adopted by the Steering Committee on Media and Information Society (CDMSI), which has helpful provisions on ensuring transparency.⁵⁶ And the goal of Guidance Note is to contribute to the operationalisation of the Committee of Ministers call to ensure the “transparency, accountability and explainability” of platforms’ algorithmic systems.⁵⁷

d. Prioritisation of professional news sources and public interest content

There have been notable recommendations on the role of prioritisation of professional news sources and public interest content as a specific tool for platforms in order to tackle mis- and disinformation. Indeed, the Parliamentary Assembly of the Council of Europe, in its Resolution 2255(2019) on Public service media in the context of disinformation and propaganda, specifically recommended that online platforms cooperate with public and private European news outlets to improve the “visibility of reliable, trustworthy news and facilitate users’ access to it”.⁵⁸ Similarly, four international special mandates on freedom of expression and media freedom issued a Joint Declaration in 2023, with a specific recommendation for large online platforms to “privilege independent quality media and public interest content on their services in order to facilitate

-
54. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, “Disinformation and freedom of opinion and expression”, (A/HRC/47/25), 13 April 2021, paragraph 81.
 55. Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, (A/HRC/38/35), 6 April 2018, paragraph 66.
 56. Guidance Note on Content Moderation: Best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation, Adopted by the Steering Committee for Media and Information Society, June 2021.
 57. Recommendation CM/Rec(2022)11 of the Committee of Ministers to member States on principles for media and communication governance, pp. 6-7.
 58. Parliamentary Assembly, Resolution 2255 (2019) on public service media in the context of disinformation and propaganda, 23 January 2019, paragraph 8.2.

democratic discourse”.⁵⁹ The Committee of Ministers has recognised that some platforms have made “efforts to give greater prominence to generally trusted sources of news and information”; however, it has also emphasised that the “impact of these measures on the free flow of information and ideas in democratic societies must be studied carefully”.⁶⁰

Crucially, the criteria for any such identification and labelling must be developed in a transparent and multi-stakeholder process. Indeed, some member states have been implementing prominence regimes, and there is research on these mechanisms, including those applicable to public service media.⁶¹ Notably, States and platforms should follow and build on examples such as the Journalism Trust Initiative by Reporters Without Borders.⁶² In this regard, in implementing this Guidance Note, any measures should be implemented consistent with the detailed and comprehensive Guidance Note on the Prioritisation of Public Interest Content Online, adopted by the CDMSI in 2021.⁶³ Notably, States should be legally restrained from obliging that their own statements and communications are made prominent by platforms, except under public emergencies as defined by Article 15 of the Convention and interpreted in the case law of the European Court of Human Rights.⁶⁴

e. Independent research

Finally, the provisions on independent research are informed by the overriding principle that effective policy-making on platform-design solutions to mis- and disinformation requires “accurate, nuanced and comprehensive knowledge” based on “rigorous and independent research”.⁶⁵ Policymakers in Europe, such as the EU’s DSA, are implementing frameworks to facilitate (and compel) access to data from certain online platforms for independent

59. United Nations Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, the Organization for Security and Co-operation in Europe Representative on Freedom of the Media, the Organization of American States Special Rapporteur on Freedom of Expression and the African Commission on Human and Peoples’ Rights Special Rapporteur on Freedom of Expression and Access to Information in Africa, *Joint Declaration on Media Freedom and Democracy*, 2023, p. 8.

60. *Recommendation CM/Rec(2022)4 on promoting a favourable environment for quality journalism in the digital age*, appendix, paragraph A6.

61. See, for example, Cole M. D. et al. (Eds), (2022), *Prominence of European works and of services of general interest*, IRIS Special, European Audiovisual Observatory.

62. Journalism Trust Initiative, <https://www.journalismtrustinitiative.org>.

63. *Guidance Note on the Prioritisation of Public Interest Content Online*, adopted by the Steering Committee for Media and Information Society, CDMSI(2021)009, 2 December 2021.

64. *Ibid.*, paragraph 21.

