European Digital Media Observatory

# Description and Analysis of Relevant Emerging Research Topics

# /

# Enhancing Content Reliability by Prominence: Indicators for Trustworthy Online Sources

M 12



This project has received funding from the European Union under Contract number: LC-01464044

| | |
|---|---|
| **Project number:** | SMART 2019/1087 |
| **Project Acronym:** | EDMO |
| **Project title:** | European Digital Media Observatory |
| **Start date of the project:** | 01/06/2020 |
| **Duration of the project:** | 30 |
| **Project website address:** | **https://edmo.eu/** |
| **The deliverable has been elaborated by:** | Centre for Media Pluralism and Media Freedom / European University Institute / EDMO Task 5 |



This project has received funding from the European Union under Contract number: LC-01464044

# Table of Contents





This project has received funding from the European Union under Contract number: LC-01464044

# Description and Analysis of Relevant Emerging Research Topics

**Enhancing Content Reliability by Prominence: Indicators for Trustworthy Online Sources**[1]

## 1 Executive Summary

*In the current online information environment, it has become increasingly complicated for users to define what information to trust: the amount of available content online exceeds the time and attention that users can invest in analysing what source is reliable and what is not. This paper seeks to analyse a topic that, within many facets, is becoming increasingly relevant as an element of present and future media policy. Moreover, it aims to inform and guide the EU approach to tackle disinformation online. Individual choices are driven both by technology-based and policy-based curation, which can limit human autonomy and freedom of choice. Therefore, new policies should take into account measures to enhance exposure to a diversity of trustworthy quality content.*

*Considering the scope of the newly instituted European Digital Media Observatory (EDMO), and its purpose of contributing to the debate on new policies to fight disinformation, the analysis of this paper concentrates on the measures foreseen by the Code of Practice on Disinformation as regards online trustworthiness and the ways to implement them (considering both its 2018 text and the guidelines to strengthen it). The Code of Practice on Disinformation foresees an important role for the promotion and prioritisation of trustworthy content by large online platforms. 'Trustworthiness' is explicitly mentioned in two pillars of the Code: Pillar A (scrutiny of ad placements) highlights the importance of indicators of trustworthiness when identifying the sites where advertisement can be placed without (unintentionally) monetising purveyors of disinformation; Pillar D (empowering consumers) mentions indicators of trustworthiness as the basis of content prioritisation and media literacy measures. The European Commission is therefore looking for indicators of trustworthiness that can provide the basis for platforms for improving the findability of trustworthy content sources and for 'diluting' the visibility (downranking) of their non-trustworthy counterparts. These indicators of trustworthiness should be based on objective criteria and endorsed by news media associations, in line with journalistic principles and processes. So far, there are four prominent projects that are often mentioned in the context of defining online content's trustworthiness in*

[1] The report was authored by Konrad Bleyer-Simon and Elda Brogi. konrad.bleyer-simon@eui.eu



This project has received funding from the European Union under Contract number: LC-01464044

*the EU: The Trust Project, the Credibility Coalition, the Journalism Trust Initiative and the NewsGuard browser extension.*

*This report will provide an overview of the indicators identified and listed by these projects. It is, for example, a common property of these projects that they look at trustworthiness as a requirement that is attached to the content creator, rather than to the content itself; moreover, they treat trustworthiness mainly as a requirement that is especially attached to news outlets, among possible content creators. While linking trustworthiness to content creators is indeed the best way to provide the basis for ex ante measures, the current focus on news media only allows for a narrow application.*

# 2 Introduction

Our societies' increased access to the internet, as well as the revolution in content production and distribution offers users an abundance of information, more than they can assimilate, it is argued. Barriers to entry have almost disappeared, thus anyone can act as a content creator – and share text, video or audio commentary with a (possibly) large audience. In this new online environment, news and information are largely consumed through intermediaries, especially the big social media or online search platforms. This constellation means that users are regularly confronted with new content sources that they have not been familiar with so far.

Amidst the rising propagation of news is the spread of disinformation. Disinformation is understood as verifiably false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, information which may cause public harm (as defined by the European Commission's Communication on tackling online disinformation[2]). In the new online context, it has become increasingly hard for users to determine what information they can trust. This document thus provides a description and analysis of a relevant emerging research topic, namely that of the trustworthiness of content online.

To tackle the above-mentioned issues, the Code of Practice on Disinformation enacted in 2018 foresees an important role for the promotion and prioritisation of trustworthy content by large online platforms. Trustworthiness is explicitly mentioned in two pillars of the Code. Pillar A (scrutiny of ad placements) highlights the importance of indicators of trustworthiness when identifying the sites where advertisement can be safely placed; Pillar D (empowering

2 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions (2018): Tackling online disinformation. A European Approach (COM/2018/236).



This project has received funding from the European Union under Contract number: LC-01464044

consumers) mentions indicators of trustworthiness as the basis of content prioritisation and media literacy measures. The European Commission's 2021 Guidance on Strengthening the Code of Practice on Disinformation reiterates the need for these indicators.

## 2.1 The Code of Practice on Disinformation

The Code of Practice on Disinformation was the first major initiative developed at EU level to fight disinformation. After agreeing on the Code, signatories (among them Google, Facebook, Twitter and TikTok) were required to regularly report on the actions taken in order to further the goals that were identified,[3] but so far, they do not seem to be complying with expectations – especially with respect to the actions of empowering consumers, which is related to the pillar that encourages the 'findability of trustworthy content' in the online information environment. Overall, there is a problem with the absence of standards for its evaluation and for reporting, lack of oversight on the compliance, lack of sanctions for non-compliance, and lack of data against which to check the statements and reports created by platforms themselves.

With the 2020 Democracy Action Plan, the Commission started steering the efforts to turn the Code of Practice on Disinformation into a co-regulatory framework, which introduces obligations and requirements for accountability on online platforms. In addition, the Digital Services Act (DSA) proposal aims to establish a powerful framework for transparency and clear accountability, which enables oversight over online platforms, especially those referred to as 'very large online platforms' such as Facebook or Google, which are in a dominant position in most EU markets. As a follow-up to these initiatives, the Commission issued guidance on enhancing the Code of Practice in 2021, mainly by creating a more robust framework for monitoring its implementation by the signatories.

