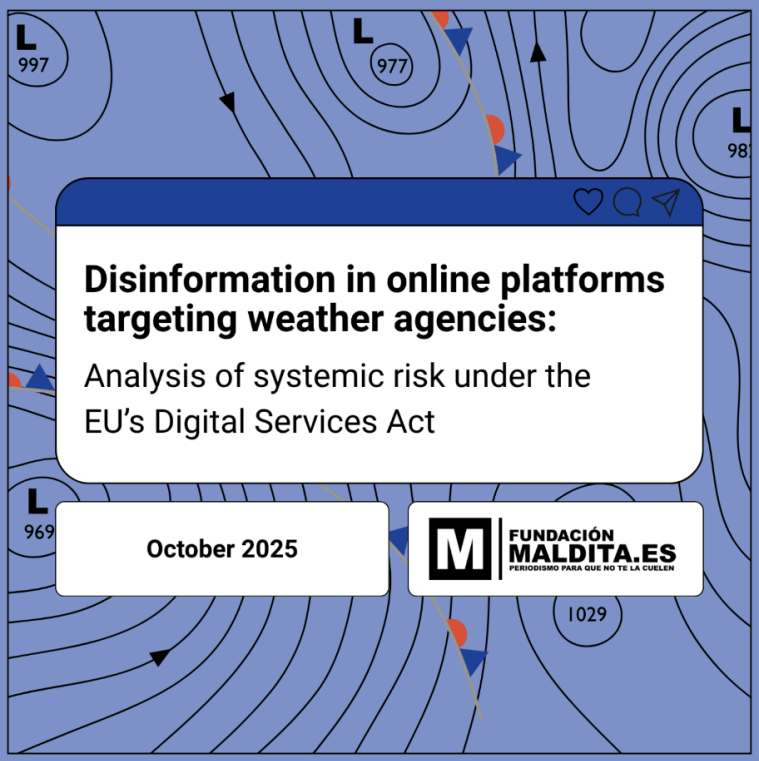


Disinformation in online platforms targeting weather agencies:

Analysis of systemic risk under the EU's Digital Services Act

October 2025

FUNDACIÓN MALDITA.ES
PERIODISMO PARA QUE NO TE LA CUELEN

Table of contents

Executive summary

On October 23, 2024, the Spanish Meteorological Agency (AEMET) issued its first alert regarding what would later become the *Valencia DANA* six days later - a natural disaster that caused more than 200 deaths and underscored the importance of information integrity and the risks of disinformation before, during, and after emergency situations.

Fundación Maldita.es had already identified that a significant portion of climate-related disinformation spread in Spain in recent years through digital platforms has aimed to discredit the State Meteorological Agency (AEMET) and its professionals, generating distrust in its warnings and public service information (particularly essential in situations like the DANA).

This report constitutes a first step toward considering the risks posed by that type of disinformation within the regulatory framework of systemic risks established by the European Union's Digital Services Regulation (DSA). To this end, Fundación Maldita.es has assessed its severity using the analytical framework recommended by the European Commission for identifying such systemic risks, based on the Rabat Action Plan.

We have also analyzed misleading content hosted on digital spaces designated by the European Commission as "Very Large Online Platforms", and therefore subject to the specific obligations of identifying and reducing systemic risks established in Articles 34 and 35 of the regulation. In this case, these include Instagram, Facebook, TikTok, Youtube and X.

This analysis shows that the risk mitigation measures implemented by these platforms in recent years have been insufficient. Only 8% of posts containing misinformation already debunked by Maldita.es received any form of verification label or additional context from the platforms. This level of effectiveness varies across services, ranging from 16.67% on Facebook to 0% on TikTok. The precision of each platform's internal policies in addressing disinformation targeting meteorological agencies is also uneven, with X not even addressing disinformation in general.

