REGULATING DISINFORMATION: RISKS AND OPPORTUNITIES

FUNDACIÓN MALDITA.ES

PERIODISMO PARA QUE NO TE LA CUELEN

ABOUT MALDITA.ES

Maldita.es is a not-for-profit media outlet established as **Fundación Maldita.es contra la desinformación: periodismo, educación, investigación y datos en nuevos formatos.** In English Maldita.es Foundation Against Disinformation: Journalism, Education, Research and Data in new formats. Maldita is the Spanish word for "damned", a reference to our first project **Maldita Hemeroteca** (Damned Archive), that confronted public leaders with their past statements. Our goal is to give people "tools not to be fooled", so we develop innovative formats particularly suited for social media and for the creation of a community that is committed to our mission and ready to help us fulfill it.

Through **Maldita Hemeroteca** (archive), **Maldito Bulo** (hoax), **Maldita Ciencia** (science) and **Maldito Dato** (Data) we fight disinformation in the public discourse using fact-checking techniques and data journalism. We also count on various projects that address disinformation on more specific topics: **Migration, Feminism, Nutrition, Climate and Fraud**. In **Maldita Tecnología** (Technology) we work for a better knowledge of technology and in **Maldita Educa** we do Media Literacy in order to build a society that everyday grows more capable to identify dis and

> **Our main goal is to create a community that is more resistant and educated against disinformation**

misinformation before sharing it. We also have teams specialized in engineering, projects, academic research and public policy.

Since 2017 we are part of the International Fact-Checking Network as signatories of their Code of Principles and in 2022 we led the process to develop an association of over 40 European fact-checking organizations, the European Fact-Checking Standards Network (EFCSN), in which we currently hold the presidency. We were the only Spanish media organization that was part of the European Commission High Level Expert Group on Fake News and Disinformation and we now take part in the Task Force for the Code of Practice on Disinformation. In 2022, we were as well in the EDMO Task Force on Disinformation on the War in Ukraine and the Group of Experts of the Civil Society on Disinformation Campaigns in the field of National Security in Spain.

Contact:
Carlos Hernández-Echevarría
Head of Public Policy and Institutional Development
chernandez@maldita.es

M FUNDACIÓN **MALDITA.ES**
PERIODISMO PARA QUE NO TE LA CUELEN

ÍNDICE

SUM-MARY

Legislation can help fight dis and misinformation, but it will not solve the whole problem. Neither does a large government body that regulates and oversees social media and digital platforms. The main venues for government action ought to be education, technology, community and journalism.

What legislation can do is to incentivize big platforms to apply mechanisms of content moderation that are more effective and credible. Invite them not to keep making those decisions by themselves, opaquely and on a case-by-case basis, but in close collaboration with those who, along them, better know how disinformation works: academics, media and independent fact-checkers.

Any legal reform ought to be done with the highest possible level of consensus, not only political but also social, so it may not be seen as an attempt to censure critical

> **It cannot be "the reform of X party or X government". Disinformation is a systemic problem and it calls for systemic solutions**

speech. It cannot be "the reform of X party or X government". Disinformation is a systemic problem and it calls for systemic solutions.

Those solutions must take into account a fundamental issue: the main responsibility of public institutions when it comes to disinformation, regardless of how they prosecute it, is education: in order to build a society that is more resilient to lies, they have to provide citizens with the necessary tools to distinguish disinformation, to doubt about what they receive and analyze it carefully before sharing it.

A KEY MOMENT IN THE FIGHT AGAINST DISINFORMATION

Disinformation is not a new problem, but fighting it has never been so urgent. The same technologies that have made the dissemination of information faster, cheaper and more accessible have equally facilitated that of disinformation. This has become even more evident during pandemic: a study by the Reuters Institute[1] has found that, in the first three months of the COVID-19, the number of debunks in English by fact-checking organizations rose by 900%. Maldita.es is a nonprofit fact-checking organization and we have also witnessed the same phenomenon: during the worst part of the pandemic, we received five and six times more requests than in the previous months. In total we have debunked more than a thousand contents related to the coronavirus.