Under its task 5, led by the European University Institute (Centre for Media Pluralism and Media Freedom), the European Digital Media Observatory (EDMO) supports research and analysis on policy activities to tackle disinformation, which includes the provision of key elements to enable continuous monitoring and independent assessment of platform activities to limit the impact of disinformation. The task includes working on a methodology, defining standards and identifying structural key performance indicators (KPIs) that allow for the assessment of the Code's impact on the spread of online disinformation[4].

EDMO strives to develop and test a methodology that is: inclusive (considering current and potential future signatories of the Code); feasible (capable of being implemented on a regular basis under different forms of regulatory regime); mixed-methods-based (combining

3 Platforms are asked to report on the implemented measures under the Code. The 2019 reports can be found on the European Commission's website. https://digital-strategy.ec.europa.eu/en/news/annual-self-assessment-reports-signatories-code-practice-disinformation-2019

4 A proposal by CMPF on this methodology is forthcoming.



This project has received funding from the European Union under Contract number: LC-01464044

quantitative and qualitative indicators); and data-informed (relying on an increased transparency of platforms and functional data access).

## 2.2 What is 'Trustworthiness'?

In April 2018, the European Communication published the communication 'Tackling Online Disinformation: a European Approach', which highlighted the importance of fostering 'credibility of information by providing an indication of its trustworthiness, notably with the help of trusted flaggers, and by improving traceability of information and authentication of influential information providers'. This attempt was followed by the EU's Code of Practice on Disinformation, which identified a number of actions for its signatories in order to address the challenges posed by disinformation.

In the Code, the term 'trustworthiness' refers first and foremost to content sources, and is often mentioned in connection with ownership transparency and 'verified identity'. Indicators of trustworthiness are expected to provide the basis for platforms that seek to improve findability of trustworthy content sources and 'dilute' visibility (downranking) of their non-trustworthy counterparts. According to the Code's text, indicators of trustworthiness should be based on objective criteria and endorsed by news media associations, in line with journalistic principles and processes. They are expected to be complemented with information from fact-checkers. The 2021 Guidance adds that these indicators should be developed by independent third parties 'in collaboration with the news media, including associations of journalists and media freedom organisations, as well as fact-checkers'.

News media is in the focus of most of the documents assessing the Code that are published or commissioned by the European Commission. The account by Valdani, Vicari and Associates (VVA – an independent contractor working for the Commission), for example, refers to the concepts of 'trustworthy news' and 'trustworthiness of sources'. 'Ranking', 'prioritising' or 'pushing up' trustworthy content is often mentioned in these documents and assessments as methods that make the best use of the indicators. This approach also makes it very likely that, (for the sake of feasibility), the focus has to be on *ex ante* measures on the level of content sources, with an emphasis on news media. Moreover, it is mentioned that the indicators of trustworthiness need to be designed in a way that they can feed into algorithmic evaluation.[5]

5 Already in the chapter on "Purposes" the Code of Practice mentions two trustworthiness-related efforts:

> (viii) Ensure transparency with a view to enabling users to understand why they have been targeted by a given political or issue-based advertisement, also through indicators of the trustworthiness of content sources, media ownership and/or verified identity.
>
> (ix) Dilute the visibility of disinformation by improving the findability of trustworthy content.

In addition, pillar A of the Code (scrutiny of ad placements) mentions:



This project has received funding from the European Union under Contract number: LC-01464044

As such, we can say that, in our context, trustworthiness, (often connected with credibility), is a term that refers to the source or publisher of a piece of information. A publisher of information can be regarded as trustworthy (or credible) when the users' chance of being exposed to false or misleading content (dis- but also misinformation) by that source is relatively low. Moreover, it is expected that a trustworthy publisher has a procedure in place to make sufficient and timely corrections, for any case wherein false or misleading content is suspected. A trustworthy source of information is, generally, transparent in its ownership, authorship and sourcing of information. In addition, it holds procedures in place to clearly label advertisement and monitor paid content, as well as to separate fact from opinion.

# 3 How to Make the Online Media Landscape Trustworthy?

Some important works in the social sciences argue that we rely heavily on trust to deal with the complexity of the world in which we are living (Luhmann, 1980). Creating an environment of trust is, therefore, in the self-interest of actors; economic history shows that trust is a precondition for functioning commercial contracts and economic prosperity (Fukuyama, 1995; Khalil 2003). Trustworthiness and the related concept of trust are also relevant challenges of the internet, where the demand for quick, up-to-date and freely available information has created a number of tools/means for information exchange and access, from chats and online forums, to crowdsourced encyclopedias and comment sections of news media. The literature on how online content is assessed is often focused on the perceived density of producers' (social) networks and other proxies that determine the content creators' or sources' trustworthiness. Many scholars suggest trustworthiness can be ascertained to some extent by evaluating and ranking (the performance of) content producers/sources. Producers' previous activity (thus, the trustworthiness of previous contents) and their connectedness can provide a pattern which would allow an educated guess related to the risks posed by their future communications (Pan and Chiou, 2011; Stavri et al., 2003). However, in a recent study by the Oxford Internet Institute and the Reuters Institute (Toff et al. 2021) highlights that the content most trusted by audiences is often not the content that underwent the best quality checks. Thus,

> The signatories recognise that indicators of trustworthiness and information from fact checking organisations and the new independent network of fact checkers facilitated by the European Commission upon its establishment can provide additional data points on purveyors of disinformation.

Pillar D (empowering consumers) adds:

> The Signatories of this Code recognise the importance of diluting the visibility of Disinformation by improving the findability of trustworthy content (...)
>
> (...) transparency should reflect the importance of facilitating the assessment of content through indicators of the trustworthiness of content sources, media ownership and verified identity. These indicators should be based on objective criteria and endorsed by news media associations, in line with journalistic principles and processes.



This project has received funding from the European Union under Contract number: LC-01464044

it is important not to confuse the trust of the audiences with the trustworthiness of content (the latter being the focus of our paper).

A given content source can be seen as trustworthy when our trust in the content or its producer is well placed, meaning that the consumers' risk to be confronted with harmful or misleading content is relatively low.