1. Risk definition under the Digital Services Act framework

Within the framework of the European Union's Digital Services Act, very large online platforms are required to assess the potential risks associated with the design, operation, and use of their services. Among the established risk categories are threats to civic discourse and public safety, as well as negative effects and consequences related to public health and the physical and mental well-being of individuals. These are closely linked to disinformation targeting AEMET, particularly because the Regulation emphasizes that platforms must monitor how their services are used to "disseminate or amplify false or misleading content, including disinformation."

a. Definition of the risk

In its report "Digital Services Act: Application of the risk management framework to Russian disinformation campaigns", the European Commission suggests a qualitative and quantitative assessment of risk severity by applying a modified version of the Rabat Action Plan as a proportionality test. We apply these points to the spread of disinformation about AEMET.

i. Context

AEMET is a public agency that plays an essential role in the current context of the climate emergency. Its functions include issuing weather warnings and forecasts to protect the population, maintaining the historical record of climate data, conducting research in atmospheric sciences, and developing climate change scenarios.

In the current scenario, weather phenomena such as heavy rainfall or extreme cold are becoming increasingly frequent and intense as the global average temperature rises due to climate change. Consequently, AEMET's warnings are more recurrent, and citizens are more familiar with the Agency.

Especially after the catastrophic impact of the DANA storm in towns across Valencia at the end of October 2024, AEMET came under public scrutiny, further fueled by successive disinformation messages. This led the Agency to present itself as an aggrieved party in the investigation into the leak of an audio recording of a conversation between a meteorologist and a technician from the Generalitat Valenciana's 112 emergency service, which had been shared in a truncated form.

Disinformation campaigns find in institutions like AEMET a clear target where narratives critical of public management converge with climate manipulation theories. Similar attacks have been detected against public agencies in other countries, such as the U.S. Federal Emergency Management Agency (FEMA), which was affected by disinformation campaigns, amplified and spread by foreign actors, after hurricanes Helene and Milton.

ii. Position, status, or intention of the speaker

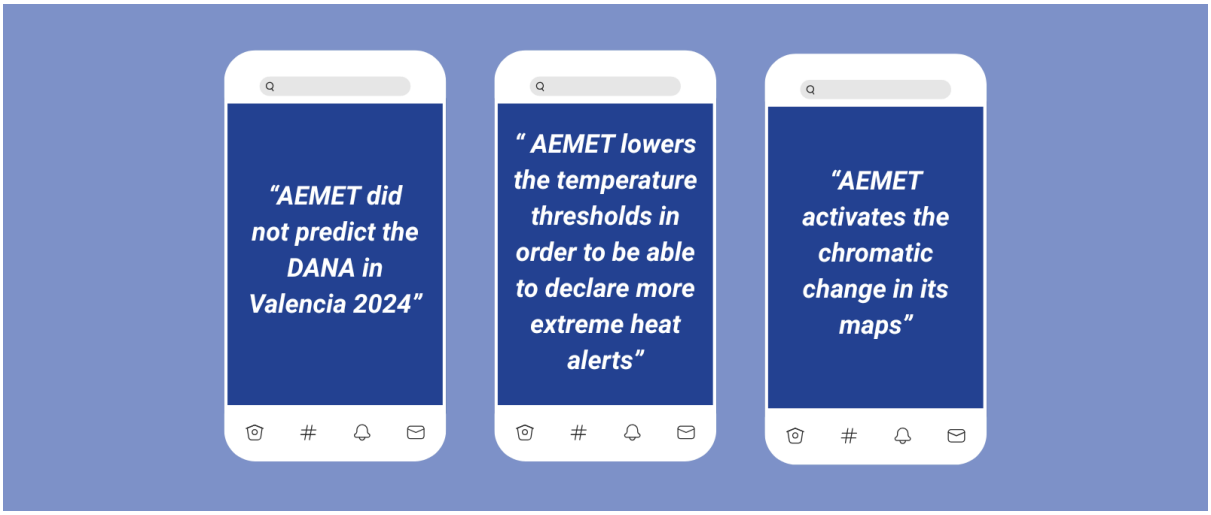
Not only after the DANA storm was AEMET questioned from political spheres, but different authorities have also made public statements calling for "sharpening" its weather forecasts or for more "rigor" after challenging a warning. This questioning of the specialized Agency, fueled by disinformation narratives, further amplifies its reach as it is sometimes also covered by national media outlets or widely read websites.

iii. Content and form of the statement

In recent years, the Maldita.es Foundation has identified and debunked various disinformation narratives targeting the State Meteorological Agency. These individual falsehoods, shared on digital platforms, collectively feed into broader discourse that share a common denominator: undermining the credibility and public perception of the Agency.

