RSC 2023/34
Robert Schuman Centre for Advanced Studies
Centre for Media Pluralism and Media Freedom

WORKING PAPER

Structural indicators to assess effectiveness of the EU's Code of Practice on Disinformation

Iva Nenadic, Elda Brogi, Konrad Bleyer-Simon

European University Institute
Robert Schuman Centre for Advanced Studies
Centre for Media Pluralism and Media Freedom

Structural indicators to assess effectiveness of the EU's Code of Practice on Disinformation

Iva Nenadić, Elda Brogi, Konrad Bleyer-Simon

RSC Working Paper 2023/34

ISSN 1028-3625

Published in April 2023 by the European University Institute.
Badia Fiesolana, via dei Roccettini 9
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu

With the support of the Erasmus+ Programme of the European Union

The European Commission supports the EUI through the European Union budget. This publication reflects the views only of the author(s), and the Commission cannot be held responsible for any use which may be made of the information contained therein.

Robert Schuman Centre for Advanced Studies

The Robert Schuman Centre for Advanced Studies, created in 1992 and currently directed by Professor Erik Jones, aims to develop inter-disciplinary and comparative research on the major issues facing the process of European integration, European societies and Europe's place in 21st century global politics.

The Centre is home to a large post-doctoral programme and hosts major research programmes, projects and data sets, in addition to a range of working groups and ad hoc initiatives. The research agenda is organised around a set of core themes and is continuously evolving, reflecting the changing agenda of European integration, the expanding membership of the European Union, developments in Europe's neighbourhood and the wider world.

For more information: http://eui.eu/rscas

Centre for Media Pluralism and Media Freedom (CMPF) Working Paper Series on 'Freedom and Pluralism of the Media, Society and Markets' benefits from contributions from the CM PF's fellows as well as from leading scholars and experienced practitioners interested in and focused on the subject matter. The Working Papers Series aims at assessing theoretical issues, specific policies, and regulatory questions.

The Centre for Media Pluralism and Media Freedom is co-financed by the European Union. This initiative is a further step in the European Commission's on-going effort to improve the protection of media pluralism and media freedom in Europe and to establish what actions need to be taken at European or national levels to foster these objectives.

The aim of the EUI Centre for Media Pluralism and Media Freedom is to enhance the awareness of the importance of freedom and pluralism of the media, to contribute to its protection and promotion and to develop new ideas among academics, policy makers, regulators, market stakeholders, journal., and all other directly involved professionals who take part in the public debate

Abstract

The key instrument of the European Union's policy against disinformation – the Code of Practice on Disinformation - in its improved version seeks to adopt structural indicators to examine the disinformation phenomenon and gauge effectiveness of the Code in suppressing it both in individual EU member states and in the EU as a whole. The paper outlines the process and the proposal for an initial set of approaches and metrics towards building such structural indicators. This initial proposal is a pioneering attempt placed in a policy framework of the self-regulatory Code, and with consideration that empirical research on online disinformation in Europe is limited and there are no systematic and cross-country comparable insights on how the problem evolves in its various dimensions. The proposal described in this paper should thus be seen as a minimum and first step in what should be a wider and more systematic attempt to monitor disinformation and related policy effectiveness in Europe.

Keywords

Code of Practice on Disinformation; Structural Indicators; EU policy; monitoring disinformation; monitoring policy effectiveness

Sponsor Information

This project has received funding from the European Union under Contract number: LC-01935415

Introduction

The European Union has been on the forefront of designing and implementing policies to tackle disinformation. In 2018, it explicitly recognised the spread of intentionally deceiving content as one of the prime problems that European democracies are facing[1]. In the same year, the central instrument of the EU anti-disinformation policy was presented: the Code of Practice on Disinformation, conceived as a form of self-regulation by the leading online platforms and representatives of the advertising industry, but initiated and supervised by the European Commission. However, after the initial period of its implementation, it became clear that the form, scope, and implementation regime of the Code needed to be revised to allow for appropriate monitoring of its effectiveness. The review of the initial Code, conducted by the European Regulators Group for Audiovisual Media Services (ERGA)[2], and the European Commission itself[3], highlighted a set of significant shortcomings, including the lack of key performance indicators against which to gauge achievements. To address this deficiency, the updated 2022 version of the Code envisions a set of key performance indicators, organised in two levels: Service-level Indicators are tied to the specific measures adopted under the Code; while Structural Indicators, still to be adopted, should examine the disinformation phenomenon and the effectiveness of the Code in suppressing it both in individual EU member states and in the EU as a whole to the extent possible. In this paper, we present the initial pilot methodology for such Structural Indicators developed by the Centre for Media Pluralism and Media Freedom within the European Digital Media Observatory (EDMO) project and discussed in a process of exchange with the Code's signatories. The paper first briefly presents the policy framework that requested the development of such indicators. A literature review examines the challenges and approaches in policy evaluation and in building structural indicators. Finally, the methodological process and the proposal for the Structural Indicators to accompany the European Code of Practice on Disinformation are discussed in detail.