In his book on *Trust and Trustworthiness* Russell Hardin (2002:29) identifies three forms of inducements that can lead an actor to be trustworthy in future interactions, be they internal, external or comprised of mixed inducements. Internal inducement refers to behaviour driven by the actor's conscious decision to do what is right; external inducements refer to legal or institutional constraints, while mixed inducements are a combination of the two. The power of internal inducements, for example, is visible when considering that it is in the interest of content creators not to violate the trust of their consumers (unless they operate with the purpose of spreading disinformation). In fact, sharing disinformation may inflict the kind of harm on a media outlet that is hard to fix (Altay, Hacquin and Mercier, 2020). Indicators of trustworthiness are external inducements: a third party evaluates news sources and attaches labels to them. This, in turn, guides audiences' interactions with them and provides an urge to produce more reliable content.

These considerations can be relevant when platforms have to make decisions that aim to contribute to a trustworthy online ecosystem. The Commission Staff Working Document titled 'Assessment of the Code of Practice on Disinformation - Achievements and areas for further improvement' points out that online platforms have supported the development of projects designing trustworthiness and credibility indicators, such as the *Trust Project*, the *Credibility Coalition* or the *Journalism Trust Initiative*,[6] but an evaluation by VVA highlighted that there is no detailed information available on the integration of these indicators in platforms' search services and recommender systems[7].

Nevertheless, the Commission's Staff Working Document lists a number of efforts from platforms to prioritise trustworthy content sources, noting that 'signatory platforms took a broad range of actions including investments in technology to give prominence to trustworthy information sources on their content ranking and recommender systems, while making it easier for users to find diverse perspectives about topics of public interest'[8]. However, there is no

6 The Commission's Guidelines also mention the *Global Disinformation Index* (GDI) as a possible source of indicators, however, this index relies mainly on the Journalism Trust Initiative for input.

7 Facebook, for example, announced that it incorporates the Trust Project's trust indicators in the publisher's info of news media pages, but these are currently not visible.

8 From the Working Document: "For instance, Facebook notifies users when they share content that was fact-checked and rated as 'false' or 'mixture' and makes it easier for users to view information, via a Context Button, about websites and publishers they see on Facebook. Via the "Full Coverage" feature in Google News, users can access context and diverse perspectives about news stories from a variety of publishers, and in September 2019 Google announced ranking updates that give more prominence in Search to articles identified as significant



This project has received funding from the European Union under Contract number: LC-01464044

mention of the criteria the platforms use to determine a trustworthiness, aside from recommendations by fact checkers, and the knowledge that Microsoft has a partnership with large, 'vetted' sources. The commitment of platforms to implement measures to promote trustworthy content is also emphasised in the Code of Practice on Disinformation and its Annex 2 on best practices.[9] It mentions four examples of best practice related to trustworthiness:

1. Facebook's trusted source strategy refers to policies that prioritise news content from sources the community rate as trustworthy. To determine what sources are 'trustworthy' Facebook has asked users in a survey about their familiarity with specific websites and whether they trust those or not. In the EU, this initiative was available in Germany, France, Italy and Spain, but there is no detailed information on how it works, how reliable it is and whether it is still active.
2. Facebook's user option to report what users see as false news.
3. The Mozilla Information Trust Initiative has led to the creation of a misinformation database, which, so far, lists 89 research articles on the misinformation topic.
4. Third-party ad-verification companies in the advertising industry authenticate online content via keywords, metadata and URL analysis to make sure that advertisement does not end up next to problematic content (e.g. Trustworthy Accountability Group, Digital Ad Trust in France).

The VVA Study for the 'Assessment of the implementation of the Code of Practice on Disinformation' also emphasises the importance of prioritising, and highlights the need to involve a range of actors (fact-checkers, publishers, etc.) in determining what content is trustworthy. However, in its interviews with traditional media organisations, VVA encountered problems, as the media themselves were unable to provide useful criteria, which might indicate that many well-known media brands would fail an assessment due to the rigidity of such criteria.

> A key objective seems to be to ensure that users get to see the highest quality and most relevant content first. In order to do so and to improve the implementation of the commitments under Pillar [D], platforms need to work with publishers, fact-checkers, and other content creators to better label the trustworthiness of different kinds of content. For instance, as suggested by stakeholders connected to the Sounding Board

original reporting, which will stay longer in a highly visible position; The "Microsoft News" service partners with over 1.000 news sources worldwide, which are all vetted by Microsoft to ensure that the service only shows licensed reputable content. At the same time, the platforms' collaboration with the fact-checking community has provided users with additional possibilities to critically assess information accessed online, and enabled the development of new features giving users more contextual information about fact-checked websites or webpages, with the aim to reduce the spread of false narratives online."

9 It writes: "Relevant Signatories commit to invest in products, technologies and programs such as those referred to in Annex 2 to help people make informed decisions when they encounter online news that may be false, including by supporting efforts to develop and implement effective indicators of trustworthiness in collaboration with the news ecosystem."



This project has received funding from the European Union under Contract number: LC-01464044

> in interviews, platforms could engage more with traditional media to develop transparency and trustworthiness indicators for information sources (which falls under Commitment no. 7), which can then be used to feed content ranking algorithms, eventually providing users with access to a plurality of credible information sources. A good practice example that can be mentioned in this regard is the Trust Project which is a consortium of top news companies led by an award-winning journalist. [The project] is developing transparency standards that help consumers easily assess the quality and credibility of journalism. Several platforms (i.e. Google, Facebook and Bing) are involved in the project.
>
> One specific solution suggested by interviewees from traditional media organisations to achieve [the above goal] is to 'push up' content from so-called 'trusted information providers', which relates to Commitment Number 8 of the Code. However, the platforms pointed towards the difficulties in distinguishing such outlets as well as the fact that not all content published by these outlets is necessarily trustworthy (e.g. even trusted information providers can publish click bait content). According to the platforms, a definition is not the only thing needed of these trusted information providers, but also a body that defines which outlets comply with these criteria. Indeed, the interviewees from the traditional media organisations did not manage to define trusted information providers and their preference for this solution might be motivated by their own presumed status as such.

