

Building Trust, Curbing Lies:
*A Digital Badge
to Political Misinformation*

AUTHORS

Claudia Caffo

Davide Bello

Ririko Hirase

Emma Bergman Karlsson

Téo Thevenin

The Future of Europe

16 December 2024

SUPERVISORS

Mariyana Angelova

Bence Hamrak

01 PROBLEM DEFINITION

Political misinformation in the realm of social media

02 CURRENT SOLUTIONS

Current measures in place to combat political misinformation

03 OUR PROPOSAL

The Digital Literacy Badge: our plan to combat misinformation

04 CONCLUSION

A step towards responsibility and trust on social media

Political misinformation in the realm of social media

- *Tiktok under accusations of interference in Romania's elections*



Călin Georgescu after its first round victory from the newspaper "franceinfo" (2024)

A recent controversy published by POLITICO (2024), surrounding TikTok's alleged role in **Romania's presidential election** serves as a stark example of political misinformation. The ultranationalist candidate Călin Georgescu's unexpected first-round victory has been linked to a surge in his popularity on the platform, where accusations of **manipulative content** and **externally sponsored accounts** have surfaced. Romanian authorities allege that "thousands of accounts" amplified Georgescu's message, potentially **swaying public perception** and **voter behavior** (Robertson, 2024).

- *Online echo chambers favor the spread of misinformation*

Politicians and politically influential people are often responsible for disseminating false information that has a profound impact on voter behavior and confidence in democratic institutions (Lewandowsky et al., 2023). **Misinformation** is defined as incorrect or misleading information, exaggerated using clickbait headlines or out-of-context details to make a story harder to ignore. Misinformation is distinguished from **disinformation** that is deliberately created and spread with the intent to mislead, deceive, or manipulate people. In this policy paper, the term "misinformation" will be used as a generic term, whether or not the intention is to deceive users. While algorithms can propagate misinformation, **online echo chambers** play a particular role in promoting and reinforcing presupposed narratives and ideologies (Samuel C. Rhodes, 2021). Online echo chambers refer to an ecosystem where users encounter and engage primarily with viewpoints that reinforce their own beliefs. Social media platforms in particular show increased echo effects around political topics, with users often engaging with ideologically similar voices (Barberá et al., 2015). This phenomenon is largely the result of **selective exposure**, where individuals gravitate towards information that supports their existing beliefs. As a result, political discussions on these platforms tend to take place within ideological boundaries, creating an environment conducive to **polarized discourse** (Rabb et al., 2023).

Political misinformation in the realm of social media

- *Affective polarization*

Social media may exacerbate emotional divides between political groups, a phenomenon often referred to as **affective polarization**. This is even accentuated when partisanship triggers positive feelings for the in-party and negative biases toward the out-party (Iyengar et al., 2012). Unlike **ideological polarization**, which is based on policy differences, affective polarization is driven by group identity. Simply identifying with a political party triggers emotional reactions that shape perceptions of the opposing group. This process occurs through "**hot cognition**", where political stimuli are stored in memory with emotional associations that influence future judgments, making objective reasoning difficult (Lodge & Taber, 2013).

- *Heightened political division*

While cross-ideological engagement does occur, **political content** circulates mainly within **ideologically homogeneous groups** (Barbera et al., 2015). Additionally, **divisive political rhetoric** can strengthen affective polarization by portraying opposing parties as **existential threats**, thereby galvanizing partisan loyalty and intensifying perceptions of threat (Iyengar & Westwood, 2014).

- *A threat to the Democracy*

Highly **polarized individuals** are usually more prone to accepting **misinformation** that supports their in-party while dismissing information coming from the out-party (Jenke, L., 2023). Misinformation poses serious **threats to democracies** and can influence the result of elections. French elections of 2017 and Indian elections of 2019 are often cited as examples where misinformation has influenced election process (Muhammed et al, 2022). As a result, it is not surprising to see that 83% of EU citizens view political misinformation as a major concern. It is therefore natural to ask whether the EU has taken the necessary measures to combat political misinformation (European Commission, 2018).

83 %

Percentage of EU citizens that view political misinformation as a major concern

What has the EU done?