According to the definition adopted by the final Report of the European Commission High Level Expert Group on Fake News and Disinformation[2], of which Maldita.es was a member, disinformation is "false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit[3]. That includes the promotion of medical remedies whose efficacy is not accredited or the communication of fake arguments against measures that prevent contagions, but other times disinformation hurts social harmony and coexistence.

> **"**
>
> **In total we have debunked more than a thousand contents related to the coronavirus**

In the same report [4], experts expressed that most of the solutions for disinformation "will be of a non-regulatory character", but the current peak in harmful disinformation has made many lawmakers wonder if laws can do something about it. At the European level it has already been approved the new Digital Services Act[5] which has among its objectives the fight against this phenomena and that it will be implemented soon. In Spain we have also recently witnessed public debates on how to better address this crisis and which role should public institutions play. In fact, Maldita.es took part in an initiative by the National Security Department that resulted in the publication of a book filled with proposals from the civil society to counter disinformation campaigns. Nevertheless, it is evident that the scale of the problem overflows borders and legislation would be more effective if implemented EU-wide.

---

[1]Brennen, J. S., Simon, F., Howard, P. N., & Nielsen, R. K. (2020). Types, sources, and claims of COVID-19 misinformation. *Reuters Institute*, 7, 3-1.

[2]de Cock Buning, M. et. al. (2018). A multi-dimensional approach to disinformation: Report of the independent High level Group on fake news and online disinformation.

[3]Wardle, C., & Derakhshan, H. (2017). Information disorder: Toward an interdisciplinary framework for research and policy making. *Council of Europe report*, 27, 1-107.

[4] see 2.

[5]European Commission (2022) The Digital Services Act package

THE DIFFICULTIES OF LEGISLATING AGAINST DISINFORMATION

Legislating against disinformation is not without risk. The most dangerous of them all is for the new regulation not to be effective or even be counterproductive if it does not have wide popular support. Since it tangentially affects a fundamental right such as freedom of speech, regulation can only be credible and effective if it is not perceived as "the government's reform" or "X party's reform". Disinformation is a systemic problem and it calls for systemic solutions.

That depends for starters on a drafting process that is transparent and inclusive, aiming for political and social consensus. It has to rely on the participation of the widest array of political actors, but also representatives of civil society and in particular of the most relevant stakeholders in the disinformation field: universities and research institutions, media and journalists, freedom of speech experts and independent fact-checkers. That collaboration will not only produce a better text, it will also make it more difficult for the regulation to be seen as a tool to outlaw speech that is critical of the government or a given political party. That would mean failure before the regulation is even enacted.

Beyond public perception of its final form, regulating disinformation online has particular difficulties. It is a sphere where change is constant and the law ought to be flexible enough to not to become obsolete in a very short period of time, and also being able to oversee very different actors. In the text of the new Digital Services Act[6], for example, there are established differentiated obligations for platforms on the nature of their business and their size

Moreover, legislators will find that any regulation will have to address several different goals that are equally crucial in the fight against dis and misinformation:

1. How digital platforms are to manage content that is patently illegal such as child pornography, frauds or hate speech.
2. What kind of treatment platforms have to apply to content that is not necessarily illegal, but it is nevertheless disinformation harmful to society: who gets to decide it is so and how that decision is made, based on what.
3. Which mechanisms platforms can adopt to avoid manipulations of their services that could have, in the words of the European Commission, "real and foreseeable negative impact on public health, public security, civil discourse, political participation and equality"
4. What to do with individuals or companies that create or disseminate disinformation
5. Which measures should public authorities themselves adopt to foster a society that is more protected and educated against disinformation.

"

Regulating disinformation online has particular difficulties

[6]European Commission (2022) The Digital Services Act package

THE RESPONSIBILITY OF BIG PLATFORMS REGARDING DISINFORMATION

A big part of the public debate about how to better fight against disinformation has focused so far on which obligations the most popular platforms have in it. Should they decide by themselves that a content is illegal or wait until judges rule? Is a complaint enough evidence to decide so? What timeframe should they have for acting on an illegal content? What should they do about content that is fake and harmful but not patently illicit? Who decides that a content belongs in that category? And once it is decided, should the platform delete or just include a label explaining why it is problematic?