The Commission indicates a preference for *ex ante* measures, e.g. when recommending the following; 'Ex ante approval by ad-placement service providers of websites selling advertisement space, possibly based on trustworthiness indicators agreed with advertisers (a 'white list' approach).' This *ex ante*, white list approach is in line with the Code's attempts to classify content producers/content sources as trustworthy and untrustworthy. There could also be scope for intermediate categories that would, for example, consider the lack of sufficient information on a given content creator, consider whether previous instances of publishing misleading content was unintentional or that the missing of safety standards and relevant safety procedures could increase the risk of publishing false or misleading content.

The European Commission's 2020 *Democracy Action Plan* highlights the importance of the Commission supporting self-regulatory initiatives promoting professional standards, reinforcing the need for indicators of trustworthiness in this endeavour. 'This includes the development of structural and procedural indicators for trustworthiness by the media sector that promote compliance with professional norms and ethics. The Commission co-funds for example the implementation of the Journalism Trust Initiative (JTI) **https://jti-rsf.org/en**.' The Democracy Action plan also refers to the Code of Practice as follows:

> [T]he strengthened Code of Practice will aim to address the following objectives: (...) support adequate visibility of reliable information of public interest and maintain a plurality of views by developing accountability standards (co-created benchmarks) for



This project has received funding from the European Union under Contract number: LC-01464044

recommender and content ranking systems and providing users with access to indicators of the trustworthiness of sources; (...) Alongside media and other relevant actors, fact-checkers have a specific role in the development of trustworthiness indicators and the scrutiny of ad placement.

# 4 Four Trustworthiness Indicators: An Overview

The documents assessed and the platforms' communications mention four major projects that can be seen as viable sources for elaborating indicators that help assess which sources to prioritise as trustworthy: these are the *Trust Project*, the *Credibility Coalition*, the *Journalism Trust Initiative* and the *NewsGuard* browser extension.[10] In the following pages, we will provide a brief overview of their work and list their indicators.

## 4.1 The Trust Project

The Trust Project (which was funded, over the years, by Google and Facebook) has interviewed news consumers, and used their input to create eight 'Trust Indicators' and the accompanying questions that a consumer of news must ask when determining whether a given content can be trusted. In this case the assessment looks at both at the content and the producer level.

1. Expertise: the journalist is an expert

   *Who wrote/created this? Do they have a good professional reputation? Are they reporting on an area they normally focus on?*

2. Labels: the purpose of the story is clear

   *Why has this been created? Does it have a clear opinion, or is it impartial? Is this sponsored, or is it advertising something? Is the purpose explicitly indicated?*

3. References: you can find and access the sources

10 Although these projects focus mainly on content sources, still, indicators to determine trustworthiness of content should not be completely disregarded, as content-level assessment can also help assess the trustworthiness of the given content source / content creator (See the Climate Feedback Process to assess widely-shared claims about the environment and the Claim Review Schema utilized by many members of the Poynter International Fact-Checking Network). The World Federation of Advertisers has launched another related project, the Global Alliance for Responsible Media (GARM) to create common standards for advertisers and media to determine which websites are safe to place advertisements.



This project has received funding from the European Union under Contract number: LC-01464044

*What's the source? For investigative, in-depth, or controversial stories, do we have access to the sources behind the claims? Can you find another source to back up what is being said?*

4. Local: the journalist uses local knowledge

*Was the reporting done with in-depth knowledge about the local situation or community? Was the journalist on the scene? Does the story let readers know when the news sources are local?*

5. Diversity: the story brings in many kinds of people

*What efforts and commitments does the newsroom put in place to bring in diverse perspectives? Are some communities included only in stereotypical ways, or even completely missing?*

6. Actionable Feedback: The news organisation allows readers to participate

*Can we participate? Can we give feedback? Does the news site invite and acknowledge contributions from the public?*

7. Methods: We can tell the process used to make the story

*How was it put together? How long did it take to create? Who else was involved in the process?*

8. Best Practices: The journalist or news organisation explains their ownership and standards

*Does the journalist or organisation have a list of rules that they have to follow? How do they check their facts? Who funds them? What is the organisation's mission and its priorities? Does the journalist or organisation make corrections if errors are discovered? Do they have a commitment to ethical/diverse/accurate reporting and how do they show they are sticking to the rules?*

The application of the Trust Project is described on its website as follows:

> The Project's Trust Indicators, easily recognized anywhere, are both available to the public on news pages and easily read by machines in the code that produces those pages. Google, Facebook [and] Bing use the indicators and their associated machine-readable signals in various ways to enhance their ability to differentiate reliable, trustworthy journalism from other information, and continue to develop new uses. Other organizations such as NewsGuard, a news literacy company; Nuzzel, a news



This project has received funding from the European Union under Contract number: LC-01464044

aggregator; and PEN America [a writers' association] use the Trust Indicators to help the public find trustworthy news with authority. Richard Gingras, vice president for news at Google, said, "The Trust Project's required disclosures and clear definitions should help the public – and Google's systems – recognize and value quality journalism." Gingras says the Trust Indicators can be helpful in guiding Google's own internal evaluations of the quality of search results. Facebook uses the Best Practices Trust Indicator in its process to index news Pages, among other uses, and Bing uses Trust Indicator labels to display whether an article is news, opinion or analysis, providing information that people need to understand an article's context. The three technology companies rely on [founder, Sally] Lehrman and the Trust Project consortium as an expert advisor in their effort to elevate accurate, dependable news in search and social media.

## 4.2 Credibility Coalition

Funded by a similar set of donors, the Credibility Coalition is also working on criteria to define what content (and what content source) deserves trust. One of its research projects came up with a set of credibility indicators. The authors of the paper 'A Structured Response to misinformation: defining and annotating credibility indicators in news articles', have recommended 16 indicators that could help in the assessment of a given content's credibility. Eight of these indicators are focusing on the content, another eight are looking at the context. The content indicators assess whether the title is in line with the content, whether it is clickbait, whether the article included quotes from outside experts, whether the argument has logical fallacies, etc. The context indicators look at originality, the existence of fact-checking, the degree in which content is in line with sources, etc.