The **European Union** has introduced several **policies** to fight **misinformation** across media platforms due to their significant influence on **public opinion** and ability to disseminate **false information** rapidly.

Regulations and Acts

The EU's regulative framework establishes rules for social media platforms, to curb illegal content and misinformation. In the last few years different regulations have been implemented like the **General Data Protection Regulation (GDPR)** in 2016 about social media companies' access of consumers data (Regulation EU 2016/679), the **Code of Practice on Disinformation** in 2018 and strengthened in 2022, to take measures against manipulative practices, improve fact-checking, and clearly label AI-generated content (European Commission, 2018), the **Digital Markets Act (DMA)** to target monopolistic "gatekeeper" platforms like Google and Meta, promoting content diversity and reducing echo chambers by curbing the dominance of single viewpoints (Regulation EU 2022/1925).

The milestone regulation regarding the issue of this paper is the **Digital Service Act (DSA)** of 2022, that imposes obligations on platforms to identify and remove harmful content, combat misinformation, and ensure user protection (Regulation EU 2022/2065). It emphasizes due diligence, transparency, and risk assessments by platforms, aiming to safeguard users' rights while ensuring content governance. It has been followed by the **European Media Freedom Act** in 2024 that protects EU journalists and media from political or economic interference by preventing online platforms from arbitrarily restricting or deleting independent media content (Regulation EU 2024/1083).

In particular, the DSA seemed to fix the failure of the previous acts with a layered enforcement system. However, challenges in tackling misinformation persist, including **harmonizing laws** across EU member states, some of which have introduced their own regulations that may undermine the DSA's unified approach. The law also faces scrutiny over **balancing platform responsibilities** with freedom of expression and avoiding excessive government control (Pírková, 2023). In addition, the success of these voluntary measures depends on **platform compliance** and does not fully address the need for enhanced media literacy. For example, X has not updated its policies to these guidelines in 2024 (Corlin, 2024).

What has the EU done?

Administrative and Coordination Efforts

To bolster its approach, the EU has implemented administrative initiatives like **cross-border alert systems** to facilitate information sharing on disinformation campaigns among member states. It has also allocated research and innovation funding through programs to develop technologies, including generative AI, that can detect and counter misinformation (European Commission official website).

Fact-Checking Initiatives

The EU has launched several fact-checking efforts, including:

- the **European Digital Media Observatory (EDMO)**, a hub for research, collaboration, and tools to tackle online disinformation;
- **EUvsDisinfo**, a project of the European External Action Service (EEAS) that tracks and debunks disinformation narratives, primarily from Russian state-controlled sources, and maintains a public database of disinformation cases;
- the **European Fact-Checking Network (ECFN)**, a pool of fact-checking organizations coordinated by the European Journalism Centre.

These initiatives aim to verify and debunk disinformation. However, challenges persist in holding social media platforms accountable and ensuring fact-checking efforts reach populations heavily influenced by false narratives. Their effectiveness often depends on widespread adoption and visibility.

Media Literacy

Media literacy is a cornerstone of the EU's strategy against misinformation. Programs like "**Media Literacy for All**" aim to empower citizens to critically evaluate information and navigate digital spaces effectively (Chapman, Bellardi, Peissl, 2020).

However, **gaps** remain, like: **platform compliance**, balancing **rights and regulations** and **media literacy for political contexts**. The initiatives underscore the **need for** robust oversight, cross-border consistency, and expanded **innovative solutions** to fully address misinformation coming directly from political influencers and its pervasive influence on social media and users' polarization.

The Literacy Digital Badge

To tackle the problem of misinformation, we are introducing the **Digital Literacy Badge program**. It is a recognition designed to **encourage responsible digital behavior** and serves as an incentive for social media users to consistently **share accurate and well-researched information**. The **Digital Literacy Badge** is aimed for political influencers, who **voluntarily** apply and commit to these guidelines:

- Participants must ensure the information they share is **backed by reliable sources**, such as government agencies, academic studies, or credible media.
- They are also expected to **practice transparency** by signaling warnings when content is disputed.
- Additionally, they should promote **balanced conversations** by avoiding one-sided and false information.