The main narratives detected include:

- Discrediting AEMET's scientific reliability: disinformation portraying the agency as incompetent, inaccurate, or incapable of accurately forecasting significant weather phenomena.
- Manipulation or alteration of climate data: disinformation suggesting that AEMET deliberately modifies data or adjusts thresholds with the aim of alarming the public or supporting specific political agendas.
- Alleged concealment or institutional negligence: disinformation insinuating that the agency or its staff hide critical information or act without due diligence in the face of significant weather events.
- Complicity in climate manipulation processes: content linking AEMET to conspiracy theories about artificial climate modification, attributing to it an active or complicit role in such practices.



iv. Reach, size, and characteristics of the audience

In Spain, social media is the second most used source of information (46%), but it becomes the main one among young people (43%), according to the Digital News Report 2025. In this environment, disinformation can spread to millions of people within hours—well before official agencies have the capacity to respond.

Among the examples of debunked disinformation posts collected for this report, there are posts on X with more than 34,000 likes accusing AEMET of lowering temperature thresholds in order to declare more extreme heat alerts, or others with over 3 million impressions claiming that the Agency had failed in its forecast of a warm and dry winter. On YouTube, a video about an alleged prediction by the "French AEMET" regarding the DANA storm surpassed 235,000 views. More than 2,500 users reshared a Facebook post claiming that AEMET had described two springs with the same average temperature as both "cold" and "warmer than normal." On Instagram, a reel with 21,800 likes referred to radar inoperability in the days before the DANA in Valencia, while a TikTok video sharing the truncated audio between AEMET and the 112 emergency service reached 1,300 likes and 400 reshares.

AEMET is a recurring target of disinformation campaigns, particularly around weather events. These narratives usually emerge during moments of heightened public attention and vulnerability, which amplifies their potential emotional impact and drives up interaction rates, thereby increasing both reach and the number of people exposed.

Given the nature of its work, AEMET's audience ranges from local communities to the entire country, including both rural areas and urban environments, each with different risks, attitudes, and information needs.

v. Probability or imminence of harm

*"After what happened with the DANA, I'm sorry but I don't give a d*mn about what AEMET says" (sic).*

Disinformation targeting AEMET and spread through major digital platforms is fueled and amplified by continuous hate attacks against the agency and its staff. Rubén del Campo, spokesperson for AEMET, has explained that "one in four messages directed at AEMET are hateful."

A study published in 2025 analyzed nearly half a million messages to assess hate speech directed at AEMET on X. The researchers revealed a significant percentage of hateful content driven by conspiracy theories and climate change denial, also tied to skepticism toward science. The same study warns that this disinformation climate "contributes to the erosion of public trust in AEMET and its professionals, which leads to a questioning of science in general."

Public perception

AEMET's functions are key: they form part of the emergency response chain by issuing warnings and forecasts, and its monitoring and research work is essential to underpin climate-related policies. If disinformation narratives about this agency take root, especially among people in decision-making positions, the consequences could be highly detrimental.

Chilling effect

Attacks in the form of insults, questioning of professional capacity, or challenges to the integrity of staff, driven by disinformation narratives, have a direct impact on those working in research. According to a study by the Spanish Foundation for Science and Technology published in 2024 on "Researchers' experiences in their relationship with the media and social networks", half of those interviewed admitted to having suffered some form of attack, and the majority (59.49%) stated that it had affected their work.

This chilling effect, meaning self-censorship or modification of public communication for fear of negative reactions, in a context where it is urgent to fill digital spaces with rigorous information, can have serious consequences and further facilitate the spread of disinformation.

b. Systemic risks under the European Digital Services Act

Risk category #1. Public health

Disinformation that discredits or undermines the work of AEMET may lead the population to fail to adopt adequate preventive measures, increasing the risk of injuries, illnesses, or even deaths (e.g., heatstroke, hypothermia). It may also result in dangerous behaviors (e.g., ignoring warnings advising people not to go outside) by downplaying the need for precautions during alerts and extreme weather events.