Most of these assessments require the work of trained people who check these criteria manually. Large-scale assessment and automation is only possible in some of the context indicators, such as reputation of citations (impact factor of the publication of the scientific study cited), number of ads on the site where the article is displayed (which was treated as an indication that the site monetises disinformation), and number of 'social calls' (requests to share the content on social media).

## 4.3 Journalism Trust Initiative

The Journalism Trust Initiative (JTI) has developed a complex set of indicators that help determine whether a newsroom is trustworthy, namely:



This project has received funding from the European Union under Contract number: LC-01464044

1. Basic requirements on Media's Identity (looking at the legal entity name, contact details and identifiers, description of media outlet, a list of all distribution channels and URLs, newsrooms' response to safety concerns, location, founding date)
2. Editorial mission (the existence of an editorial mission statement)
3. Public Service Media (description of public service media mission, governance and independence)
4. Disclosure of Type of Ownership (privately held, state or public owned, publicly traded company, other)
5. Requirements on Owners' Identity (names of owners and board members, contact details of direct and indirect owners, names of shareholders, percentage of shareholdings; the indicator also includes some limited criteria for member-owned media outlets)
6. Disclosure of Identity of the Management Team and its Location (management directory, location of branches and offices)
7. Disclosure of Editorial Contact Details (social media, newsroom contact details, consumer service contact details)
8. Disclosure of Revenue Sources and Data Collection (sources of revenue, disclosure of data collection; which personal data are processed, how and for what purpose)
9. Accountability for Journalism Principles (description of editorial guidelines, purpose of guidelines, guidelines and journalism principles, conflicts of interest)
10. Accuracy (information on the process for ensuring accuracy, process review, statistics and external content, identification of journalists, agencies, location reporting, automatically reported content, algorithmic dissemination and curation, treatment of explicit content)
11. Responsibility for Content Provided by the General Public (how newsrooms deal with user generated content / eyewitness news, editorial guidelines for UGC / eyewitness news, opinion guidelines)
12. Responsibility for Sources (newsrooms describe the ways in which they protect sources, including their anonymity, privacy rights, guidelines on sources and newsroom independence, diversity of sources)
13. Professionalism for Affiliations (sponsored content policies, sponsored content indicators, separation of news and opinion)
14. Internal Accountability (dealing with inaccuracies, publishing corrections, contact and process for complaints, internal process for complaints, independence of ombudsperson, powers of ombudsperson)
15. External Accountability (external oversight, compliance with external accountability, absence of external oversight, contact details of external accountability bodies, other associations)
16. Professionalism in the Media Outlet (recruitment and training, working conditions, contact policy and labour relations, staff welfare)
17. Training (training in editorial guidelines, continuous training, support and advice)



This project has received funding from the European Union under Contract number: LC-01464044

18. Publication of self-assessment (self-assessment available for the general public, in a machine-readable form, so that advertisers, platforms, etc. can use them for the assessment of media trustworthiness)

These indicators add up to a standard of trustworthiness, which was developed in cooperation with the European Committee of Standardization (CEN). This standard is defined as 'a normative, non-proprietary benchmark for internal and external assessment of media outlets' which covers the 'institutional and process level of journalistic production' (ie. looks at the characteristics of media companies and the internal procedures of content production).

This process can be certified by third-party audits to make the self-assessment credible. The self-assessment is intended to help governments and regulators in making decisions on subsidies, media development actors in providing financial support, advertisers in making informed decisions about the placement of their ads. The questionnaire is designed in a way that allows machine readability so that platforms can use it as the basis of their evaluation of a content producer's trustworthiness. In the long-run, JTI hopes to turn the CEN standard into a global International Organization for Standardization (ISO) standard.

The JTI's indicators are also used as the basis of the Global Disinformation Index, a project that provides disinformation risk ratings for news sites, promising a 'gold standard' to assess disinformation risks. While the indicators used by the Index are based on JTI's indicators, their assessment is supplemented by a blind review of randomly selected news content from the websites scrutinised.

### 4.4 NewsGuard

*NewsGuard* is a browser extension that assigns a red or green rating to websites, thereby signalling their trustworthiness to users. In the NewsGuard process, websites have to reach a score of 60 out of 100 in order to be listed as trustworthy. The scores are assigned based on the following (weighted) criteria:

Credibility:

- Does not repeatedly publish false content: The site does not repeatedly produce stories that have been found—either by journalists at NewsGuard or elsewhere—to be clearly and significantly false, and which have not been quickly and prominently corrected (22 Points. A score lower than 60 points gets a red rating).
- Gathers and presents information responsibly: Content providers are generally fair and accurate in reporting and presenting information. They reference multiple sources, preferably those that present direct, first hand information on a subject or event or from



This project has received funding from the European Union under Contract number: LC-01464044

credible second hand news sources and they do not egregiously distort or misrepresent information to make an argument or report on a subject (18 Points).

- Regularly corrects or clarifies errors: The site makes clear how to report an error or complaint, has effective practices for publishing clarifications and corrections and notes corrections in a transparent way (12.5 Points).
- Handles the difference between news and opinion responsibly: Content providers who convey the impression that they report news or a mix of news and opinion distinguish opinion from news reporting, and when reporting news, do not egregiously cherry pick facts or stories to advance opinions. Content providers who advance a particular point of view disclose that point of view (12.5 Points).
- Avoids deceptive headlines: The site generally does not publish headlines that include false information, significantly sensationalise, or otherwise do inaccurately report what is actually in the story (10 Points).

Transparency:

- Website discloses ownership and financing: The site discloses its ownership and/or financing, as well as any notable ideological or political positions held by those with a significant financial interest in the site, in a user-friendly manner (7.5 Points).
- Clearly labels advertising: The site makes clear which content is paid for and which is not. (7.5 Points)
- Reveals who is in charge, including possible conflicts of interest: Information about those in charge of the content is made accessible on the site (5 Points).
- The site provides the names of content creators, along with either contact or biographical information: Information about those producing the content is made accessible on the site (5 Points).