When it comes to the management of illegal content, it is important that platforms act quickly on the ones that pose an immediate danger to some people. It is evident that a video that contains child pornography or incitation to murder shall have no place online or offline. However, there are many

"

What should they do about content that is fake and harmful but not patently illicit?

other cases which are more difficult to determinate, such as hate speech that the Council of europe has defined [7] as "the advocacy, promotion or incitement (…) of the denigration, hatred or vilification of a person or group of persons, as well as any harassment, insult, negative stereotyping, stigmatization (…) and the justification of all the preceding types of expression, on the ground of "race", colour, descent, national or ethnic origin, age, disability, language, religion or belief, sex, gender, gender identity, sexual orientation and other personal characteristics or status".

9

There have been several recent initiatives across Europe that focused on timeframes: how quickly should platforms be in removing illegal content once it is flagged. Some have already become law in Germany[8] or France[9] while other parliaments, such as the Spanish [10], have approved motions to urge their governments to follow suit. In Maldita.es we believe that it is important that illegal content is acted upon but, besides defining timeframes, regulatory efforts must focus on creating procedures to identify those contents and moderate them in a manner which is effective, objective and mindful of freedom of speech. Credible processes to manage illicit content, but also other disinformation that through a broad agreement has been considered harmful to society.

Currently, almost every big digital platform makes its moderation decisions by itself and in an opaque manner, and that makes them subject to accusations of acting selectively and capriciously against disinformation. Some experts argue that the best solution is to create a specific governmental body to regulate platforms. Others such as the US conservative think tank Cato Institute[11] predict that any public intervention is deemed to fail.

In Maldita.es we do not believe that a government body imposing its decisions is the best way to decide how to moderate content online, neither is for the platforms to remain deciding by themselves, without clear and fact-based procedures. We believe society needs to have a voice in deciding how to regulate that sphere. Our proposal is to create an independent and collaborative tool that includes the most relevant stakeholders: platforms, academia, media, journalistas and fact-checkers. A body who can establish rules and oversee compliance while legislation is used for more detailed areas such as microtargeting, political manipulation and algorithmic transparency.

This model of collaborative regulation is consistent with the principle of "co-regulation" established by the UE's Digital Services Act[12] and it is also consistent with the path that platforms like Facebook or Google are starting to walk. Both already include the work of independent fact-checkers as the basis of their moderation decisions.

Other companies such as Tik Tok have also started to work with independent fact-

FUNDACIÓN
MALDITA.ES
PERIODISMO PARA QUE NO TE LA CUELEN

"

We do not believe that a government body imposing its decisions is the best way to decide how to moderate content online

checkers, but the results of that collaboration are not yet verifiable and quantified. Twitter is making efforts to reduce the dissemination of disinformations, but its processes are still opaque and closed. Some other platforms such as Telegram or YouTube are not doing much to stop disinformation, at least in Spanish.

What any disinformation legislation must require from big platforms in the first place is their commitment to a

collaborative regulatory framework and an enforceable duty of transparency. For starters, that collaboration should translate into information and data: disinformation experts need to better understand how disinformation flows inside platforms, assess the impact of their automatic algorithmic moderation systems and the effectiveness of academic research and verification.

"

Disinformation experts need to better understand how disinformation flows inside platforms

---

[7] ECRI: General Policy Recommendation N°15 on Combating Hate Speech (2015). *Council of Europe*

[8] Netzwerkdurchsetzungsgesetz vom 1. September 2017 (BGBl. I S. 3352), das zuletzt durch Artikel 9 Absatz 2 des Gesetzes vom 30. November 2020 (BGBl. I S. 2600) geändert worden ist

[9] LOI n° 2020-766 du 24 juin 2020 visant à lutter contre les contenus haineux sur internet

[10] Proposición no de Ley sobre la prevención de la propagación de discursos de odio en el espacio digital (161/001536).

[11] Samples, J. (2019). Why the government should not regulate content moderation of social media. *Cato Institute Policy Analysis*, (865).

[12] European Commission (2022) The Digital Services Act package.