Risk category #2. Public safety, civic discourse, and democratic processes

Extreme weather events often trigger civil protection and evacuation protocols. If disinformation discredits the meteorological authority, it may generate widespread disobedience to these orders, hindering the work of emergency services. This not only endangers individuals but also overwhelms public resources (rescue, healthcare, transport, or recovery). In addition, the spread of contradictory or manipulative messages during weather emergencies can cause collective panic, unrest, or inefficient evacuations. This is the case, for example, of a false rumor that immigration status would be checked at

evacuation shelters after Hurricanes Harvey and Irma, which could have affected the decision-making of many citizens.

On the other hand, disinformation campaigns portraying agencies like AEMET as manipulative, corrupt, or incompetent undermine trust in public institutions. In other words, the attempt to discredit impacts not only AEMET but the entire network of public organizations and agencies with different functions. Furthermore, it also affects the perception of scientific evidence, eroding the quality of democratic debate by replacing facts with conspiratorial narratives.

2. Evaluation of risk mitigation by platform

Proper mitigation of the risks posed by harmful disinformation directed at a public body like AEMET requires a dual strategy from digital platforms. First, they need to adapt their community standards to the specific nature of this type of content and explain clearly how these are applied. Second, it is essential to implement visible measures for users that enable them to make informed decisions.

Among these actions, key measures include the official verification of AEMET's institutional accounts as trusted sources, as well as the use of labels providing specific, explanatory information based on scientific evidence to effectively reduce the virality of harmful content. During crises, it is also necessary for platforms to integrate panels with official information that users can easily access. Overall, for these measures to achieve their goals, they must be applied consistently.

The analysis carried out shows that compliance with these measures is uneven across different platforms.

	Facebook	Instagram	X/Twitter	YouTube	TikTok
Disinformation posts with additional context information	16,67%	7,69%	5,45%	4,35%	0%

a. Meta: Facebook and Instagram

i. Terms and conditions

Meta's community standards, applicable to both Facebook and Instagram, have a specific policy on misinformation, with moderation actions that include deletion or labeling. However, most of the content detected against AEMET does not clearly fit the categories of this policy, with the exception of possible "manipulated content" that could be misleading, which would be labeled with an informative label.

In the standards regarding Meta's fact-checkers program, different cases of relevant content are described in more detail, such as:

- "Claims (…) that cannot be considered an interpretation of something that actually happened or was actually said,"
- "Conspiracy theories explaining events as the secret work of individuals or groups, citing true or unverifiable information but presenting an unlikely conclusion,"
- "Inaccuracies or miscalculations regarding numbers, dates, or times, but which could be considered an interpretation of something that actually happened or was said,"
- "A fragment of authentic multimedia content (…) that distorts the meaning of the original content to implicitly make a false claim," or
- "Use of data or statistics to imply a false conclusion."

ii. Visible actions on Facebook

- 16.67% of Facebook posts containing disinformation attacking AEMET had a label with information provided by independent fact-checking organizations—the highest percentage among the platforms analyzed.
- These labels correspond to different disinformation messages that have been debunked. However, more than half of the posts containing these same messages were not labeled, reflecting some inconsistency.
- On AEMET's official Facebook page, the "Government Organization" badge is visible in the page information section.
- Facebook did not deploy any special information panels during the floods in Valencia in 2024 or the wildfires in Spain in August 2025.

iii. Visible actions on Instagram

- 7.69% of Instagram posts containing disinformation attacking AEMET had a label with information from independent fact-checking organizations—the second highest percentage after Facebook.
- AEMET's official account is not verified by Instagram.
- Instagram did not deploy any special information panels during the floods in Valencia in 2024 or the wildfires in Spain in August 2025.

b. TikTok

i. Terms and conditions

Within its community guidelines under "Integrity and Authenticity," TikTok has a specific section on "Disinformation." It states that it does not allow "disinformation that may cause significant harm to individuals or society, regardless of intent." There are two levels of moderation with content types relevant to this case:

- Not allowed – disinformation that: "poses a risk to public safety," "may cause panic about a crisis or emergency situation," "related to climate change," or "specific conspiracy theories targeting a specific individual."
- Excluded from FYF: "Misrepresented credible sources," "general unfounded conspiracy theories claiming that certain events or situations are the result of covert or powerful groups," "unverified claims related to an emergency or ongoing event."