So far, *NewsGuard* is closest to being used as a tool that informs online platforms' users about the trustworthiness of websites they are visiting. Microsoft provides a free NewsGuard plug-in for the Microsoft Edge web browser (and an opt-in news rating feature for the Edge mobile application). Users can see 'NewsGuard ratings right next to links on search engines and social media feeds across all major platforms'.[11]

### 4.5 Compatibility of Indicators with the Code of Practice

In the Staff Working Document, the Commission indicates a preference for *ex ante* measures, e.g. when recommending the following option: '*Ex ante* approval by ad-placement service providers of websites selling advertisement space, possibly based on trustworthiness indicators agreed with advertisers (a 'white list' approach)'. This *ex ante* approach and white list, is in

11 More information can be found under this link: https://www.newsguardtech.com/edge/



This project has received funding from the European Union under Contract number: LC-01464044

line with the Code's attempts to classify content producers/content sources as trustworthy and untrustworthy .

Based on the above listed indicators elaborated by *NewsGuard*, the *Trust Project*, the *Journalism Trust Initiative* and, in the context of the *Credibility Coalition*, we will now highlight key indicators that are relevant in identifying trustworthy content sources. This is a short (and not necessarily complete) list:

1. Past conduct of publisher

- The content publisher has not been found to publish verifiably false information repeatedly

2. Sourcing of articles

- Diversity of sources used in published items
- Transparent sourcing of articles (references, hyperlinks, quotes from identified sources) / openness of methods used to acquire information
- Reliance on reader feedback
- Logical soundness of content published

3. Correction and labelling

- Timely correction and clarifications in case errors or inaccuracies were spotted
- Labelling of advertising and sponsored content / separation of fact and opinion / number of ads and calls to share content on social media

4. Clear indication of funders and content creators

- Disclosure of ownership and financing of media organisation
- Disclosure of authors, incl. contact details (email)

The approaches of *NewsGuard*, the *Credibility Coalition*, the *JTI* and the *Trust Project* can all contribute to the creation of an environment where users have easy access information from trustworthy sources. Some of these indicators can be checked automatically (e.g. existence of a masthead, owner information, as well as additional indicators, such as being registered with the country's media authority, or checking the average number of outside links, corrections, etc.) or providing the basis of self-reporting (such as the machine-readable, detailed questionnaire of JTI). Others require the active work of users and fact checkers (such as reporting suspicious content by users and flagging by fact-checkers).



This project has received funding from the European Union under Contract number: LC-01464044

# 5 Our Assessment of the Trustworthiness Indicators

## 5.1 Problems for media pluralism

Despite the possible solutions mentioned above, using these indicators as the single means for determining trustworthiness of content sources may create a media environment in which established players gain further competitive advantage, while new players face unprecedented barriers to entry. This may lead to serious problems for media pluralism and could distort the media market in a way that news players will find only limited access to the advertising market or other revenue sources. An over reliance on these indicators could silence diverging or non-mainstream voices, as has been seen in the past when alternative/non-mainstream newsrooms suddenly lost a sizable percentage of their readers, due to some tweaks in platform algorithms.

At discussions among stakeholders, representatives of publishers have also signalled that reporting about one's trustworthiness (or even auditing these reports) based on indicators like the ones developed by JTI or the Trust Project cannot be made mandatory. Thus, they argue, media outlets should not be labelled untrustworthy simply for not being party to such a project or initiative. Not to mention that the Code itself highlights that measures should be consistent with Article 8 of the European Convention on Human Rights (the right to respect private and family life), the fundamental right of anonymity and pseudonymity, and the proportionality principle – these could all be violated by overly stringent reporting requirements. In addition, the Code also highlights Article 10 of the European Convention on Human Rights (freedom of expression), as decisions on prioritisation might limit users' access to relevant ideas and information.

In its interview-based assessment, VVA highlighted that most representatives from traditional media organisations were unable to provide a working definition of what constitutes 'trusted information providers'. This implies that many outlets that enjoy a good reputation in society and among decision makers may fail to meet the requirements set up by the creators of transparency indicators.

The problems of trust are greater in less wealthy, less developed countries. While most of the largest legacy newsrooms can afford filling out detailed questionnaires, small newsrooms with a handful of journalists, especially in newer member states of the EU, may not have the same capacity to show compliance. Moreover, the criteria might be designed in a way that fits a healthy Western news environment, but is not necessarily feasible in media landscapes where the news media are less developed or journalists face immense economic or political pressures.

Furthermore, niche outlets, especially outlets by and for underrepresented social groups, such as linguistic, ethnic, sexual, etc. minorities might find it hard to fulfil all formal requirements. Steensen and Westlund (2020) argue that news in the 21st century has been separated from



This project has received funding from the European Union under Contract number: LC-01464044

journalism (or, in any event, from journalistic platforms). 'Today, news is something that you find in formats and on platforms of your own choosing. News is more often than not deprived of edited contexts and fixed genres and formats, and reaches you in mash-ups containing journalistic news, public relations news, advertisements, news from politicians, celebrities, sports idols, and artists, personal news from your friends and family, professional news from your colleagues and professional associations, and perhaps also fake news from bots' (Steensen and Westlund, 2020:8). The narrow focus on trustworthiness indicators on news media also means that other sources of information, such as most blogs, social media pages or individual users on social media are not included in the assessment.

The Commission's Guidance reflects on some of these issues, by highlighting both the voluntary nature of these indicators and the need for transparency about the indicators' methodology so that users can be aware of their limits and possible bias. The Guidance says:

> Signatories could facilitate access to such indicators providing users with the choice to use them on their services. In this case, the strengthened Code should ensure that signatories provide transparency regarding such third-party indicators, including about their methodology.
>
> The implementation of such trustworthiness indicators should be fully in line with the principles of media freedom and pluralism. To this end, it should be left for the users to decide if they want to use such tools.

Since the employment of indicators is 'left for the users to decide', it is expected that platforms will rely on not just one, but a diverse set of indicators that they can offer. This would be in line with what Leclercq et al. (2020) describe as 'a competitive environment for indicators, leaving users free to change, as they do with privacy settings'. Still, it has to be discussed whether and in what ways platforms should nudge users to use some form of trustworthiness indicators, or even to be confronted with default settings, when consuming content online.