65. *Recommendation CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression*, preamble.

research into aspects of platforms' systems.⁶⁶ In this regard, in implementing the provisions in the Guidance Note on guaranteeing independent research, States should follow the detailed provisions in Section 6 of the Committee of Ministers' Recommendation CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression. These include on (i) ensuring researchers can access data held by platforms in ways that are secure, legal and privacy-compliant; (ii) the role of competent authorities to create secure environments that facilitate research; (iii) accessing individual-level data available for independent research; (iv) vetting; (v) liability; and data-sharing agreements between platforms and researchers. And the report of the European Digital Media Observatory's Working Group on Platform-to-Researcher Data Access is an excellent resource for how platforms can share data with researchers.⁶⁷

4. Empowerment of Users

This section explains the context, reasoning and principles underpinning the recommendations on the empowerment of users contained in the Guidance Note.

a. Current situation

Enhancing the integrity of online information requires a holistic strategy that goes beyond reducing exposure to mis- and disinformation. In open, democratic societies it is challenging to completely prevent the production and diffusion of low-quality information without jeopardising freedom of expression. While the other pillars of this Explanatory Memorandum highlight ways in which the contemporary media ecosystem can enhance the circulation of high-quality information and limit the diffusion of low-quality content, ultimately the impact of such information depends on whether users (a) are regularly exposed to, and recognise the importance of, high-quality content; (b) are capable of distinguishing between high-quality and low-quality content; (c) act responsibly towards others in sharing and discussing different kinds of information they may encounter; and (d) enjoy strong safeguarding protections for their human rights, know how to exercise them, and are confident they can make a positive difference, as well as protecting themselves, by exercising them.

In this sense, the role of Public Service Media in Europe is more relevant than ever. By fulfilling its mission to provide accurate and objective information that is freely accessible to all sectors of society, Public Service Media can ensure that

66. See, for example, Digital Services Act, Article 40 (Data access and scrutiny).

67. EDMO, (2022), [Report of the European Digital Media Observatory's Working Group on Platform-to-Researcher Data Access](#).

most individuals regularly encounter reliable news. As the business models of commercial media have been challenged by technological and social changes, the centrality of Public Service Media as the “gold standard” in informing and empowering the public becomes even more crucial.⁶⁸

Empowering users is especially relevant when online communication occurs in private and semi-private platforms, such as WhatsApp, Facebook Messenger, Snapchat, and Telegram, and via messaging services such as iMessage and Android Message. These apps are very popular and constitute an important component of people’s everyday social lives, as they are used to maintain connections with close friends and family, more distant acquaintances, and, at least in some cases, larger groups mostly made up of strangers who are interested in a certain topic or live in a certain area. As communication on these apps is private and often end-to-end encrypted, the scope for applying technologically driven design solutions and interventions against the spread of mis- and disinformation is much more limited than on public social media platforms. At the same time, there is evidence that personal messaging users are reluctant to challenge and correct others who share disinformation on these apps because they fear that they may instigate conflict with others with whom they have close relationships of kinship or friendship, or that they may embarrass them vis-a-vis others when the interaction occurs in groups.⁶⁹ Hence, empowering users so they are equipped to discern between true and false information, responsible in what they share, and capable of having respectful and constructive discussions with those who may share low-quality content is even more important on private messaging than on public social media.

If low-quality information circulating online reaches aware and empowered users, the potential harms resulting from this exposure are likely to be minimal. Such users will be able to identify the markers that usually characterise false or inauthentic messages, will be capable of verifying them, will be less likely to share them, will be more prepared to correct the falsehoods they contain by addressing the person who shared them in a way that is appropriate to the social context of the interaction, and will be aware of any tools at their disposal to report this content to the platform where it circulates, thus contributing to the platform’s ability to recognize and reduce the spread of falsehoods. Thus, user empowerment could facilitate a virtuous circle whereby users not only

68. See the UK Office for Communications’ recommendations on the future of Public Service Media, available at : https://www.smallscreenbigdebate.co.uk/data/assets/pdf_file/0023/221954/statement-future-of-public-service-media.pdf, accessed 26 September 2023.