A maturing EU policy against disinformation

EU policy against disinformation took shape in 2018, following the establishment of a High-Level Expert Group on Fake News and Online Disinformation (HLEG)[4] in early 2018 and the Group's Report, published two months later, advising the EC against simplistic solutions in tackling the phenomenon. The Commission also ran public consultations[5] and dialogues with relevant stakeholders, and the specific Eurobarometer[6] poll was conducted in 2018 in all EU member states indicating a high level of concern among the respondents for the spread of online disinformation in their country (85%) and seeing it as a risk for democracy in general (83%). Based on all these inputs, in April 2018 the European Commission published *Communication on Tackling online disinformation: a European Approach*, a document that set the tone for future actions in this field. The key output of the Communication and of a related Action Plan[7] was the Code of Practice on Online Disinformation, presented as a self-regulatory instrument that should encourage proactivity of online platforms in ensuring transparency of political advertising and restricting the spread of disinformation. The Code applied within the framework of existing laws of the EU and its Member States, including Directive 2000/31/EC, and, in particular, to liability rules under Articles 12 to 15. The Code was signed in October 2018 by the online platforms Facebook, Google, Twitter and Mozilla, as well as by advertisers and other players in the advertising industry. Microsoft joined in May 2019, and TikTok in June 2020. The key categories of the Code were: demonetizing the spread of misinformation, ensuring the transparency of political advertising, empowering users through greater transparency around algorithmic content recommenders, demoting misinformation and promoting 'trustworthy'

1 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0236
2 https://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf
3 https://digital-strategy.ec.europa.eu/en/library/assessment-code-practice-disinformation-achievements-and-areas-further-improvement
4 https://ec.europa.eu/digital-single-market/en/news/experts-appointed-high-level-group-fake-news-and-online-disinformation
5 https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-fake-news-and-online-disinformation
6 https://europa.eu/eurobarometer/surveys/detail/2183
7 https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX:52018JC0036

sources, strengthening cooperation with fact-checking organisations and ensuring access to data from leading Internet platforms for researchers.

The Code was brought to light as a unique mechanism with the potential to expand globally and to include other relevant actors. Even if acknowledged as a significant step forward, after the initial period of its implementation, it became clear that the form, scope, and the implementation regime of the Code needed to be revised to allow for appropriate monitoring of its effectiveness. The Sounding Board, composed of academics, media, and civil society organisations, who reviewed the establishment of the CoP in 2018, already then presented a critical view on its content and the commitments laid out by the platforms, stating that it "contains no clear and meaningful commitments, no measurable objectives" and "no compliance or enforcement tool"[8]. This was confirmed two years later in the evaluation provided by ERGA[9], and the Commission[10] itself. Both reports highlighted a set of significant deficiencies, including difficulties in understanding the activities implemented and their potential impact, especially as functional data access was not provided by platforms, and the key performance indicators were not established as to adequately gauge achievements.

The Code was therefore revised, and in the new edition[11], published in summer 2022, it includes Service-level Indicators. The signatories of the new Code's edition further committed to work together within the Task-force as well as with EDMO, ERGA, and relevant third-party experts to develop a first set of Structural Indicators that can help assess the impact and success of the Code[12]. Principles on self-regulation and co-regulation require the setup of mechanisms to monitor "the degree and success of their implementation, using objective criteria and reliable indicators defined in advance and specified according to sectors and objectives".[13] The Task-force is designed as a permanent feature of the new Code where signatories, together with EDMO, ERGA and relevant third-party experts regularly review these indicators over time. The new Code is wider in scope in such a way that it includes a number of new signatories, representatives of various stakeholders, above all fact-checkers and civil society. As such, the Code is an unusual form of self-regulation which, in any case, increasingly takes the form of co-regulation as the full implementation of the Digital Services Act (DSA)[14] approaches. The DSA, which is a horizontal framework for regulatory oversight, accountability, and transparency of platforms and search engines in the EU, could strengthen the Code of Practice and transform it into a Code of Conduct where it becomes an instrument for fulfilling legally prescribed obligations, namely, to assess and mitigate the risk of disinformation.

Another relevant dimension of the DSA is that it contains legal obligation to ensure access to data of very large online platforms and very large search engines for vetted researchers (including civil society organizations that conduct scientific research in the public interest) and in a way that is proportionate and appropriate to protect rights, such as the personal data of users (Art. 40). This is a key precondition for effective monitoring of the online information environment and the effectiveness of the Code. So far, there has been an asymmetry of information between online platforms (as information holders) and researchers, regulators, journalists, and the public (as information seekers). Furthermore, there has been a dimension of regional asymmetry in data access: when some states, public authorities, and researchers obtain better access due to their political power, legal competencies, and research capacities, while others are lagging behind. Considering the importance of building policies on evidence, and having in mind the diversity of political and media systems in the EU member states, the data access provided should be more inclusive, and the local capacities should be strengthened to benefit from such access. Functional data access is the key dimension of transparency of platform operations, and the precondition for any meaningful assessment of the

8 THE SOUNDING BOARD'S UNANIMOUS FINAL OPINION ON THE SO-CALLED CODE OF PRACTICE 24 SEPTEMBER 2018

9 https://erga-online.eu/wp-content/uploads/2020/05/ERGA-2019-report-published-2020-LQ.pdf

10 https://digital-strategy.ec.europa.eu/en/library/assessment-code-practice-disinformation-achievements-and-areas-further-improvement

11 https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation

12 https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation p. 37

13 Opinion of the European Economic and Social Committee on Self-regulation and co-regulation in the Community legislative framework (own-initiative opinion) (2015/C 291/05) https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52014IE4850&rid=3

14 https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2065

impact platforms' policies may have on information environments.

The role of EDMO

The European Digital Media Observatory (EDMO) is a project, financed by the European Commission, that brings together fact-checkers, media literacy experts, and academic researchers to understand and analyse disinformation, in collaboration with media organisations, online platforms and media literacy practitioners. It deploys a platform to support the work of a multidisciplinary community with expertise in the field of online disinformation. The main pillars of EDMO are: the creation of community platforms and tools for fact-checking organisations and researchers; the support and coordination of fact-checking and open source investigations activities; the support and coordination of media and information literacy activities; the support and coordination of research activities; the fostering of a framework for online platforms' data access for research purposes; and policy research and analysis, which is a pillar coordinated by Centre for Media Pluralism and Media Freedom. EDMO counts on the work of its 14 national and regional hubs who are active in all EU member states and Norway. EDMO is also part of the Task-force of the Code of Practice on Disinformation.