## 5.2 Size to be considered

One of the problems related to trustworthiness is that on the internet only big, established enterprises and strong brands can be easily judged by their reputation. Legacy media like *Der Spiegel* or *The Guardian*, well-known online natives such as *Mediapart* or *De Correspondent* are automatically put in the category of trustworthy media. The same applies to well-known advertising agencies or verified accounts of public personalities (unless they abuse the authority provided by their position). So far, processes and initiatives that offer indicators to assess trustworthiness do not go much further than that, especially when it comes to the assessment of content sources that are not promulgating news.

This does not mean that these processes are not valuable. Yet, in some situations, even established media deliberately spread disinformation or disinformation (in the EU, the most



This project has received funding from the European Union under Contract number: LC-01464044

visible examples are the public service media and a part of the private media landscape in Poland or Hungary). Moreover, the trustworthiness of a media outlet can and does change: outlets can improve their track record once they have better funding or more expertise, or their standards can fall, due to a change in leadership. Thus, even in the case of content producers deemed trustworthy, there is a need to efficiently and regularly double-check context and content indicators, and, if needed, adjust the assessment of trustworthiness accordingly.

Even if there are understandable concerns by industry representatives about trustworthiness self-assessments and audits, we recommend asking content sources with a large enough audience (what constitutes large should be determined among experts) to provide sufficient information about their compliance with indicators. As some of the indicator creators themselves have recommended, while non-compliance should not be punished with downgrading, compliance could be rewarded with upgrading (prioritising) one's content. In parallel, fact-checkers should monitor content[12] (or react to cases when users report content) provided by these outlets. Those who are caught repeatedly publishing misinformation or disinformation would be downgraded in rankings.

The detailed assessment of trustworthiness is not feasible in the case of small or new players. Therefore, they should have the chance to use social media to reach audiences without constraints, as long as there is no sign of malicious use of the content-sharing platforms. Social media platforms themselves have already introduced some transparency requirements for users or accounts that come into play once they aim to monetise their content or boost their messages.[13] These requirements can also provide a basis to identify which content providers should be subject to increased scrutiny (defining thresholds based on number of followers, reach of content, as well as reports about earlier spreading of mis- or disinformation).

## 5.3 Platforms' Compliance

In addition to indicators of trustworthiness for content sources, facilitating access to trustworthy sources requires compliance by platforms as well. The Code of Practice on Disinformation requires platforms to regularly report about their efforts to tackle

12 Here, the Global Disinformation Index provides a good example for manual assessments of news media websites. It looks at selected news articles' "credibility, sensationalism, hate speech and impartiality." Their sample is based on an "anonymised review of 10 of the top-shared articles on a domain that have been randomly selected. The review is done by a researcher and the source of the articles is not disclosed to them."

13 See for example: "YouTube Channel Monetization Policies," Google.com, https://support.google.com/youtube/answer/1311392?hl=en#zippy=%2Cfollow-adsense-program-policies.

"Facebook Community Standards," Facebook.com, https://www.facebook.com/business/help/185404538833362?id=2520940424820218&recommended_by=321041698514182.



This project has received funding from the European Union under Contract number: LC-01464044

disinformation and empower users. This self-assessment has so far failed to live up to the expectations, as platforms were arbitrarily interpreting their requirements, while the numbers and data they provided were often incomplete[14].

To overcome this problem, the next iteration of the Code of Practice will require well-defined key performance indicators to assess platforms' actions (service-level Key Performance Indicator – KPIs and overall improvements in the online information environment (system-level or structural indicators). These KPIs need to be inclusive (considering current and potential future signatories of the Code); feasible (capable of being implemented on a regular basis under different forms of regulatory regime); mixed methods based (combining quantitative and qualitative indicators); and data informed (relying on an increased transparency of platforms and functional data access). In order to ensure platforms' compliance and keep track of developments related to trustworthiness, KPIs should also focus on both platforms' actions to prioritise trustworthy content in their algorithms and the overall improvements in the media system, with regard to content sources.

Example of such KPIs could be qualitative Service-level indicators that look at the existence of a working definition of trustworthiness utilised by platforms, the criteria used to determine whether a source/content is trustworthy and should be given prominence, the description of the measures taken to improve findability of trustworthy content, and the description of changes made to algorithms in order to improve findability of trustworthy content.

Quantitative Service-level KPIs could look at the number of content sources flagged, removed or suspended due to being identified as untrustworthy by platforms or by fact-checkers (either independent or contracted by the platform).

Structural Indicators could consider the change in number of available content sources flagged as not trustworthy, as well as traffic and engagement with them. If the assessment uses sample groups composed of internet audiences, that would also allow the tracking of a change in engagement with untrustworthy content producers and a share of untrustworthy content sources in users' online media diet.

## 5.4 Policy Developments to be Considered

In parallel with the discussion of trustworthiness of online content sources, it must be acknowledged that the EU audiovisual policy is facing the challenges of defining standards for the online environment and is proposing, as of its most recent revision in 2018, not only 'prominence' of European works as an obligation for all on-demand AVMS (Article 13(1),

14 Find the 2019 self-assessment reports here: https://digital-strategy.ec.europa.eu/en/news/annual-self-assessment-reports-signatories-code-practice-disinformation-2019



This project has received funding from the European Union under Contract number: LC-01464044

Recital 35 AVMSD), but also that 'Member States may take measures to ensure the appropriate prominence of audiovisual media services of general interest' (Article 7(a), Recital 25 AVMSD). Member States are still in the process of adopting national prominence framework; approaches vary significantly from country to country. Some built on long standing traditions regarding public service media (PSM),[15] others consider the use of 'quality labels'[16]. A number of EU states still lack a policy framework and the developments will be interesting to evaluate, particularly in a convergent perspective.

The Council of Europe, moreover, will release a Guidance Note on the Prioritisation of Public Interest Content, that will provide principles on prominence online, to establish to what extent relevant internet intermediaries can, or should prioritise certain forms of content over others, and under what conditions of transparency, accountability and liability.

# 6 Conclusion

We have seen that in the current online information environment it has become increasingly complicated for users to define what information to trust; the amount of available content online exceeds the time and attention that users have to assess their veracity. This study sought to analyse a topic that, within many facets, is increasingly becoming relevant as an element of the present and future media policy and of policies to tackle disinformation online; specific in the EU approach, is the consideration of how individual choices are driven by a technologically-curated information environment and how this can limit human autonomy and freedom of choice. Moreover, we looked at how new policies should take into account measures to enhance exposure diversity of trustworthy quality content in the abundance of content online.