69. Chadwick, A., Vaccari, C., & Hall, N. A. (2023), “What Explains the Spread of Misinformation in Online Personal Messaging Networks? Exploring the Role of Conflict Avoidance”, *Digital Journalism*, pp. 1-20.

avoid falling into the trap of low-quality content, but also protect and alert other users who may do so, as well as improve digital platforms' accountability and capacity to combat the spread of mis- and disinformation.

When these conditions occur, individuals and societies develop resilience against mis- and disinformation. In turbulent informational environments, and particularly during crises, resilience is a key resource against the individual and societal harms that can result from the spread of falsehoods.

At the individual level, resilience to mis- and disinformation can be expected to be higher among citizens with higher levels of educational attainment, interest in news, digital and information literacy, availability of and familiarity with technology, and access to quality news, among others. Policy interventions to enhance resilience at the individual level should therefore focus primarily on enhancing those characteristics among individuals and groups that lack them.

We can also think about resilience against mis- and disinformation as a societal resource. From this perspective, resilience can be seen as the result of political, media, and economic conditions. In the political system, higher levels of political polarisation and prominent populist parties and leaders may decrease resilience. Foreign and domestic actors can also decrease resilience by actively targeting a country with disinformation and campaigns aimed at cultivating distrust in experts and journalists. Media systems with low trust in news among citizens, weak Public Service Media, and highly fragmented audiences may also be conducive to less resilient societies. As regards economic conditions, the business conditions in which news organisations operate, the size of a country's market for advertising and the levels of social media use may reduce resilience.⁷⁰ Policies aimed at enhancing resilience against mis- and disinformation at the societal level should address some of the systemic factors that facilitate its development.

b. Key challenges

At any time in any given society, some users will already benefit from high levels of empowerment, while others will not. The processes through which humans acquire awareness of problems, learn new skills, and develop new habits are complex and require motivation, time, and resources. Moreover, different groups enjoy different levels of empowerment and require specific approaches to enhance it. One key distinction is between those groups who can pay to access quality information and those who cannot. Another is

70. Humprecht, E., Esser, F., & Van Aelst, P. (2020), "Resilience to Online Disinformation: A Framework for Cross-National Comparative Research", *The International Journal of Press/Politics*, 25(3), pp. 493–516.

between younger and older users of digital media: the former tend to possess higher levels of technical skills but lower levels of interest in news and politics, while the opposite tends to apply for the latter. Other societal divisions (based on religion, ethnicity, gender, and education, among others) also shape the conditions under which individuals can benefit from the opportunities of the information ecosystem. These complexities give rise to some key challenges to the empowerment of users against mis- and disinformation.

First, empowerment should be prioritised as a medium-to-long term goal, with limited dependence on nudges and short term/ immediate solutions in crises and emergencies contentious election. Secondly, policies to enhance user empowerment require long-term investment by governments, reform of the education curricula (including lifelong learning and further education), and wide-ranging collaboration among a variety of stakeholders. Thirdly, any initiative aimed towards empowerment needs to start from users and communities, adapting to their specific needs and contextual conditions, particularly with respect to the most vulnerable and disconnected groups (starting from children and young adults). Programs that establish meaningful connections with the communities they aim to serve, including by co-designing curricula and interventions, are more likely to succeed than programs that adopt a one-size-fits-all, top-down approach. The use of ombudsman-like schemes to independently address user grievances in a flexible and impartial way could also help promote awareness among different communities and safeguard their rights. Fourthly, many initiatives aimed at enhancing empowerment, such as literacy programs, do not reach some of the most vulnerable groups, such as elderly people and those who lack access to digital technologies.⁷¹ Limited engagement by the main platforms and substantial inequalities among member states in the quality, diversity, and accessibility of literacy programs are other key challenges that need to be addressed.⁷² Finally, because human attitudes, skills, and behaviours are complex and difficult to change, the effects of user empowerment programs and interventions are difficult to measure and demonstrate. This problem is compounded by the fact that initiatives in this space, for instance digital and media literacy programs, often forgo the crucial step of evaluating their effects, partly due to lack of funding and short timeframes. Furthermore, programs that undertake robust evaluations of their

71. See for instance the overview of UK-based programmes in the UK Government's Department for Digital, Culture, Media & Sport's [Online Media Literacy Strategy](#).