The Policy Research and Analysis task by EDMO, coordinated by the Centre for Media Pluralism and Media Freedom at the European University Institute, was asked by the European Commission to support the permanent Task-force of the 2022 Code of Practice on Disinformation in developing an initial set of Structural Indicators to measure the contribution of the Code in tackling disinformation. This endeavour first followed the instructions provided by the Commission's Guidance on Strengthening the Code of Practice on Disinformation[15]: while Service-Level Indicators are tied to the specific measures adopted under the Code, Structural Indicators should serve the evaluation of various dimensions of mis- and dis-information in a more comprehensive, objective, and longitudinal way, as well as the overall role of the Code in suppressing it (the Code's implementation).

We operated within the requirement and the constraints that this initial proposal for Structural Indicators needs to be feasible and rapidly applicable. This pioneering attempt has been conducted in a challenging framework in which the parameters of monitoring the implementation of the Code and thus also the implementation of the Structural indicators were unknown. The exact framework and the resources available for the challenging task of implementing this monitoring in all EU member states at regular intervals is yet to be established. Based on our extensive experience in developing and implementing a holistic but feasible methodology in assessing the state of play of media pluralism in all EU member states and candidate countries on a regular basis, we remain convinced that such monitoring and Structural Indicators should be comprehensive in a way that includes quantitative insights based on platform data and other research and contextualised with the assessment of local social, political and policy context. However, considering the need to provide a first iteration of Structural Indicators for 2022 Code in a narrow timeframe, we tabled a proposal that included six areas of measurement: prevalence, sources, audiences, demonetization of disinformation, as well as collaboration and investments in fact-checking, and investment in the implementation of the Code. The proposal explored the potential of engaging the existing networks and resources within the Code (i.e. various signatories) and EDMO (i.e. fact-checking community, local EDMO Hubs).

This first proposal is not a fully-fledged methodology to understand the disinformation phenomenon in the EU and its member states and to assess the effectiveness, to the extent possible, of the Code of Practice on Disinformation. It should be seen as a starting point for triggering a discussion with the signatories of the Code of Practice and with a broader academic community. As it was stressed also by the experts consulted for this proposal, research on disinformation in Europe is limited, and would need extensive investments and resources to contribute an assessment that can be safely used for policy purposes.

15 https://ec.europa.eu/commission/presscorner/detail/en/ip_21_2585

Concepts, theories, and instruments in policy evaluation

In his 2021 book[16] on regulating platforms, Terry Flew points out that a pragmatic approach of assessing platform governance measures "would consider the policy objectives, the measures proposed, their possible or probable effectiveness, and the risks associated with any unintended consequences". A systematic approach proposed to assess the impact of wider media policies is regulatory impact analysis[17]. It is widely used by multilateral institutions, including the European Commission for the establishment of monitoring systems, but this approach is limited to forms of assessment that weight costs and benefits of a given intervention, and thus cannot be easily adapted to the context and the intended objectives of the assessment of the Code of Practice. Our tool of choice, the development and assessment of structural indicators, is mentioned in several European Union documents (such as on the monitoring of the Lisbon Strategy, the EU 2020 Strategy and the monitoring of education and training reforms); it is utilised to track improvements in specific policy fields across countries. In the context of educational policy, Parveva et al. (2022)[18] mention that their structural indicators provide contextual information to the analysis of the field where policy developments took place: being created based on research findings in the relevant field, and guided by EU communications and recommendations. However, this and other such documents lack a proper definition of structural indicators, and they do not make clear whether the assessment is focused on statistical data or the mapping and assessing of existing policies.

In general terms, structural indicators can be said to be used to monitor the implementation of certain policies (e.g. employment, health, media etc.), and in relation to the objectives to be achieved. Most commonly, such indicators use socio-demographic variables and key structural components to measure them over time, with the objective of observing structural transformations.

Furceri and Mourougane (2010)[19] clearly elaborate how structural indicators can be differentiated according to a number of criteria, for instance: they can be perception-based (in this context this would be audience studies - surveys) or fact-based (in this context: relying on verifiable platform data). Fact-based, as the authors also point out, does not mean objective: the choice of questions embodies some level of subjectivity and there should be a way to validate the data provided. Indicators can be single or composite. A composite indicator combines different sub-indicators or variables into a single measure, as such it allows for the same concept to be measured by different data sources, which may increase the reliability of the measurement, but the challenge with composite indicators is how to decide the weight given to its components. According to Furceri and Mourougane (2010), the indicators can also be policy versus outcome measures. For this specific exercise, it is difficult to start with outcome measures as the Code has just been adopted, but the indicators could quantify the disinformation phenomenon in some clearly defined segments of the digital environment and assess whether the resources invested by the signatories to implement commitments are adequate.