Considering the scope of EDMO and its purpose of contributing to the debate on new policies to fight disinformation, the analysis of this paper concentrated on the measures foreseen by the Code of Practice on Disinformation as regards trustworthiness and the ways to implement them. The Code of Practice on Disinformation foresees an important role for the promotion and prioritisation of trustworthy content by large online platforms. Trustworthiness is explicitly

15 See e.g. Germany with its Medienstaatsvertrag (Article 84 MStV), the first and most advanced example of regulation in this area. For a detailed analysis on this, see Mazzoli and Tambini (2020), *Prioritisation uncovered. The discoverability of public interest content online* https://rm.coe.int/publication-content-prioritisation-report/1680a07a57

16 See e.g. the experiences in Bulgaria, Luxembourg and the recommendations made by the Flemish Media Regulator. A more detailed overview of the current implementation status of the AVMSD, and its prominence provisions can be found in the ERGA's "Overview document in relation to Article 7a of the Audiovisual Media Services Directive" (https://erga-online.eu/wp-content/uploads/2021/01/ERGA_SG3_2020_Report_Art.7a_final.pdf).



This project has received funding from the European Union under Contract number: LC-01464044

mentioned in two pillars of the Code. Pillar A (scrutiny of ad placements) highlights the importance of indicators of trustworthiness when identifying the sites where advertisement can be safely placed; Pillar D (empowering consumers) mentions indicators of trustworthiness as the basis of content prioritisation and media literacy measures.

The European Commission is looking for indicators of trustworthiness that can provide the basis of platforms for improving findability of trustworthy content sources, and for 'diluting' visibility (downranking) of their non-trustworthy counterparts. These indicators of trustworthiness should be based on objective criteria and endorsed by news media associations, in line with journalistic principles and processes. So far, there are four major projects that are often mentioned in the context of defining trustworthiness in the EU: the Trust Project, the Credibility Coalition, the Journalism Trust Initiative and the NewsGuard browser extension. In our overview of the indicators identified and listed by these projects, we found that it is a common property of these projects that they look at trustworthiness as a requirement that is attached to the content creator, rather than the content itself. Moreover, they treat trustworthiness mainly as a requirement that is attached to news outlets.

While connecting trustworthiness to the level of the content creators is indeed the best way to provide the basis for ex ante measures, the current focus on news media only allows for a narrow application. We found that using these indicators as the single source of determining trustworthiness of content sources – and therefore of which sources to downrank or prioritise – may create a media environment in which established players gain further competitive advantage, while new players face unprecedented barriers to entry. This could lead to serious problems for media pluralism and could distort the media market in a way that new players will find their access to the advertising market or other revenue sources further limited. The overreliance on these indicators can silence diverging or non-mainstream voices, as we have seen in the past when alternative/non-mainstream newsrooms suddenly lost a sizable percentage of their readers, due to tweaks in platform algorithms. Therefore, we recommend an approach in which content sources with a large enough following or readership would be asked to provide sufficient information about their compliance with indicators. In this system, non-compliance should not be punished with downgrading, but compliance could be rewarded with upgrading (prioritising) one's content. In parallel, fact-checkers should monitor content (or react to cases when users report content) provided by these outlets. Those who are caught repeatedly publishing disinformation (or misinformation) would be downgraded in rankings. Small and newly established outlets would, in the meantime, get a chance to use social media to reach audiences without constraints, as long as there is no sign of malicious use of platforms.



This project has received funding from the European Union under Contract number: LC-01464044

# 7 References


Altay, Sacha, Anne-Sophie Hacquin, and Hugo Mercier (2020): Why Do so Few People Share Fake News? It Hurts Their Reputation. New Media & Society.

European Commission (2018a): Tackling online disinformation. A European Approach

European Commission (2018b): Code of practice on Disinformation

European Commission (2020a): Democracy Action Plan

European Commission (2020b): Digital Services Act

European Commission (2020c): Commission Staff Working Document 'Assessment of the Code of Practice on Disinformation - Achievements and areas for further improvement'

Fukyama, Fancis (1995): Trust. New York: Simon and Schuster.

Khalil, Elias L. (2003): Trust. Cheltenham: Elgar.

Hardin, Russell (2002). Trust and Trustworthiness. New York: Russell Sage Foundation.

Leclercq, Christophe, Marc Sundermann and Paolo Cesarini, 'Time to act against fake news,' Euractiv, 24, November, 2020. https://www.euractiv.com/section/digital/opinion/time-to-act-against-fake-news.

Luhmann, Niklas (1980): Trust and Power. New York: John Wiley.

Mazzoli, Eleonora and Tambini, Damian (2020): Prioritisation uncovered. The discoverability of public interest content online.Council of Europe study. DGI(2020)19. https://rm.coe.int/publication-content-prioritisation-report/1680a07a57

Pan, Lee-Yun & Chiou, Jyh-Shen (2011). How Much Can You Trust Online Information? Cues for Perceived Trustworthiness of Consumer-generated Online Information, Journal of Interactive Marketing, vol. 25(2): 67-74.

Stavri, P Zoë et al. (2003): 'Perception of quality and trustworthiness of Internet resources by personal health information seekers.' AMIA Annual Symposium proceedings. AMIA Symposium vol. 2003: 629-633.

Steensen, Steen, & Westlund, Oscoar (2020). What is Digital Journalism Studies? (1st ed.). Routledge. https://doi-org.eui.idm.oclc.org/10.4324/9780429259555




This project has received funding from the European Union under Contract number: LC-01464044

Toff, Benjamin et al. (2021): Listening to what trust in news means to users: qualitative evidence from four countries. Oxford: Reuters institute. https://reutersinstitute.politics.ox.ac.uk/listening-what-trust-news-means-users-qualitative-evidence-four-countries

Valdani, Vicari and Associates (2020): Study for the assessment of the implementation of the Code of Practice on Disinformation



This project has received funding from the European Union under Contract number: LC-01464044