72. Cabrera Blázquez, F., Cappello M., Talavera Milla J., Valais S. (2022), [User empowerment against disinformation online](#), European Audiovisual Observatory.

effects do not always disseminate the results of these exercises, which makes it challenging to develop cumulative knowledge on these issues.⁷³

A key component of user empowerment is critical thinking, or the tendency to question and seek verification for the information one encounters, and to be prepared to always reflect upon and challenge one's beliefs in light of new and reliable evidence. Critical thinking is a desirable feature of democratic citizenship because it promotes vigilance and responsibility among users, reduces the risk that mis- and disinformation will cascade quickly through online and offline social networks, and ultimately disincentivizes political and media elites to produce and distribute falsehoods, as they are more likely to be called out and sanctioned for doing so.

On the other hand, public discourse emphasising the prevalence of mis- and disinformation and initiatives aimed at increasing citizens' awareness of the problem may also conceivably cultivate cynicism, particularly among some groups already disconnected and distrustful of public institutions and the media.⁷⁴ Cynicism is the generalised belief that others are untrustworthy and that it is impossible and even undesirable to establish the difference between truth and falsehood. Overall, this attitude is democratically dysfunctional because it encourages an "anything goes" mentality whereby no one feels responsible for the information they distribute and for correcting the information shared by others. Cynicism is thus the opposite of empowerment, as it promotes a lack of responsibility and dismisses the notion that users have agency over the environment in which they communicate. Importantly, the sense that cynicism is endemic in a population may also generate what scholars call the "liar's dividend";⁷⁵ which enables political and other actors accused of lying to defend themselves from these charges by claiming that it is impossible to credibly establish the truth. Leveraging and cultivating this form of cynicism is a key principle of authoritarian regimes' approach to propaganda in the twenty-first century.⁷⁶

It is important that any initiatives aiming to promote digital and media literacy and other forms of user empowerment cultivate critical thinking while avoiding

73. The UK Office for Communications has developed a toolkit for evaluating media literacy interventions, available at : <https://www.ofcom.org.uk/research-and-data/media-literacy-research/approach/evaluate/toolkit>, accessed 28 September 2023.

74. Boyd, D. (2017), Did media literacy backfire?, *Journal of Applied Youth Studies*, 1(4), pp. 83–89.

75. Chesney, B., & Citron, D. (2019), "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security", *California Law Review*, 107(6), pp. 1753–1820.

76. Pomerantsev, P. (2017), *Nothing is true and everything is possible: Adventures in modern Russia*. Faber & Faber.

boosting cynicism among citizens. Being exposed to inaccurate content, even if just to show how it can be identified as misleading,⁷⁷ or to discussions of the problem and prevalence of mis- and disinformation in contemporary media ecosystems,⁷⁸ or to news coverage of large-scale disinformation campaigns,⁷⁹ can lead to decreased levels of trust in news and the democratic process, which suggests cynicism rather than critical thinking.

c. Addressing the challenges

There are at least three broad categories of endeavours that can enhance user empowerment against mis- and disinformation. First, digital platforms can provide tools and services that give users the means to control the content made available and recommended to them,⁸⁰ to verify sources and their reliability,⁸¹ and to get swift, fair, and effective redress if they feel their human rights have been limited by the platforms or other users. These initiatives improve the user experience of specific digital platforms by making information available to them that they would not have otherwise encountered and that may support them in making better decisions or protecting their rights. However, absent stringent forms of regulation or co-regulation, these initiatives depend on platforms' decisions and there is limited transparency on how they are implemented.⁸²