Another pivot in the process is the intended goal of the intervention. Since the Code of Practice doesn't set clear benchmarks related to the desired degree (or a complete absence of it) of disinformation on platforms and in information diets (in part due to the difficulties assessing the current level), structural indicators can draw – a still limited – picture of the problem, and register the changes over time, as subsequent implementations of the indicators take place. A guiding principle can be the work of Ortiz and Leal (2020)[20] which, in the context of energy policy, describes the conceptual differences between the assessment of the effectiveness, efficiency and efficacy of measures, and then moves on to find specific indicators for a comprehensive assessment. An efficiency-based assessment would

16 Flew, Terry. *Regulating platforms*. John Wiley & Sons, 2021., p. 166

17 Oermann, Markus, and Wolfgang Schulz. "Assessing policy III: regulatory impact assessment." *The Palgrave handbook of methods for media policy research* (2019): 575-593.

18 Parveva, T., Motiejunaite, A., & Noorani, S. (2022). Structural indicators for monitoring education and training systems in Europe 2022: overview of major reforms since 2015. https://op.europa.eu/en/publication-detail/-/publication/1e14056b-6701-11ed-b14f-01aa75ed71a1/language-en

19 Furceri, D., & Mourougane, A. (2011). Structural Indicators: a critical review. *OECD Journal: Economic Studies*, *2010*(1), 1-34.

20 Ortiz, D., & Leal, V. (2020). Energy Policy Concerns, Objectives and Indicators: A Review towards a Framework for Effectiveness Assessment. Energies, 13(24), 6533.

compare the outcome of a measure to a specified goal, effectiveness can allow for the evaluation of the degree of "success" (how much has been achieved and how the outcome relates to the original objective), while efficiency aims to understand the relationship between the outcome and the means invested. While efficacy-based assessments are out of the scope of this exercise, structural indicators provide a proxy for the effectiveness of measures (the change registered in the area assessed by the indicators). Moreover, by supplementing the CoP's qualitative reporting elements and service-level indicators (which look at measures taken by individual platforms to fulfil their commitments) with structural indicators (that can help better understand the state of disinformation on online platforms and track developments/transformations between reporting periods), the exercise would also provide an indication of efficiency (to what extent the measures and means described in the first two sets of indicators can contribute to the favourable outcome). Such a set would be useful for ex-post evaluations, but in some cases, it could also be useful for ex-ante evaluation and thus assist the design of better policy programmes. As such, the assessment measures under the Code of Practice would come close to the holistic and pragmatic approach outlined by Flew. Therefore, the main motivations of this work have been to create a support document that enables a more structured and comprehensive planning process (and that does not leave important aspects forgotten); planning programmes that have a higher chance of succeeding, in practice; and supporting the good use of resources (public and/or private).

Still, the limits of such an approach should not be disregarded. While structural indicators provide a good proxy to assess the state of disinformation on online platforms, the approach will inevitably have its limits. As shown by Tsfati et al. (2020) in their review of literature exploring the role of mainstream news media in the dissemination of inaccurate and misleading information, circumstantial evidence suggests that more people learn about fake news stories from mainstream media than from social media. Similarly, Altay et al. (2023)[21] point out that the prevalence and spread (or "circulation") of disinformation is not limited to social media: news websites and offline media are also used to spread mis- and disinformation. They point out that the working definition of mis- and disinformation, the sample (whether the focus is on political or social issues or a larger pool of posts, including personal news) as well as the availability of data (as posts that are less shared tend to be excluded) will impact the findings. Moreover, the authors add that prevalence of or even interactions with disinformation may not allow for an accurate assessment of the influence of mis- and disinformation or the informedness of social media users. In addition, we need to highlight an aspect mentioned by Flew (2021): the unintended consequences. Our approach – at least until the group of signatories is limited – cannot assess whether and to what extent disinformation increases on platforms which are not party to the Code of Practice on Disinformation, for example due to purveyors' conscious move to new platforms. As some recent empirical research shows, current platform interventions may result in (temporarily) reducing users activity but increasing toxicity and the sharing of biased news in the long term and at other places by stimulating users migration to alternative platforms[22]. Furthermore, platforms are known for transience – a concept used to describe how platforms frequently and suddenly change their policies, procedures, and affordances[23]. It is, at the moment, unknown whether and how these frequent interventions interact with each other and what their individual and cumulative impact may be at different points in time. A holistic policy assessment in this area should consider trends cross-platform and in a systematic longitudinal way.

21 Altay, S., Berriche, M., & Acerbi, A. (2023). Misinformation on misinformation: Conceptual and methodological challenges. Social Media+ Society, 9(1), 20563051221150412.

22 Trujillo and Cresci. "Make Reddit Great Again: Assessing community effects of moderation interventions on r/The_Donald." *The 25th ACM Conference On Computer-Supported Cooperative Work And Social Computing (CSCW'22)*.

23 Barrett B, Kreiss D. Platform transience: changes in Facebook's policies, procedures, and affordances in global electoral politics. *Internet Policy Review* [Internet]. 2019 [23 March 2023]; 8(4).

A methodology for developing CoP Structural Indicators

The process of developing the proposal for Structural Indicators to accompany the Code of Practice on Disinformation consisted of broad consultations with policy and technology experts, academic researchers, non-profit and civil society organisations, members of the EDMO Executive and Advisory Boards, and ERGA. It also included a literature review of various approaches, methods and metrics used in empirical studies measuring mis/disinformation (any aspect of it), and at scale. This process was primarily informed by a repository of European academic research studying disinformation in digital media at scale, produced under EDMO, and by the MediaWell citation library of the Social Science Research Council that aggregates emerging and foundational research on topics related to digital dis/misinformation.

Consultations with experts

In the process of understanding the scope and resources available for the implementation of Structural Indicators, we have relied on the Commission's Guidance on Strengthening the Code of Practice on Disinformation and the direct consultations with the European Commission.