• *A practical example of the Digital Literacy Badge*

In the image below, our team of experts have represented what the **implementation** of the Digital Literacy Badge would look like on social media, in this case Instagram. Here, a French content creator named “*hugodecrypte*” covers the Ukrainian war and interviews Ukrainian President Volodymyr Zelensky. In order for political influencers and politicians to earn this badge, they shall only **share reliable and accurate information** to their followers.



Figure 1: example of an Instagram post of *hugodecrypte*, a French content creator

Please be aware that this is a fictionnary reproduction of how the digital literacy badge could be implemented on social media platforms

The Literacy Digital Badge

Why do social media users want to earn the Digital Literacy Badge?

The users with the Badge would **gain public recognition** as trustworthy and responsible social media users. The Badge would also **enhance visibility** for users, as content from accounts with this badge would be prioritized. In this way, the accounts of political influencers can contribute to fostering a **fairer electoral competition**, where **truthful content** is amplified, and false information is diminished.

Why would social media platforms cooperate?

Social media platforms have incentives to support the badge program, as it encourages **responsible content sharing without censorship**, allowing platforms to highlight credible voices and **reducing the risk of legal challenges** related to misinformation. Furthermore, we believe that users will feel more confident engaging with content from accounts with the badge and will be **more inclined to stay** on the platform if it is backed by reliable sources, helping to avoid what happened with X, where a large decline in users occurred due to reduced trust and the spread of misinformation (The Economic Times, 2024).

Feasibility

Independent Oversight and Certification: An EU-based independent body would oversee the accreditation process as well as review content for **misinformation**. If a badge-holder spreads misinformation, their **badge** will be **revoked**. Since participation in the Digital Literacy Badge program is **voluntary**, the badge system **avoids backlash** from regulations regarding freedom of speech, allowing political influencers to choose to engage responsibly.

Social Impact

The **Digital Literacy Badge program** is expected to **enhance social welfare** and **reduce political polarization**. By fostering a more balanced political environment and decreasing misinformation, it could ultimately contribute to more **stable democracies** as disinformation is seen to undermine democracy (European Parliament, 2021; Au et al., 2021).

To support these outcomes, the control process mirrors established models such as **Hazard Analysis Critical Control Point (HACCP)** and **Fair Trade** certifications, which rely on independent audits and robust accountability mechanisms. These measures not only ensure the reliability of the badge system but also reinforce public trust, making the program a sustainable tool for combating misinformation and strengthening democratic institutions.

*The Digital Literacy Badge will be **funded** through:*

- **Public funding**, which will come from the **Digital Europe Programme** and the **Citizens, Equality, Rights and Values Programme (CERV)**. It should also be secured through **partnerships with tech companies**, like Google, that already contribute to misinformation efforts, as part of their corporate social responsibility commitments. The targeted **social media platforms** should also be involved to contribute to fund this crucial policy.

Implementation

The **successful implementation** of the **Digital Literacy Badge** system will depend on effective **collaboration** between the **European Commission** and **social media platforms**. Additionally, it will require the establishment of an **Independent EU Oversight Body** and an **EU Marketing and Public Engagement Body**, which will coordinate their efforts with the aforementioned entities.

Funding for the Digital Literacy Badge will be sourced from both **public** and **private resources**.

*The **European Commission** will be central in implementing the Digital Literacy Badge by:*

- Defining the **Digital Literacy Badge** within the **Digital Services Act**, holding social media platforms accountable for actively preventing misinformation disseminated by political influencers.
- Implementing **rigorous standards** for **badge holders** in collaboration with the *Independent Oversight Body*, social media platforms and the existing *Digital Services Coordinator* of each Member State, ensuring that all content shared by badge holders is verifiable and strictly adheres to these official standards.

***Social Media Platforms** will be responsible for:*

- Collaborating with the **Independent Oversight Body** and the **EU Marketing and Public Engagement body** to develop an effective strategy leveraging their respective competences and ensure smooth implementation and continuous assessment for improvement.
- Promoting **badge-holders' content** more organically as part of their **strategy** to combat **misinformation** in the **political field**, by aligning the badge system with their existing content ranking algorithms. Certified content will receive higher rankings in user feeds. Initial pilot programs in select member states will serve as testing grounds, chosen to reflect diverse linguistic, cultural, and digital contexts, refining processes and gathering feedback through standardized metrics and real-time feedback loops. A full rollout will follow, supported by public-private partnerships and additional EU funding to ease integration costs for platforms.