-
77. Vaccari, C., & Chadwick, A. (2020), "Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news", *Social Media + Society*, 6(1), available at: <https://journals.sagepub.com/doi/full/10.1177/2056305120903408>, accessed 28 September 2023.
 78. Van Duyn, E., & Collier, J. (2019), "Priming and fake news: The effects of elite discourse on evaluations of news media", *Mass Communication and Society*, 22(1), pp. 29-48.
 79. Ross, A. R., Vaccari, C., & Chadwick, A. (2022), "Russian meddling in US elections: How news of disinformation's impact can affect trust in electoral outcomes and satisfaction with democracy", *Mass Communication and Society*, 25(6), pp. 786-811.
 80. For instance, Facebook offers users various ways to control the content they see on their news feed (available at: <https://www.facebook.com/help/964154640320617>, accessed 28 September 2023) and explanations for why they were shown a particular advertisement (available at: https://www.facebook.com/help/562973647153813?helpref=faq_content, accessed 28 September 2023).
 81. For example, in some regions and languages YouTube enhances some search results with information panels featuring content from fact checkers (available at: <https://support.google.com/youtube/answer/9229632?hl=en>, accessed 28 September 2023).
 82. Commitment 22 in the [European Union's Strengthened 2022 Code of Practice on Disinformation](#), which was signed by most but not all tech companies who adhered to the code, compels platforms to provide users with these types of tools. Commitment 24 in the Code requires platforms to provide adequate information on actions taken against them and mechanisms for appealing those decisions.

Secondly, a variety of public and private actors can design tools – including browser plug-ins,⁸³ educational games,⁸⁴ how-to guides,⁸⁵ and support chatbots⁸⁶ – that assist users in understanding the dynamics of information circulation online, verifying the content they see, and engaging constructively with other users.⁸⁷ These tools can support users in their everyday digital lives but require users to be aware of their existence and willing to engage with them. This is why collaboration by the main digital platforms in promoting and disseminating these tools to their users is essential. The European Union has supported the development of various such tools as part of its programs to tackle disinformation.⁸⁸ One of these tools is EDMO, with a holistic approach that includes fact-checking in all the languages of the European Union, protection of sensitive events in the EU such as elections, as well as research and verification of the engagements taken by the platforms. Public actors can also support a healthy information ecosystem and enhance user empowerment by means of making available at their own initiative information that they hold in line with the requirements of [Recommendation Rec\(2002\)2 of the Committee of Ministers to member States on access to official documents](#) and the [Council of Europe Convention on Access to Official Documents \(CETS No. 205 – the Tromsø Convention\)](#).

-
83. For instance, the InVID Verification Plugin assists users in verifying content on various social media (available at: <https://www.invid-project.eu/tools-and-services/invid-verification-plugin/>, accessed 28 September 2023).
 84. For instance, “Bad News” (available at : <https://www.getbadnews.com/books/english/>, accessed 28 September 2023) guides users in understanding the key factors that facilitate the spread of disinformation and “Go Viral” (available at : <https://www.goviralgame.com/en>, accessed 28 September 2023) focuses on Covid-related disinformation.
 85. The International Federation of Library Associations and Institutions developed infographics that summarize key strategies users can employ (available at: <https://repository.ifla.org/handle/123456789/167>, accessed 28 September 2023). First Draft offers resources for users and journalists in verifying information (available at: <https://firstdraftnews.org/training/>, accessed 28 September 2023). UNESCO has published a handbook for journalism education and training on how to identify disinformation. Ireton (C.), Posetti (J.), (Eds.), (2018), *Journalism, fake news & disinformation: handbook for journalism education and training*, available at: <https://unesdoc.unesco.org/ark:/48223/pf0000265552>, accessed 28 September 2023.
 86. The “Fake News Immunity Chatbot” (available at: <http://fni.arg.tech/>, accessed 28 September 2023) helps individuals identify invalid arguments in different types of news. Various fact checkers, such as Maldita in Spain, have experimented with chatbots that automatically respond to user queries about the veracity of information they report (available at: <https://www.europeanpressprize.com/article/maldita-es-whatsapp-chatbot/>, accessed 28 September 2023).
 87. The RAND Corporation maintains a list of tools against disinformation online (available at: <https://www.rand.org/research/projects/truth-decay/fighting-disinformation/search.html>, accessed 28 September 2023).
 88. A list and description of funded projects is available at: https://commission.europa.eu/strategy-and-policy/coronavirus-response/fighting-disinformation/funded-projects-fight-against-disinformation_en, accessed 28 September 2023.