In the process of developing the proposal for Structural Indicators we acted as a hub: engaging and discussing with a number of policy and technology experts, academic researchers, non-profit and civil society organisations, members of the EDMO Executive and Advisory Boards, and the European Regulators Group for Audiovisual Media Services (ERGA), collecting their input, ideas, experiences and suggestions on how Structural Indicators for this purpose and within the specific framework should be developed. We further engaged with the experts and stakeholders throughout the whole process of preparing and revising this proposal.

One of the main challenges in drafting any indicators that seek to assess the disinformation phenomenon is that there is no broad-based consensus and common definition of the phenomenon, mostly due to its complexity, and thus various platforms approach disinformation in different ways (e.g. as content or as behaviour). The Commission's Guidance on Strengthening the Code, following the definition of the High-Level Expert Group on Fake News and Online Disinformation[1] defines disinformation as *false or misleading content spread with an intention to deceive or secure economic or political gain and which may cause public harm*. The focus in this definition is on content. Experts and scholars have generally distinguished at least four dimensions of disinformation that require attention: (1) harmful content, (2) manipulative actors, (3) the variety of techniques and technologies used by actors to increase the reach and impact of their campaigns, and (4) digital architecture of online platforms and their internal policies that affect the potential spread of misinformation[24].

The 2022 Code, following the Commission's Guidance, expands its scope to include misinformation, further to disinformation. The Guidance defines misinformation *as false or misleading information spread without a malicious intent, but the effects can be still harmful, when there is a significant public harm dimension*. Misinformation indeed can be problematic, especially when viral and in high-intensity events, and it can often overlap with disinformation in terms of contents, or they can build on each other. However, it is already difficult to achieve an agreement (even if just within the EU) over what kinds of contents and behaviours constitute disinformation, as interpretations of what contents are problematic, harmful, unacceptable, or even illegal, depend largely on specific political and cultural context and legal traditions of different countries. The key differentiation between disinformation and misinformation is to be found in the existence of intent, but it is hard to imagine how to determine, in all possible cases, whether there is or no intention to deceive or to do any other harm with the spread of content that is false or misleading. Furthermore, it is not clear who should and how to establish whether "there is a significant public harm dimension".

24 see, for example, François, 2019; and Alaphilippe, 2020; ERGA, 2020; Gillespie, 2022, Leersen, 2023

According to the Preamble, within the 2022 Code, disinformation is considered to include **misinformation** (*false or misleading content shared without harmful intent though the effects can be still harmful*), **disinformation** (*false or misleading content that is spread with an intention to deceive or secure economic or political gain and which may cause public harm*), **information influence operations** (*coordinated efforts by either domestic or foreign actors to influence a target audience using a range of deceptive means, including suppressing independent information sources in combination with disinformation*), and **foreign interference** (*coercive and deceptive efforts to disrupt the free formation and expression of individuals' political will by a foreign state actor or its agents*). This way, the signatories are applying a broad notion and operationalization of disinformation (content, behaviour, manipulative actors, techniques, and technologies).

Literature Review

Research on disinformation has focused on exploring the characteristics and dynamics related to the different dimensions of the phenomenon. Lately, the research in the US is increasingly observing ideological and socio-demographic asymmetries in disinformation production and reception (see, for example, Freelon et al. 2020). This more comprehensive approach, focusing on content, actors, behaviours, and technologies is also elaborated by Alice Marwick and Rebecca Lewis (2017), who further consider also vulnerabilities in the news media ecosystem as a factor to increasing the visibility of and audience for their disinformation, as well as motivations of the actors involved (both purveyors and receivers of disinformation). We are not aware of a similar stream of research in Europe.

To understand what kind of methodological approaches and metrics have so far been used in empirical studies measuring mis/disinformation (any aspect of it), and at scale, we explored the repository of European academic research studying disinformation in digital media at scale, produced under EDMO[2]. We have also explored the MediaWell citation library of the Social Science Research Council that aggregates emerging and foundational research on topics related to digital dis/misinformation. More specifically, we examined research focusing on topical streams classified as "Algorithms and Automation" and "How Misinformation Spreads".

Some of the highlights from the literature review:

- SPREAD, PREVALENCE:

  - Bruns, Axel; Harrington, Stephen; Hurcombe Edward (2020) traced the dissemination dynamics of rumours (on the links between the pandemic outbreak and the rollout of 5G technology), using CrowdTangle. In order to assess the likely spread, visibility and thus impact that particular posts achieved, they focused on the number of followers that the public pages, groups and verified profiles had at the time of posting. Authors highlight that following a page does not mean that a user would see every post from that page - the absolute number of Facebook users who saw a post (unique views) would be a better metric. Another metric they suggest is sharing and the potential reach in sharing (a post appearing mainly in pages and groups with a few hundred followers would be far less likely to reach a large audience than a post that appeared in pages and groups with millions or tens of millions of followers).

  - Allcott, H., Gentzkow, M., & Yu, C. (2019) measured trends in the diffusion of content from 569 fake news websites (sources) and 9540 fake news stories (content) on Facebook and Twitter. Metrics used: monthly Facebook engagements (defined as the sum of shares, comments, and reactions such as likes) and Twitter shares. The authors suggest that, ideally, exposure would be measured using data on views, but such data were not publicly available.

  - Fletcher et al. (2018) also focused on the sources of disinformation (actors, disinformation purveyors), measuring average monthly reach, total time spent with selected false news websites each month as compared to the time spent with selected news websites, and engagement as seen through the total number of comments, shares, and reactions generated

by false news outlets compared to the engagement generated by the most popular news brands in the same period and on the same platform. They also measured the type of reactions provoked by junk news (e.g. likes or anger, laughter…).