Implementation

*Establishing an **Independent EU Oversight Body** will ensure consistent, EU-wide supervision of the Digital Literacy Badge system, addressing fragmentation across Member States. This institution will:*

- Include **independent law, philosophy, and IT experts**, ensuring diverse representation.
- Collaborate with **platforms** to develop an **automated monitoring system** that provides real-time tracking of badge-holder content, flags potential misinformation for manual review to ensure fairness, and defines how badge-holder content is prioritized while enforcing EU Commission standards.
- Revoke badges when necessary and issue an '**Orange Badge**' to signal the need for **corrective action** against misinformation.
- Implement an **appeals process** for individuals to challenge decisions.
- Publish regular **performance reports** to enhance public trust, outlining badge-holder compliance rates, addressing reported misuse, and ensuring transparent, evidence-based accreditation free from political influence.

*The **EU Marketing and Public Engagement Body (EMPEB)**, composed of members from the Directorate-General for Communication, will be responsible for:*

- Collaborating with **tech experts** to design a badge system that appeals to a **broad pool of political influencers**. This system will offer **benefits** such as increased **visibility, networking opportunities, and public recognition**.
- Launching a **marketing campaign** to raise **awareness** and target **diverse political ideologies**, including **both mainstream and anti-establishment figures** —especially as **initial adopters**.
- Addressing **misconceptions** about the badge by emphasizing its role in enhancing **accountability without limiting diverse voices**.
- Evaluating the badge system to ensure its **effectiveness** in reducing **misinformation** from political influencers across various demographics.

Fostering Accountability and Trust: A Scalable Solution to Political Misinformation in the EU

The proposed **Digital Literacy Badge** offers an innovative solution by incentivizing **responsible digital behavior** and promoting credible, transparent content shared by **political influencers**. By establishing clear criteria, such as using verified sources and fostering cross-ideological discourse, the badge system can empower political figures, to foster a healthier online ecosystem.

Limits of our proposal

Despite its potential, the Digital Literacy Badge system faces critical limitations that must be acknowledged to ensure realistic expectations. Perceived bias in accreditation could undermine public trust, and voluntary platform participation risks inconsistencies if major players opt out. Reliance on influencers introduces agent hazards, as their integrity cannot be guaranteed, while comprehensive monitoring risks inefficiency without a trust mechanism. Furthermore, involving multiple institutions enhances expertise but creates vulnerability to single-point failures. Public skepticism toward establishment endorsements, especially from followers of misinformation-prone influencers, poses another challenge. These issues highlight the need for continuous evaluation and adaptive strategies to maintain credibility and effectiveness.

Bigger picture and complementary solutions

The Digital Literacy Badge system is a cornerstone initiative in the broader fight against misinformation, but its success depends on integration with complementary measures and long-term strategic planning. Strengthening regulatory frameworks, including provisions under the Digital Services Act (DSA), will bolster accountability for platforms disseminating misinformation. As the system matures, global outreach efforts could position the EU as a global leader in misinformation governance by showcasing scalable solutions and fostering international partnerships to safeguard democratic values.

Reference list (1/2)

