

THÉMA



Summary

Combating the dissemination of false information on online platforms:

an evaluation of the application and effectiveness of the measures implemented by operators in 2019

July 2020



Summary

Online platform operators subject to the duty of cooperation established under **Title III of the Law of 22 December 2018** on combating the manipulation of information submitted their first declarations to the CSA (*Conseil supérieur de l'audiovisuel* - Superior Audio-visual Council) regarding the means they have introduced during 2019 to **combat the dissemination of false information**.

Under Article 11 of this law, operators are required to introduce a reporting mechanism as well as additional measures which may relate to algorithm transparency, the promotion of content from press companies and news agencies and from audiovisual communication services (hereinafter EAP-SCA)¹, combating accounts responsible for the large-scale propagation of false information, information for users concerning the sponsored content of information relating to debates of general interest, and media and information literacy (EMI). In its **recommendation of 15 May 2019**, the CSA published guidelines on the means of implementing these measures.

This document provides an evaluation of the application and effectiveness of the measures introduced by operators in accordance with the provisions of the 4th paragraph of Article 17-2 of the Law of 30 September 1986 on the freedom of communication, as a result of the above-mentioned Law of 22 December 2018.

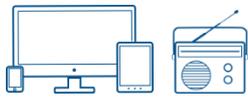
The timetable initially established by the Council for this first year was modified due to the COVID-19 health crisis. The deadline for the submission of declarations was extended to **23 April 2020**², by which time the majority of operators had actually sent their documents to the CSA, having drawn in particular on the questionnaire published by the latter in February 2020. In addition, a number of operators sought to provide the Council with additional information relating to specific measures introduced within the context of the health crisis. This evaluation is based on all of this information, and on additional exchanges between the Council's services and the operators, and on input from a committee of experts on online disinformation.

Eleven operators sent in declarations concerning, in total, **sixteen** services. These consist of online rating and referencing services for content, goods and online services, provided by third parties, sand multi-party contact services used for the sale of goods, the provision of services, and the exchange or sharing of content, goods or services. The CSA's analysis took into account the diversity of the platforms, both in terms of their activities and their business models, and their numbers of visitors and turnover.

The CSA wishes to commend the **high level of cooperation on the part of the operators** and the high quality of the dialogue that was established with them. **In this connection, the vast majority had appointed a contact person**, in accordance with the provisions of Article 13 of the Law of 22 December 2018.

¹ « Entreprises et agences de presse et services de communication audiovisuelle ».

² Instead of 31 March as established in Article 7 of the Recommendation of 15 May 2019.



The overall quantity and quality of the information declared is noteworthy. However, the level of detail in the responses varies. By and large, **little information was provided regarding the human and financial resources deployed to combat the manipulation of information, and the intelligibility³ of algorithms.** Although **business secrecy** could legitimately be invoked in certain cases, the CSA reiterated that it had given operators the opportunity to inform it about confidential information that they did not wish to be published.

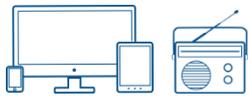
This lack of information hampered the CSA's capacity to assess the measures introduced. The Council, trusting the operators' **sense of responsibility**, therefore requests that they provide more complete declarations in the future, particularly with regard to its questionnaire.

As a result of their analysis, the CSA found that, in general, operators have responded to the challenge of combating the dissemination of false information by following a number of its recommendations; nevertheless, these measures require further development.

³ Intelligibility is understood here to mean the "explicability" of algorithms, a concept which refers to the publication of information relating to the purpose of the processing for which the algorithms have been programmed, in order to provide these details to the users to whom this processing applies. Some of the communicable information under this principle includes the rules defining this processing, the main features of its implementation, the data processed and its sources, the processing parameters and their weighting, the operations carried out by the processing and, where applicable, the degree and type of contribution towards decision-making, for example by offering one content rather than another. In particular, this principle has been mentioned in the context of administrative relations/users, in Decree No. 2017-330 of 14 March 2017 on the rights of persons making individual decisions based on algorithmic processing, pursuant to Article L. 311-3-1 of the Code of relations between the public and the administration, as amended by Article 4 of the Law of 7 October 2016 for a Digital Republic.



Summary.....	3
1. A mechanism for reporting false information likely to threaten public order or undermine the fairness of a ballot	6
2. Transparency of algorithms.....	7
3. Promotion of content from press companies and news agencies and from audiovisual communication services (EAP-SCA)	9
4. Combating accounts disseminating false information on a massive scale	10
5. Information of users on the nature, origin and modalities for dissemination of content, and the identity of individuals providing remuneration in return for the promotion of information content	11
6. Media and information literacy (EMI)	12



1. A mechanism for reporting false information likely to threaten public order or undermine the fairness of a ballot

The declarations included particularly well-developed information regarding the means employed in mechanisms for reporting false information.