- AUDIENCE CHARACTERISTICS:

  o Freelon et al. (2020) examined the socio-demographic characteristics (more specifically race) as a predictor of disinformation engagement: they examined it for both the producers and consumers of IRA Twitter accounts (Russian government-funded "troll farm").

  o Marchal, N., Kollanyi, B., Neudert, L. M., Au, H., & Howard, P. N. (2020) explored trends in user interactions with junk and professional content (classified manually based on the criteria developed within this OII project) on Facebook. Using the CrowdTangle, they measured the volumes of interactions (comments, likes, and shares) with content produced by the eighteen most popular sources of junk and professional news in their dataset during the selected timeframe.

- AUDIENCE STUDIES:

  o Surveys, recruited panels, experiments: exposing samples of users to factual and false news (often just headline, image, by-line and source). In Martel, C., Mosleh, M., & Rand, D. G. (2021) participants were asked whether they would share an article on social media publicly. Those who shared at least one false article (71%!) were presented with corrective messages of different styles and depth to see how participants would react to them. In a similar effort, Humprecht et al. (2021) ran a survey on users' willingness to disseminate disinformation. Respondents were presented 3 disinformation claims and asked about their assessment of them, as well as their willingness to share them.

Audience studies are envisioned in the Guidance for Strengthening the Code, but to conduct surveys or experiments with representative samples in all the member states and on a regular basis would require significant resources, which, at the moment, are not made available. However, it remains crucial to explore possibilities for supporting and developing simulations and off-platform, as well as on-platform experiments in order to understand patterns of use and users' behaviour.

Structural indicators for the Code of Practice on Disinformation

Proposal

1. Structural Indicator: **Prevalence of disinformation**

A total number of contents identified as disinformation and harmful misinformation (to be kept separate in reporting). This should be clearly linked with the Transparency Centre where the signatories explain, in an accessible way, how they define disinformation and how they identify sources and content of disinformation (some methods may be legitimately kept from the public, not to provide too much information to bad actors – however, these decisions need to be justified and methods are still to be shared with the evaluators). The number of dis- and harmful misinformation contents should be contextualised with the total number of public contents disseminated on the service.

A random sample of public content weighted by views (10 000 views - but to be adapted to population size of a member state) in the monitored period, per member state and language (to estimate the prevalence of disinformation), including also:

§ reach (total unique views in the monitored period per member state)

§ engagement (total number of interactions - depending on the service in question: i.e. comments, shares, and reactions with disinformation in the monitored period per member state).

A sample of TOP N (indicative number: 500) pieces of disinformation in a country, using the following metrics:

§ reach (total unique views in the monitored period per member state)

§ engagement (total number of interactions - dependent on the service in question: i.e. comments, shares, and reactions with disinformation in the monitored period per member state).

2. Structural Indicator: **Sources of disinformation**

A total number of identified sources of disinformation. This should be clearly linked with the Transparency Centre where the signatories explain, in an accessible way, how they define disinformation and how they identify/detect sources/purveyors of disinformation. The signatories should distinguish the accounts and the users, as some users can create large number of accounts for a single operation. The number should be contextualised with the total number of accounts on the service.

A sample of sources of disinformation (for example, based on detected/flagged disinformation under SI-1) per member state, focusing on:

1) The originating sources: who was the first to publish a piece of disinformation
    a) reach of, exposure to, and engagement with their content
    b) the size of their network
    c) frequency of publication
2) Superspreaders: receiving the biggest reach/visibility
    a) reach of, exposure to, and engagement with their content
    b) the size of their network
    c) frequency of publication

Considering GDPR related issues, this sample does not need to be public but can be made available to the vetted researchers for the analysis of characteristics of purveyors of disinformation. Furthermore, when naming individual sources, or individual pieces of content from individual sources as disinformation, there is a risk of legal action from those named so. The interest here thus is not in names but in the characteristics of such sources.

The samples under SI-1 & SI-2 could serve for a third-party validation (e.g. these contents and sources can be compared to those detected by the EDMO fact-checking community, vetted researchers, EDMO local hubs, or other signatories of the Code who may engage in this task in an independent way).

Acknowledging that disinformation is a multidimensional phenomenon; that the 2022 Code itself considers disinformation to include misinformation, disinformation, information influence operations, and foreign interference (so, actors, content, behaviour, techniques and technologies); and knowing that in their policies platforms apply different approaches to disinformation, we suggested focussing in SI-1 and SI-2 (initially) on measuring the content and source dimension of disinformation. The main reason for this is that these are the most common dimensions considered in various platform policies against mis/disinformation and can be validated by fact-checkers. At the same time the broader efforts to measure the impact of manipulative tactics such as the coordinated and inauthentic behaviour carried out by specific disinformation actors should be taken into account in final evaluation of SI-1 and SI-2.

Additional information considered for indicators 1 & 2:

a. Defining disinformation could be done externally, e.g by the EDMO network of fact-checkers: the fact-checking organisations would identify pieces (content) and/or sources of disinformation and the platform signatories would provide data on the reach of, exposure to, and engagement with that content/sources in the monitored period. However, this approach would require significant additional resources for the fact-checking community if they are to identify disinformation in each member state on a regular base (for each monitoring period) and in various platforms - signatories of the Code.

b. The designation of which site is a “purveyor of disinformation” should ideally be made by neutral 3rd parties with no tie to the publisher and no financial stake in the advertising process. Whoever engages in this task should comply with the highest integrity and transparency standards and be available for an independent scrutiny.

c. As designing structural indicators is understood as a process, not as a product, we could use this process to increase transparency over platforms’ policies and strategies on disinformation. Thus, instead of forcing outside harmonisation (of definitions and metrics used) we could aim to achieve internal harmonisation (to the extent possible) through transparency. More specifically, in the pilot stage of testing the structural indicators platform-signatories would explain how they define and operationalise the definition of disinformation (bridging with the Transparency Centre) and provide data on the spread and prevalence of disinformation per MS in the monitored period with some flexibility on the most appropriate metric for their service(s).

d. It is important to utilise the expertise of third party independent non-governmental organisations and research projects and institutions to assess the commitment level of signatories as an independent scrutiny mechanism.