- Afp. End of a love affair: news media quit X over “disinformation.” The Economic Times, November 15, 2024.
<https://economictimes.indiatimes.com/news/international/business/end-of-a-love-affair-news-media-quit-x-over-disinformation/articleshow/115340267.cms?from=mdr>
- Au, C. H., Ho, K. K. W., & Chiu, D. K. W. (2021). The role of online misinformation and fake news in ideological polarization: Barriers, catalysts, and implications. *Information Systems Frontiers*.<https://link.springer.com/article/10.1007/s10796-021-10133-9>
- Barbera, Pablo, et al. (2020). The Role of Social Media in Political Polarization. *Journal of Political Communication*.
- Colomina, C., Sánchez Margalef, H., & Youngs, R. (2021). The impact of disinformation on democratic processes and human rights in the world. Policy Department for External Relations, Directorate-General for External Policies of the Union. European Parliament.
[https://www.europarl.europa.eu/thinktank/en/document.html?reference=EXPO_STU\(2021\)653635](https://www.europarl.europa.eu/thinktank/en/document.html?reference=EXPO_STU(2021)653635)
- Commission of the European Union. Citizens, Equality, Rights and Values Programme Overview. European Commission. Accessed December 4, 2024.
- Corlin, Peggy. (2024). Elon Musk’s X evades compliance with landmark EU antitrust regulation.
- Digital Strategy. Impact Assessment of the Digital Services Act. European Commission. Accessed December 4, 2024.
- Digital Strategy. The Digital Programme. European Commission. Accessed December 4, 2024.
- European Commission. (2018). Flash Eurobarometer 464: Fake news and disinformation online. Directorate-General for Communications Networks, Content & Technology.
- Full Fact. The £350 Million a Week Claim and Boris Johnson’s Misuse of Statistics. Full Fact, November 27, 2024. <https://fullfact.org/>
- Google Search Fact-Checking Resources. Google Blog. Accessed December 4, 2024. <https://blog.google/products/search/google-search-fact-checking-resources/>
- Hobolt, S. B., Lawall, K., & Tilley, J. (2024). The Polarizing Effect of Partisan Echo Chambers. *American Political Science Review*, 118(3): 1464-1479.
- Instagram post: <https://www.instagram.com/p/C6CKZsOLPXv/>
- Iyengar, Shanto, and Sean J. Westwood. (2014). Fear and Loathing Across Party Lines: New Evidence on Group Polarization. *American Journal of Political Science*, 58(4): 690-707.
- Iyengar, Shanto, et al. (2012). Affective Polarization in the American Public. *Political Science Research and Methods*, 1(1): 3-18.
- Jenke, L. (2023). Affective polarization and misinformation belief. *Political Behaviour*

Reference list (2/2)

- Lewandowsky, Stephan, Ullrich K.H. Ecker, John Cook, Sander van der Linden, Jon Roozenbeek, & Naomi Oreskes. (2023). Misinformation and the epistemic integrity of democracy. *Current Opinion in Psychology*, 54.
- Libby, Laura. (2024). Misinformation and Partisan Bias: How Political Sophistication Shapes Media Consumption. *Journal of Political Behavior*, 46(2): 245-268.
- Lodge, Milton, & Charles S. Taber. (2013). The Automaticity of Affect for Political Leaders, Groups, and Issues: An Experimental Test of the Hot Cognition Hypothesis. *Political Psychology*, 34(5): 703-726.
- Morris, Chris, (2024). Elon Musk's Attacks on Federal Employees Prompt Fears of Targeted Harassment. CNN, November 27, 2024. <https://edition.cnn.com/>
- Muhammed, T. S., & Mathew, S. K. (2022). The disaster of misinformation: a review of research in social media. *International Journal of Data Science and Analytics*.
- Pírková, Eliška, (2023). The EU Digital Services Act won't work without strong enforcement, <https://www.accessnow.org/eu-dsa-enforcement/>
- Rabb, N., Cowen, L., & de Ruiter, J. P. (2023). Investigating the effect of selective exposure, audience fragmentation, and echo chambers on polarization in dynamic media ecosystems.
- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC.
- Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act).
- Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).
- Rhodes, S. C. (2021). Filter Bubbles, Echo Chambers, and Fake News: How Social Media Conditions Individuals to Be Less Critical of Political Misinformation. *Political Communication*.
- Simonovits, Gábor, et al. (2022). Partisan Identity and Support for Norm-Eroding Policies: Evidence from a Large-Scale Experiment. *Journal of Politics*, 84(3): 1175-1191.
- Starcevic, Seb, and Pieter Haeck. Brussels to Mediate Between TikTok and Romania as Election Row Builds. *Politico*, November 28, 2024. <https://www.politico.eu/article/brussels-mediate-tiktok-romania-election-row-builds/>
- Stanford Internet Observatory. Election Integrity Partnership: Final Report. Stanford University, March 2021. <https://purl.stanford.edu/tr171zs0069>

Acknowledgement: Chat GPT has been used for language purposes (only).