All the operators, with one exception⁴, had introduced a reporting mechanism, with varying ergonomics. The majority of these tools are in the immediate vicinity of the content, but sometimes a form must be filled in, which may render the procedure more burdensome. In addition, **although certain platforms have added a specific category for reporting content as “false information”, others have not** and, where appropriate, include such reports in the category entitled “other”. Two arguments are put forward to justify this decision: on the one hand, certain operators do not consider that they are entitled to define content as “false information”; on the other hand, the existence of a “false information” category would lead to a flood of unsubstantiated reports.

One operator outsources the processing of these reports to *fact-checking* agencies, whilst the others use in-house teams of moderators. Investigations may also be partly automated. The CSA notes, however, that **the information provided regarding the details of investigation procedures and the human resources employed remains insufficient and does not allow it to fully evaluate the scope of the means introduced to meet the legislation’s objectives.**

Operators refer to several types of measures that may be taken against reported content, **ranging from reducing its visibility to removing it, or even taking action to target accounts.** User information and the availability of an appeal vary according to the operators involved.

Lastly, **the CSA is concerned about the shortage of statistics** provided by operators as regards reports and their processing, although it notes the practical difficulties associated with gathering these, as highlighted by the operators. This data is in fact essential for the quantitative analysis of the effectiveness of their measures.

The CSA makes the following recommendations:

1. Operators, particularly content aggregation platforms, are invited to continue to improve their reporting tools as regards **accessibility, clarity and ergonomics**. In particular, the CSA requests them to simplify their tools so that the submission of reports can be completed in a maximum of three clicks.
2. Platforms are encouraged to include a **“False information” category** or clearly-stated equivalent in their reporting form, and ensure that their reporting system allows users to provide **context** (explanations and screenshots).

⁴ Wikipedia, on account of its collaborative mode of operation.



3. By acquiring efficient **counting tools**, they could provide valuable, reliable and exhaustive statistical data on the reporting of false information and its processing.
4. Platforms that carry out internal reviews of content reported as false information are invited to adopt standards and reference sources, including considering **partnerships with third-party fact-checkers**. In this connection, they are encouraged to create **publicly-available documentary repositories** of such non-confidential reference documents. This would, on the one hand, allow operators to pool their resources and, on the other hand, allow users and civil society to examine them.
5. Platforms are encouraged to carry out internal **evaluations** of the processing of reports, and, amongst their users, evaluations of their reporting procedures.
6. **User information** could be improved in a number of areas:
 - platforms' policies on publications that propagate false information,
 - their procedures for processing reports,
 - the information required for monitoring to be carried out (standards and sources),
 - partnerships with *fact-checkers*,
 - the means of appeal against decisions taken by platforms (conditions, deadlines, and internal procedures for processing appeals), in general and, following the processing of a report, for the users concerned.
7. Operators are encouraged to develop and formalise **emergency procedures** for reporting false information under exceptional circumstances.

2. Transparency of algorithms

With two exceptions, operators use algorithms to **sort, reference and select** content, even if this three-part process is not expressly included in their declarations. They also refer to using algorithms to control accounts responsible for the large-scale propagation of false information, to detect and process false information and to increase or reduce the visibility of certain content.

Certain services allow users to personalise content that is offered to them and obtain information on the algorithms used, but this transparency is limited.

The operators' responses differ very little from the information already accessible on their websites and do not allow the CSA to carry out an exhaustive analysis. Despite the necessary confidentiality of certain information regarding operators' business models, which are largely based on the performance of their algorithms, the CSA is concerned about **the lack of clarity regarding the intelligibility of these algorithms** and **the incompleteness of the information submitted**. At a time when automated systems are playing an increasingly important role in the processing of false information, an understanding of the systems being introduced involves greater transparency vis-à-vis the regulator, without compromising business secrecy.



The CSA makes the following recommendations:

1. In future declarations, it would like operators to provide more information on **intelligibility and accountability**⁵, and on the algorithms they use to organise, select and sort content. It also invites them to provide it with the **principles of these approaches** as well as with **lists of the criteria and data used** by these algorithms. The CSA is aware that some platforms' business models are largely based on the performance of their algorithms. Nevertheless, it believes that it should be possible for them to provide information that would allow it to understand the operating principles and objectives of certain operators, without revealing their business models.
2. **The CSA recommends greater transparency as regards the functioning of algorithms vis-à-vis users.** At this stage, this information is provided within dense headings such as those concerned with help centres or privacy policies; this does not always facilitate accessibility and understanding. In general, it believes that **operators should be proactive in providing such information** as is necessary to the understanding and proper use of their services.
3. **The CSA also recommends greater transparency as regards the extent to which the users of these platforms may adjust these algorithms.** Contextual settings⁶, already offered by a number of operators, enable the level of transparency to be improved. Attention should also be paid to ergonomics in order to ensure that these types of settings are easily accessible and require a minimum of actions on the part of the user.