3. Structural Indicator: **Audience of disinformation**

For each monitoring period, and per member states, relevant signatories should:

Pull a list of users (internally, not to share) who have had at least **X exposures** to disinformation, **providing in an aggregated and anonymised way**:

socio-demographic and psychographic characteristics of such disinformation audience; geolocation; history of platform use; frequency of platform use; frequency of exposure to disinformation; the size of network (friends/followers); whether they were following the source of disinformation or not when they were exposed to the content (algorithmic recommendations); probability that it is a bot or manifesting any other inauthentic behaviour.

Pull a list of users (internally) who have had at least X engagement with disinformation, **providing in an aggregated and anonymised way**:

socio-demographic and psychographic characteristics of such disinformation audience; geolocation; history of platform use; frequency of platform use; frequency of exposure to disinformation; the size of network (friends/followers); whether they were following the source of disinformation or not when they were exposed to the content (algorithmic recommendations); probability that it is a bot or manifesting any other inauthentic behaviour.

To contextualise, relevant signatories should also pull a list of users (internally) who have had at least **X exposure and/or engagement** with the most popular news brands, **providing in an aggregated and anonymised way**:

socio-demographic and psychographic characteristics of such news brands audience; geolocation; history of platform use; frequency of platform use; the size of network (friends/followers); whether they were following the news brand or not when they were exposed to its content (algorithmic recommendations).

NOTE: For validation, the process of studying disinformation audiences should be opened up to the independent authorities and qualified and vetted external researchers. There should be sufficient funding ensured for such research, and for building capacities of researchers and research institutions in parts of Europe where they have less opportunities and skills to engage in such research.

4. Structural Indicator: **Demonetization of disinformation**

A viable approach would assess for each monitoring period, and per member state, the monetisation strategies used by purveyors of disinformation, the revenues gained in a monitored period (transaction euro amounts) with different monetisation strategies, as well as the reach of, exposure to and engagement with disinformation content that was in a monetisation program.

While acknowledging that there are multiple ways of monetising disinformation, due to concerns of feasibility, first attempts to assess this indicator might be limited to advertising. The activity could also include a list of companies that facilitate the monetisation of disinformation on online platform services, assess the total number of monetized contents identified as disinformation and their share in overall monetised contents (relying on the sample used in the first structural indicator on prevalence). Ideally, access to unredacted sellers.json files would enable the identification of purveyors of disinformation that utilise platforms advertising services.

5. Structural Indicator: **Collaboration and investments in fact-checking**

Overall availability of fact-checking organisations in a member state; the extent to which platform signatories collaborate with fact-checking organisations per a member state; and funding by platform signatories for fact-checking per member state in a monitored period.

For each monitoring period, and per member state, the relevant signatories should provide data on:

- collaboration with fact-checkers **to be contextualised with an overall availability of fact-checking organisations in that state**
- funding provided to each fact-checking organisation with which they established a collaboration:
    - Alternatively, this can be provided as a total funding that a platform signatory invested in collaboration with fact-checking organisations in a monitored period per member state + each fact-checking organisation reporting the share of annual budget received from each platform signatories of the CoP from which they receive compensation for work

For each monitoring period, and per member state, relevant signatories should provide data on:

- Reach of and engagement with fact-checks

NOTE: This can be contextualised with the reach of and engagement with disinformation and with the most popular news brands in the country in the same period (SI-1).

For each monitoring period, and per member state, relevant signatories should pull a list of users who have had at least **X exposure and engagement** with **fact-checks**, providing in an aggregated and anonymised way:

- socio-demographic and psychographic characteristics of such disinformation audience; geolocation; history of platform use; frequency of platform use; the size of network (friends/ followers); whether they were following the source of disinformation or not when they were exposed to the content (algorithmic recommendations).

NOTE: This can be contextualised with characteristics of the audience of disinformation and the audience of the most popular news brands in the country in the same period (SI-3).

6. Structural Indicator: **Investments in the overall implementation of the Code**

For a monitoring period, and per member state, the relevant signatories should provide data on:

- **Total financial resources invested to meet the commitments and objectives set under the Code**
- Human resources invested to meet the commitments and objectives set under the Code

Rationale: this is in line with the policy indicators by Furceri and Mourougane (2010), and can serve to evaluate the commitment to implementing the policy by examining the resources invested by the signatories.

These indicators are to be assessed based on platform data collected in each monitoring period (every 6 months for VLOPs and once a year for other signatories), ideally both per member state (MS) and per official languages of the EU. The indicators are composite and designed in a way that enables them to evolve with time. In addition, to enable a comprehensive assessment, where possible, indicators are triangulated: the reach of and engagement with disinformation is contextualised with the reach of and engagement with most popular news brands in a country, as well as with the reach of and engagement with fact-checks. A similar triangulation is applied to the understanding of audiences of disinformation, where the audiences of the most popular news brands and audiences of fact-checks are included in the assessment. The data are to be sourced from major online platforms – who are signatories of the Code of Practice on Disinformation – and need to be analysed by independent researchers.