Thirdly, media and information literacy is crucial to enhance user empowerment. There are various definitions of different types of literacy: for instance, the Committee of Ministers in Recommendation CM/Rec(2022)4 on promoting a favourable environment for quality journalism in the digital age, states that media and information literacy involves the development of cognitive, technical and social skills and capacities that enable people to: (a) effectively access media content and critically analyse information, thus empowering them to understand how media content is produced, funded and regulated, as well as to have the confidence and competence to make informed decisions about which media they use, and how they use them; (b) understand the ethical implications of media and technology; and (c) communicate effectively, including by interpreting, creating and publishing content.⁸⁹ The UK Office for Communications defines “media literacy” as “the ability to use, understand and create media and communications in a variety of contexts.”⁹⁰ UNESCO defines “media and information literacy” as “an interrelated set of competencies that help people to maximise advantages and minimise harm in the new information, digital and communication landscapes”⁹¹ Scholars have defined “news literacy” as an awareness of 5 “C’s” pertaining to how news is produced, distributed, and understood: context (the environment in which news exists), creation (the process of news production), content (what the news says), circulation (how the news is distributed), and consumption (whether audiences pay attention to it and how they make sense of it).⁹² As all these types of literacy are fundamental to citizens’ ability to protect themselves against mis- and disinformation and exercise their rights, their inclusion in educational curricula at all levels is a necessary first step.⁹³ However, most members of a society at a given time will not be enrolled in formal education, hence it is equally as important that opportunities to enhance these skills be provided to people from all walks of life, particularly those who are more likely to be

89. Recommendation CM/Rec(2022)4 on promoting a favourable environment for quality journalism in the digital age, appendix, paragraph 9.

90. Available at: <https://www.ofcom.org.uk/research-and-data/media-literacy-research>, accessed 28 September 2023.

91. Available at: <https://www.unesco.org/en/media-information-literacy/about>, accessed 28 September 2023.

92. Vraga, E. K., Tully, M., Maksl, A., Craft, S., & Ashley, S. (2021), “Theorizing News Literacy Behaviors”, *Communication Theory*, 31(1), pp. 1–21.

93. The European Union has developed a set of “Guidelines for teachers and educators on tackling disinformation and promoting digital literacy through education and training”. Available at: <https://education.ec.europa.eu/news/guidelines-for-teachers-and-educators-on-tackling-disinformation-and-promoting-digital-literacy-through-education-and-training>, accessed 28 September 2023.

vulnerable and less likely to come into contact with educational institutions.⁹⁴ Public Service Media can be crucial to reach these users with content that can enhance their levels of literacy.

It is important to emphasise that any initiative aimed at enhancing user empowerment should not be seen in isolation, but as part of a broader, multi-pronged strategy to improve the health of information ecosystems and resilience against mis- and disinformation in democratic societies. This is not only because, as discussed earlier, there are limits to what any initiative can accomplish in the short to medium term, but also because placing the emphasis solely or predominantly on citizens' skills may inadvertently play into narratives that "blame the victim", i.e., the targets of disinformation, for being unable to defend themselves against the harms caused by its spread, and risks overlooking the importance of promoting structural conditions that make high-quality content easily available to all members of society.