The Council plans to **develop these two points in the questionnaire** it will be sending platforms, with a view to drafting an evaluation of the application of the measures implemented in 2020.

4. Lastly, as was pointed out by the members of the committee of experts on online disinformation, a dynamic analysis of operators' practices appears to reveal an **increased reliance on algorithms, particularly to assist in the moderation of content**. This trend may have been amplified by the health constraints associated with COVID-19, which have led to an increase in the use of automated systems. In their next annual declaration, the CSA invites operators to **provide any relevant information on the respective roles and extent of human and algorithmic curation, in both the detection and processing of false information**.

⁵ Accountability as regards algorithms is defined as *"the explicit attribution of the responsibilities implied by the operation of algorithms"* according to the CNIL (French Data Protection Authority, *Commission nationale de l'informatique et des libertés*); this definition is cited on page 51 of its report *How can humans remain in control? The ethical issues raised by algorithms and artificial intelligence (Comment permettre à l'homme de garder la main ? Les enjeux éthiques des algorithmes et de l'intelligence artificielle)*, published in September 2017. For any given algorithmic processing, it is a question of knowing who is responsible for its definition and implementation and for any resulting decisions, as well as its stages and levels of intervention with regard to the machine. This is all the more important since recently-developed algorithms are based on shared learning (*"machine learning"*), the details of the outcomes of which, for example in terms of signal classification, are not even understood by the designers of such models.

⁶ Contextual settings or tools: features which are directly accessible to users from content offered when using the service referred to, for example by clicking on them. This type of feature differs from what are known as general settings or tools, which are accessible within help pages, settings sections or privacy policies.



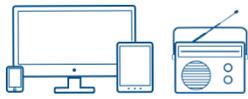
3. Promotion of content from press companies and news agencies and from audiovisual communication services (EAP-SCA)

Combating false information also involves the **promotion of reliable content**. Most operators identify such content using visual signs, and some allow users to obtain further information on its origin. Some operators **use algorithms** to **boost the visibility of content that is considered to have come from reliable sources and/or reduce the visibility of content where this is less likely**.

On the other hand, although **the importance of fact-checkers** in combating false information is generally acknowledged, the way operators use them and promote their work tends to vary. Some use *fact-checkers* directly or via a downloadable module, whilst others carry out verifications within their own teams. Still others cite the nature of their services or the implementation of editorial policies to justify their decision not to use them.

The CSA makes the following recommendations:

1. It welcomes and encourages practices which not only **provide users** with the names and logos of the organisations responsible for publishing content, but also provides additional information and context regarding featured content from EAP-SCAs.
2. Where platforms have not identified EAP-SCAs, it would be useful for the CSA and the users to know whether the platform has verified the reliability of the latter's identities and, if applicable, by what process.
3. The CSA also recommends that platforms that identify content sources as "reliable" or "authoritative" should show transparency to their users regarding the criteria they used to reach at this categorisation.
4. The CSA encourages **labelling procedures for EAP-SCAs** and considers that it is helpful for the users of platforms to be aware of them.
5. In general, the CSA strongly encourages **the development of partnerships between platforms and fact-checkers**.
6. Amongst the existing initiatives, the system introduced by Facebook involves a particularly interesting approach. A dedicated space allows strictly-selected and trusted third-party fact-checkers to access content circulating on the platform's services, confirm its veracity and intervene where appropriate to explain to what extent such content is likely to contain false information. This ensures the dissemination of reliable and independent information without making the platform solely responsible for evaluating the veracity of a piece of information.
In general, the CSA recommends that operators who use these types of systems include an emergency procedure which allows them to respond appropriately in exceptional circumstances (e.g.: an urgent, critical event).



7. The ClaimReview feature used by Bing, Facebook, Google and YouTube, which allows publishers to attach the label “*fact-checking*” to the *fact-checking* content they publish, is also an interesting tool. It is based on a lower level of access control and greater openness (no selection process), which allows for wider use but results in greater uncertainty as to the identity of the source of tagged content.

8. Lastly, the NewsGuard module, in partnership with Bing, which uses colour-coding to indicate the reliability of an information source, also helps to provide users with accurate information. It should be available by default rather than as an option.

4. Combating accounts disseminating false information on a massive scale

The notion of “accounts disseminating false information on a massive scale” covers **various situations, depending on the operators**. They include fake accounts, accounts whose behaviour contravenes a platform’s internal rules on the large-scale propagation of false information, and those responsible for the mass-production of content.

These accounts are detected by human and/or automated means. Several types of measures may be taken against them, from **the deletion of content to the temporary or even permanent blocking of accounts**. Certain operators mention increased vigilance vis-à-vis **government accounts**, a measure sometimes documented in detail in their declarations.