ii. Visible actions

- Although removal is the main moderation measure on this platform, a search for content debunked by Maldita.es revealed that disinformation messages continue to circulate.
- None of the TikTok videos containing disinformation about AEMET have any visible actions such as warnings or labels. On average, these posts received 445 likes and 90 shares. Some of them have up to 55,000 views.
- TikTok provides search functions for terms related to climate or climate disinformation that redirect users to United Nations information on climate change and instructions for reporting content. No similar feature exists for searches using the term "AEMET."
- During the flood emergency after the DANA storm in Valencia, TikTok deployed a generic information panel about natural disasters that provided road condition updates. Similarly, during the wildfires in Spain in August 2025, a panel appeared when searching for "fires," linking to the civil protection page and its guidance for support during tragic events.
- The platform offers verification badges, although AEMET does not have an official account on TikTok.

c. YouTube

i. Terms and conditions

YouTube has specific policies on "disinformation," "electoral disinformation," and "medical disinformation" in its community guidelines. Only the first policy includes the prohibition of manipulated or miscontextualized content that poses a high risk of serious harm, which applies to some cases of disinformation against AEMET.

ii. Visible actions

- Only one of the videos on YouTube containing disinformation attacking AEMET displayed an information panel on climate change linking to United Nations content. Another avoided the panel by altering the word "climático" in the title ("clim4tico").
- These generic panels also appear in certain searches related to the climate emergency, although users can choose to disable them so they do not appear.
- AEMET's official channel does not have any badge identifying it as a public agency.
- YouTube did not deploy any special information panels during the floods in Valencia in 2024 or the wildfires in Spain in August 2025.

d. X/Twitter

i. Terms and conditions

X does not have specific definitions or policies for disinformation within its rules and policies. Under the "Authenticity" section, it only prohibits "Altered Content" or "Miscontextualized Content," mainly focused on multimedia material, similar to YouTube.

The online platform also offers the Community Notes feature, allowing users to voluntarily suggest additional context for posts that may need it. Whether these notes are displayed alongside posts containing misinformation depends on an algorithm that determines whether people with a tendency to disagree agree with the usefulness of a given note.

ii. Visible actions

- 5.45% of posts containing disinformation related to AEMET had a Community Note, despite other types of visible actions existing.
- The Community Notes labeling system relies on contributions from volunteer users without requiring validation of scientific knowledge. The voting system on the usefulness of notes causes many notes with relevant contextual evidence to never be shown to users.

- X does not provide any search functions or panel systems with official information from external and/or official sources.
- AEMET's official account has a gray verification badge, as it is classified as an "account of a governmental or multilateral organization."

3. Conclusions and Recommendations

The application of the framework recommended by the European Commission to implement the risk assessment principles of the Digital Services Act highlights the role of platforms in spreading harmful disinformation that damages public perception of meteorological agencies such as AEMET. Furthermore, a preliminary analysis of the mitigation measures currently carried out by the platforms studied shows that they are insufficient—although with differences among them—given the potential impact of the spread of this content on public safety and health, civic discourse, and democratic processes.

These are some of the actions platforms should take to curb this risk:

- Prioritize the use of visual cues (e.g., fact-checking labels) to provide relevant and specific evidence alongside misleading messages, empowering users to make informed decisions. This approach should take precedence over content removal, which should be reserved for posts considered illegal.
- Provide specific information from authoritative sources during extreme weather events according to the user's location.
- Adapt platform policies and terms of service to mitigate the risk of spreading harmful disinformation and hate speech about these agencies and their employees, and apply them consistently and coherently.
- Promote collaboration among scientific institutions, government agencies, and platforms to develop concrete strategies to limit the spread of hate speech and disinformation that hinders the work of these entities.

4. Methodology

The definition of risk follows the framework proposed by the European Commission in its report "*Application of the risk management framework to Russian disinformation campaigns*," as well as the categories of systemic risk identified in Article 34 of the Digital Services Act affected by this type of disinformation.

To assess mitigation possibilities, an analysis was first conducted of the community standards and current policies of Meta (Facebook and Instagram), TikTok, YouTube, and X/Twitter relevant to the practices observed in this type of disinformation. This was complemented by an analysis of the visible actions taken by platforms that help limit the effects of this risk. A database was created of posts shared on these platforms containing disinformation about AEMET that had been debunked by Maldita.es, to verify whether any type of moderation measure with contextual information visible to users existed. In addition, the use of verification badges or additional panels during crises related to natural disasters was also examined.