There are many examples across the world of digital tools and initiatives aimed at promoting user empowerment among different sectors of the population. UNESCO's Media and Information Literacy Alliance⁹⁵ promotes much-needed international cooperation on these issues and provides expertise and resources. The Digital Future Society summarized relevant international success stories.⁹⁶ The European Audiovisual Observatory published a comprehensive report that highlights key policies and initiatives at both national and European levels.⁹⁷ The European Platform of Regulatory Authorities (EPRA) hosts materials and presentations from an event on news literacy organised in 2022 that brought together multiple stakeholders.⁹⁸

As this is a fast-moving and diverse field, it would be impossible to comprehensively and accurately summarise the main initiatives in this space and their outcomes but pooling together information and resources – particularly evaluations of the effects of different actions – would be very beneficial to all stakeholders involved. This is why, in Paragraph 36 of the Guidance Note,

-
94. Recommendation CM/Rec(2022)13 on the impacts of digital technologies on freedom of expression offers suggestions on what aspects digital literacy programs should include. The UK Government's Department for Digital, Culture, Media & Sport's [Online Media Literacy Strategy](#) lists five key principles that support strong media literacy capabilities.
 95. Available at: <https://www.unesco.org/en/media-information-literacy/alliance>, accessed 28 September 2023.
 96. Digital Future Society (2020), [Dealing with disinformation: Strategies for digital citizen empowerment](#).
 97. Cabrera Blázquez, F., Cappello M., Talavera Milla J., Valais S. (2022), [User empowerment against disinformation online](#), European Audiovisual Observatory.
 98. The events of EPRA are available at: <https://www.epra.org/attachments?category=mil-taskforce&page=1>, accessed 28 September 2023.

we recommend that the main actors involved in promoting a healthy informational ecosystem collaborate in creating a public informational resource that collates extensive, standardised, and up-to-date data and evidence on any initiatives aiming to promote user empowerment across member States.

Finally, there is a need for further research on what works and what does not in this area. As previously discussed, our knowledge on the subject is limited by the lack of robust and systematic evaluations of existing interventions, but also by the fact that scientific research on these subjects is still in its infancy and limited to a narrow subset of wealthy liberal democracies. Of particular urgency is research into the factors that predict successful outcomes among different societal groups, particularly the most vulnerable in the population (for instance children, older people,⁹⁹ individuals living in poor information ecosystems, and non-users of digital technologies), and in contexts where democratic institutions and norms are historically weaker or are being put under strain by domestic and foreign actors that seek to benefit from democratic disruption.

5. Conclusion

This Explanatory Memorandum has sought to explain the context, reasoning and principles underpinning the recommendations contained in the Guidance Note. In summing up, as a practice and profession, fact-checking must adapt and innovate to combat the proliferation of mis- and disinformation. Studies indicate that international standards on fact-checking are still in development, and that the professional identity of fact-checkers is evolving also. Most importantly, the fact-checking landscape must be mindful of empirical evidence regarding the effectiveness and effects of their work. A growing body of literature indicates that citizens do not always accept corrected information, but that their response to fact checking heavily depends on political ideology and the way fact-checked information is presented. Further, on platform-design solutions, there are challenges to overcome in tackling mis- and disinformation, in terms of the lack of transparency and accountability of platforms' algorithmic systems. Helpfully, there are considerable European and international standards upon which the Guidance Note builds upon which can contribute to ensuring an enabling and favourable online environment for a pluralistic public debate, which can serve as a crucial counterweight to mis- and disinformation. And finally, empowerment of users can facilitate a virtuous circle whereby users

99. See Hermans, A. (2022), [The digital era ? Also my era !, Media and Information literacy: a key to ensure seniors' rights to participate in the digital era](#), Information Society Department, DGI (2022)03, Council of Europe.

not only avoid falling into the trap of low-quality content such as mis- and disinformation, but also protect and alert other users who may do so, as well as improving the accountability and capacity of the platforms to combat the spread of mis- and disinformation.

www.coe.int

The Council of Europe is the continent's leading human rights organisation. It comprises 46 member states, including all members of the European Union. All Council of Europe member states have signed up to the European Convention on Human Rights, a treaty designed to protect human rights, democracy and the rule of law. The European Court of Human Rights oversees the implementation of the Convention in the member states.

COUNCIL OF EUROPE



CONSEIL DE L'EUROPE