Users may access information in help centres and privacy policies but are not always informed about the consequences of creating such accounts.

Although operators emphasise the difficulty of identifying these accounts, the CSA is concerned about the disparities in the figures declared.

The CSA makes the following recommendations:

1. It requests operators, in future, to provide it with **details of their procedures for detecting and processing accounts** responsible for the large-scale propagation of false information.
2. It would like to obtain information on **advertising revenue**, even where minimal, generated by accounts that are responsible, or potentially responsible, for the large-scale propagation of false information and that have not been detected and deactivated since their creation.
3. In the interests of user awareness and transparency, the CSA encourages operators to **develop user information** on the measures for controlling such accounts.



5. Information of users on the nature, origin and modalities for dissemination of content, and the identity of individuals providing remuneration in return for the promotion of information content

A number of operators did not answer the section of the questionnaire relating to this objective, on the grounds that they do not allow content relating to debates of general interest. However, these questions covered all sponsored content⁷.

All the actors involved identify sponsored content and some of them help users to understand why they are exposed to this content. Although certain have a specific mechanism for reporting sponsored content, the majority rely on the general reporting system.

Operators have adopted **advertising policies**, which may involve validating advertising content or banning certain content. Little information was provided regarding **brand safety** and measures that enable sponsored content to gain visibility.

Approaches to information content that relates to debates of general interest differ between operators. Some of them prohibit its promotion, either permanently or during elections; others carry out specific checks on the advertisers concerned, reduce their advertisement targeting options or use a specific label.

A number of operators have introduced **advertisement libraries** that provide access to all or part of sponsored content.

The CSA makes the following recommendations:

1. The CSA encourages platforms to continue to introduce **clear and directly-accessible labelling of** promoted content, which is easily visible to users and provides specific details concerning advertisers' identities.
2. In the interests of clarity and transparency for users, it is essential that advertising policies, including those concerned with misleading or inappropriate advertising, should be accessible **in French** and include a **section on the manipulation of information**. These documents could be cross-referenced from various locations (for example, from the content).
3. User transparency tools and tools to empower consumers should be encouraged, such as those that **help them understand why they were targeted and set their advertising preferences**. However, the settings configurations must be accompanied by clear and easily accessible information in order to ensure that their implications are fully understood by the user.

⁷ The notion of sponsored content is understood here as advertising in the broad sense, encompassing "standard" advertising and the promotion of content relating to debates of general interest.



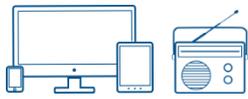
4. The CSA requests that operators make an extra effort in the future towards transparency regarding **measures that enable sponsored content to gain visibility**. An understanding of these mechanisms is essential to combating false information.
5. It calls on platforms to adopt **measures relating to brand safety and the demonetisation of pages** propagating false information, in order to prevent the creators of such content from deriving financial benefit from it. It also draws advertisers' attention to the negative effects of associating their brands and images with content that propagates false information.
6. In order to avoid situations in which advertisers could be blocked from some platforms but not others, it recommends **dialogue between operators**, to allow them to work towards a greater harmonisation of the approaches adopted regarding the promotion of information content relating to debates of general interest.
7. To this end, it invites operators to ensure that their *advertisement libraries*, especially those containing promoted information content relating to debates of general interest, **are easily accessible** and user-friendly. They must comply with a greater degree of transparency, for example by providing detailed information on the identities of advertisers, expenditures, the audiences targeted and affected and the number of publications, and by ensuring that thorough research can be carried out.

6. Media and information literacy (EMI)

Some operators undertake **EMI initiatives**, as corollaries to the above-mentioned measures. Providing their audiences with keys to understanding the rich and varied content to which they are exposed is a major challenge for these services.

Certain initiatives are developed **in partnership with public institutions** and non-profit organisations. They allow audiences both to master the use of platforms and develop their critical skills.

Online platform operators also provide financial support for several types of initiatives, and some of them mention links to the world of research, mainly through the provision of financial assistance.



The CSA makes the following recommendations:

1. It invites operators to develop **multi-dimensional approaches** to media and information literacy, aimed at all audiences (8-12 years old, 12-15 years old, 16-18 years old, 19-25 years old, 26-40 years old, 40-65 years old and the over 65s, parents, teachers, etc.)
2. As regards EMI, it encourages them to create **long-term partnerships** with recognised institutions and major stakeholders in the world of education (public institutions, associations, etc.).
3. It encourages them to **assess the impact** of their actions on the behaviour of users.
4. The CSA strongly encourages them to build **protocols** with the world of research **for the operational sharing** and transparent exploitation of **data**.

Lastly, a number of online platform operators mention **additional measures** similar to the measures used for combating *deepfakes*, for protecting electoral processes, for advanced search functions and for placing restrictions on “*live*” features.