Conclusions

This paper outlines the process and the proposal for an initial set of approaches and metrics towards building Structural Indicators with an aim to examine the disinformation phenomenon and the effectiveness of the Code in suppressing it both in individual EU member states and in the EU as a whole. Adoption and implementation of Structural Indicators is one of the commitments contained in the Code of Practice on Disinformation, where Signatories recognise the importance of assessing the impact and success of the Code and thus commit to work together with the Task-force as well as with EDMO, ERGA, and relevant third-party experts to develop a first set of such indicators. It is further acknowledged in the Code that Structural Indicators need to first be tested and where necessary adjusted over time.

The problem of disinformation is often a symptom rather than a disease itself and a way more complex than the indicators outlined in this initial proposal capture. There are many components of this multifaceted problem that should be considered and studied, including the automation and artificial intelligence in producing and amplifying fakes. This proposal contains a matrix of initial and simpler areas and measuring approaches to test and pilot towards a more comprehensive methodology that should include evaluation of techniques and technologies in producing and amplifying fakes, as well as the impact of platform business models on the extent of the problem. This gradual approach is taken considering: (1) policy needs to adopt and test initial set of feasible indicators within a framework of the self-regulatory Code of Practice, which can later be further developed in its scope, methodology, and model of implementation; and (2) that at the moment empirical research on online disinformation in Europe is limited and there are no systematic and cross-country comparable insights on how the problem evolves in its various dimensions. The proposal described in this paper should thus be seen as a minimum and first step in what should be a wider and more systematic attempt to monitor disinformation and related policy effectiveness in Europe. In parallel with building a methodology for Structural Indicators.

Structural Indicators in the context of the Code should be developed and validated in cooperation with a wider array of stakeholders and experts, as the methodological choices on what to measure (and what to exclude) shape the quality and reliability of our understanding of where the problems are and what their degree is. Furthermore, the findings and indications deriving from the application of Structural Indicators per member state should be read in relation to the contextual factors in the given state, such as the levels of media literacy, trust in media, the extent and purpose of social media use, and the availability of specific policies against disinformation. While some of these contextual factors are available from comparative research projects, the available data is still not sufficient and, in general, research on disinformation in Europe, as well as the expertise among key stakeholders should see further development.

References

Alaphilippe, Alexandre (Brookings, 27 April 2020) Adding a 'D' to the ABC disinformation framework <https://www.brookings.edu/techstream/adding-a-d-to-the-abc-disinformation-framework/>

Allcott, H., Gentzkow, M., & Yu, C. (2019). Trends in the diffusion of misinformation on social media. *Research & Politics*, 6(2), 2053168019848554.

Altay, S., Kleis Nielsen, R., & Fletcher, R. (2022). Quantifying the "infodemic": People turned to trustworthy news outlets during the 2020 coronavirus pandemic. Journal of Quantitative Description: Digital Media, 2. https://doi.org/10.51685/jqd.2022.020

Bruns, A., Harrington, S., & Hurcombe, E. (2020). 'Corona? 5G? or both?': the dynamics of COVID-19/5G conspiracy theories on Facebook. *Media International Australia*, *177*(1), 12–29.

ERGA (2020). Notions of disinformation and related concepts. https://erga-online.eu/wp-content/uploads/2021/03/ERGA-SG2-Report-2020-Notions-of-disinformation-and-related-concepts-final.pdf

Fletcher, R., Cornia, A., Graves, L., & Nielsen, R. K. (2018). Measuring the reach of" fake news" and online disinformation in Europe. *Australasian Policing*, 10(2).

François, C. (2019). Actors, behaviours, content: A disinformation ABC (Transatlantic High Level Working Group on Content Moderation Online and Freedom of Expression Series). Annenberg Public Policy Center, University of Pennsylvania; Annenberg Foundation Trust, Sunnylands. Institute for Information Law, University of Amsterdam.

Freelon, D., Bossetta, M., Wells, C., Lukito, J., Xia, Y., & Adams, K. (2022). Black trolls matter: Racial and ideological asymmetries in social media disinformation. *Social Science Computer Review*, 40(3), 560-578.

Gillespie, T. (2022). Do Not Recommend? Reduction as a Form of Content Moderation. *Social Media+ Society*, 8(3), 20563051221117552.

Humprecht, E., Esser, F., Aelst, P. V., Staender, A., & Morosoli, S. (2021). The sharing of disinformation in cross-national comparison: analyzing patterns of resilience. *Information, Communication & Society,* 1-21.

Leerssen, P. (2023). An End to Shadow Banning? Transparency rights in the Digital Services Act between content moderation and curation. *Computer Law & Security Review*, 48, 105790.

Martel, C., Mosleh, M., & Rand, D. G. (2021). You're definitely wrong, maybe: Correction style has minimal effect on corrections of misinformation online.

Marchal, N., Kollanyi, B., Neudert, L. M., Au, H., & Howard, P. N. (2020). Junk News & Information Sharing During the 2019 UK General Election. arXiv preprint arXiv:2002.12069.

Marwick, A. E., & Lewis, R. (2017). Media manipulation and disinformation online

Authors

NENADIĆ, Iva

Centre for Media Pluralism and Media Freedom / Robert Schuman Centre

Iva.Nenadic@eui.eu

BROGI, Elda

Centre for Media Pluralism and Media Freedom / Robert Schuman Centre

Elda.Brogi@eui.eu

BLEYER-SIMON, Konrad

Centre for Media Pluralism and Media Freedom / Robert Schuman Centre

Konrad.Bleyer-Simon@eui.eu