THE RESPONSIBILITY OF INDIVIDUALS, COMPANIES AND INSTITUTIONS REGARDING DISINFORMATION

Although there are people who manufacture disinformation for financial gain or ideological reasons, it is very difficult to pinpoint the exact origin of a fake content and even more to determine the intention of the person who created it. In most cases, it is impossible to separate someone who is simply mistaken from other who has done it willingly and fully aware of the harm it was causing. That is why punitive responses to disinformation ought to be moderate in nature.

However, there are also systematic efforts to disseminate disinformation, intentional and carefully organized, that can be acted upon by the law. There are countries that execute disinformation campaigns as a tool of influence or aggression and democratic governments have the duty to be vigilant and deactivate them. In addition to those, there are companies that work full time producing disinformation and monetizing it.

Some platforms have already taken measures against this kind of webs or networks and, at the very least, they should be prevented

" it is very difficult to pinpoint the exact origin of a fake content and even more to determine the intention of the person who created it

from making a profit from those activities. Of course, those measures will have more credibility if adopted through a structured and transparent procedure, with specific parameters and ways to appeal the resulting decisions. It is the platform that is in a position to evaluate if its services have been abused, but the law can create common standards and incentivize collaboration with independent fact-checkers.

As for governmental institutions, they have a crucial role to play in the fight against disinformation and one that they could be currently exercising more effectively. Their first priority ought to be being

more transparent themselves: in Spain, the government successfully responds to only 60% of the information requests it receives, according to the statistic produced by its own transparency website [13]. Secretism, especially when it comes from official sources, create opportunities for disinformation since there is always someone ready to fill that vacuum.

Aside from that, the most important action governments can take against disinformation is to support a more educated society, more resistant to it. A citizenry that thinks twice before sharing (and believing) something they have received in their cell phones or seen on social media. Any serious public endeavor to fight disinformation must include, before anything else, a strong effort in media literacy to promote critical thinking and protect citizens from manipulation.

The current crisis in the media business model and other economic uncertainties affecting journalism have created an opening for disinformers. Society must demand more from its reporters, political representatives and big digital platforms that today serve as an extension of the public sphere. It ought

"

Any serious public endeavor to fight disinformation must include, before anything else, a strong effort in media literacy

to be much less credulous than in the past as well. And of course, independent fact-checkers such as Maldita.es have to find ways to be more relevant, to reach more people and to make our job even more useful.

---

[13] Portal de Transparencia de la Administración General del Estado (2022). Datos del Derecho de acceso a la información pública.

CONCLUSIONS: FREQUENTLY ASKED QUESTIONS ABOUT REGULATING DISINFORMATIONS

Is it possible to regulate away disinformation?

Probably not

As of today, we do not have enough knowledge to regulate platforms effectively to end disinformation, nor could we do it without jeopardizing the fundamental right to freedom of expression. This is, for example, one of the concerns expressed by the Council of Europe[14] regarding the draft amendments to the Turkish Penal Code that tried to introduce a provision on "false or misleading information" and that the special rapporteur on Freedom of Opinion and Expression has condemned in other countries.[15] And even if we did, the Internet is an ever-changing environment that calls for extremely flexible regulation in order not to be obsolete on day 1. What the law can and should do is to **create collaborative and credible processes of content moderation, to incentivize the most relevant stakeholders to work together and also to promote media literacy** and technological tools that make society more resilient against disinformation, leaving strict legislation to more precise issues such as microtargeted political manipulation or algorithmic transparency.

Wait, should platforms regulate themselves?

Not really

In 2018, an EU Code of Practice on Disinformation was drafted and signed by some of the most popular digital platforms. Although it was a good first step, the European Commission itself considered that the Code had fallen short in its results. Four years later, a new version of this Code[16] has been introduced, with more demanding obligations, including the collaboration with fact-checkers and civil society organizations, and with the European Commission in charge of accruing its implementation. In **Maldita.es**, that took part on the negotiations and is a signatory of this code and part of its permanent group, we believe that many

FUNDACIÓN
MALDITA.ES
PERIODISMO PARA QUE NO TE LA CUELEN

of the solutions can be in a middle ground between self-regulation and the creation of a governmental regulatory organism: **independent tools with presence of the platforms, universities and research institutions, media, independent fact-checkers and civil society**

If done wrong, is there a risk to change things for the worse?

An enormous risk

Almost any initiative to combat disinformation affects tangentially to freedom of expression, so from the very beginning it has to be the product of the widest possible political and social consensus. Disinformation is a state problem and it calls for state policies. **If a reform comes from just the government or only one party, it will be perceived as an attempt to censor online critics of that government or that party.** To be effective, legislation on disinformation must be credible and that starts with a drafting process that is transparent, inclusive and committed to reach a wide consensus.

Then, what should we be demanding from Internet platforms?

First: Transparency

What we first need from big internet companies is transparency, information, data. They need to acknowledge that their actions against disinformation should no longer be decided in-house, shrouded in secrecy, without any input from academia, fact-checkers, the civil society and even public servants. And to better help them in that fight **we need information on the scale and impact of disinformation strategies,** on the algorithm automated responses and also on the effectiveness of current fact-checking strategies. Companies such as Facebook and Google are taking steps in that direction. Others such as Tik Tok have started to work with independent fact-checkers as well, but the results of that collaboration are no yet public or quantifiable. Twitter makes efforts to reduce the circulation of disinformation inside the network but its moderation processes are still opaque and closed. Telegram or YouTube are doing little to fight disinformation in Spanish.

Why should they work with independent fact-checkers?

Because it works

The notion that people online "will believe what they choose to believe" are commonplace but untrue. Scientific research tells us that what we do is working, that debunks actually hit the brakes on the dissemination of disinformation and also help people who believed a hoax to replace it with quality, fact-based information. That is why platforms have to work with us.

Why don't they just use content from trusted sources?

Not a great idea

If we tell platforms to only allow content from "trusted sources", we are putting them in the position to decide who is "trustworthy" and who is not. It will also be very damaging to competition, making it harder for new voices to be heard if traditional media have that kind of advantage. Fact-checkers should not do it either: in Maldita.es we would rather stick to checking facts, not outlets, and let the public judge by themselves.

Should they just delete disinformation every time it pops up?

That is problematic

It is true that platforms could simply delete a content the moment a verified fact-checker flags it as disinformation, but we do not recommend it except in cases when not doing it represents an immediate danger to human beings.It is often more useful to attach a clear warning to that content with a link to the debunk, so people can actually understand why it was problematic in the first place. Alerting citizens about false information does not jeopardize freedom of speech, but deleting can be perceived as censorship. Our mission as fact-checkers is to provide quality information that will help citizens make informed decisions.

Shall we punish those who spread disinformation?

No, in most cases

Responses to misinformation have to be measured even if the statements targeted and we should always remember that, in most cases, someone who shares a hoax is most likely a victim of the deceit and not the author of it. Even if we could find the original first source of a piece of disinformation, which is extremely difficult, sometimes is just someone who was wrong about something online and must not receive the same treatment as those who do it on purpose, with an agenda, or putting others in danger willfully.

Are the platforms the only problem?

Not at all

In the fight against disinformation, there are many other areas with room to improve besides the behaviour of a few large companies: there are political actors who foment disinformation and many institutions whose lack of transparency make it flourish. We also need education to boost media literacy that allows people to be more resilient and less credulous when they come across disinformation. There is also the question of how the depletion of the media traditional business model affects this problem, for example when we find online ads for scams with the appearance of news placed in widely-considered trusted sources. And of course us independent fact-checkers have to find ways to be more relevant and increase the impact of our work.

---

[14] Venice Commission - Council of Europe (2022) Urgent Joint Opinion of the Venice Commission and the Directorate General of Human Rights and Rule of Law (DGI) of the Council of Europe on the draft amendments to the Penal Code regarding the provision on "false or misleading information".

[15] UN Human Rights Council, Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, Irene Khan (2021) Disinformation and Freedom of Opinion and Expression. A/HRC/47/25. https://documents-dds-ny.un.org/doc/UNDOC/GEN/G21/085/64/PDF/G2108564.pdf?OpenElement

[16] European Commission (2022) The 2022 Code of Practice on Disinformation

M

FUNDACIÓN
MALDITA.ES
PERIODISMO PARA QUE NO TE LA CUELEN