Funded by the
European Union

SMIDGE

Social Media Narratives : Addressing Extremism in Middle Age

D7.1 Guidelines, Recommendations, and Policy Briefs

[15/12/2025]

Disclaimer

This project has received funding from the European Union's Horizon Europe Programme Under Grant Agreement no 101095290.

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Research Executive Agency (REA). Neither the European Union nor the granting authority can be held responsible for them.

UK participant in Horizon Europe Project SMIDGE is supported by UKRI grant number 10056282 (De Montfort University).

More info and contact: info@smidgeproject.eu | www.smidgeproject.eu

Funded by the European Union

Document Control

Deliverable	D7.1
WP/Task Related	WP7/ T 7.2
Delivery Date	December 2025
Dissemination Level	Public
Lead Author	UMIL
Lead Partner	UMIL
Contributors	
Reviewers	UCPH, DMU
Abstract	
Key Words	policy, recommendations, guidelines

Revision History

Version	Date	Author(s)	Reviewer(s)	Notes
0.1	15.12.2025	Giovanni Ziccardi, Paulina Kowalicka, Marco Castelli	Line Nybro Petersen (UCPH), Sara Wilford (DMU)	
0.2	19.12.2025	Giovanni Ziccardi, Paulina Kowalicka		

Table of Contents

Executive Summary and Methodology

This deliverable presents a set of five policy frameworks developed as part of the SMIDGE Project activities dedicated to analysing and combating misinformation, conspiracy narratives, and online extremism, with reference to different stakeholder groups. These frameworks are a central output of the project and reflect a structured effort to translate research findings into operational recommendations relevant to public policy. In addition to the data collected during the SMIDGE project, a survey was conducted between June and October 2025 to gather stakeholders' suggestions and assess their interest, in compliance with the guidelines of a Responsible Research and Innovation (RRI) approach. The twenty responses received allowed us to redefine the five stakeholder categories, enabling the production of more consistent documents. From the outset, the Project has adopted a guiding principle of avoiding uniform and generalized solutions, favouring instead the development of a coherent but differentiated framework, tailored to the specific roles, responsibilities, and constraints of different stakeholder groups.

All five policies share a common methodological basis, founded on a risk-based approach that respects fundamental rights, consistent with European values and the legal frameworks of the Union.

This approach includes a clear distinction between supervisory, accountability, and enforcement functions; an explicit focus on avoiding censorship-oriented language, favouring instead transparency, proportionality, accountability, and empowerment; and the alignment with the current European debate on disinformation, systemic risk, and social resilience.

At the same time, each policy framework has been drafted at a level of operational detail appropriate to its target audience, using language and concepts that are accessible, credible, and relevant to the respective stakeholder group.

The five policy frameworks are aimed respectively at: i) the general public; ii) policymakers and regulators; (iii) social media platforms; iv) actors in the information ecosystem, including journalists, media organisations, and content creators; and v) law enforcement and the judiciary. While differentiated according to their target audience, the policies are designed to be mutually consistent, avoid overlaps or conflicts of responsibility, and reflect a shared understanding of

disinformation and conspiracy theories as a systemic and societal challenge, rather than merely a content issue.

General definitions

Definitions tailored to the level of specialization and differentiated for each stakeholder are provided in the individual policies. For the general purposes of this work, these definitions may be used as a common reference:

- Disinformation: information that is verifiably false or misleading, created, presented, and disseminated intentionally with the aim of deceiving or manipulating the public, influencing political or social decisions, damaging the reputation of individuals, institutions, or organisations, or obtaining economic or strategic advantage, and that may cause public harm, including by undermining trust in democratic processes, public institutions, or justice systems.
- Misinformation: false or misleading information that is created or disseminated without the intent to deceive or cause harm, often shared by individuals who believe it to be true, but which may nonetheless distort public understanding, influence public perception, or interfere with democratic processes, judicial systems, or law enforcement activities, and can cause real-world harm despite the absence of malicious intent.
- Malinformation: information based on real or accurate facts that is used deliberately in a misleading way, including by being taken out of context, selectively disclosed, framed, or manipulated, with the aim of causing harm, such as damaging reputations, generating public confusion, obstructing justice, or contributing to political or social instability.
- Fake news: false or misleading information presented as legitimate news, often intended to mislead or influence public opinion.
- Conspiracy narratives (or conspiracy theories): narrative frameworks or story-based constructs that allege the existence of secret plots by powerful actors or hidden groups controlling events, are typically resistant to falsification or contrary evidence, and are used to delegitimise institutions, democratic processes, judicial or law-enforcement bodies, scientific knowledge, or social groups. These narratives are sociological and communicative in nature and must not be confused with the criminal-law concept of conspiracy. They often

undermine public trust, foster social polarization, and may generate online harassment or contribute to the legitimisation of offline violence.

- Online extremism: the online dissemination of ideological content that legitimises or promotes violence, hatred, or the systematic denial of fundamental rights, including content associated with violent extremism, terrorism, radicalisation narratives, or the rejection of democratic values. Such content often targets specific out-groups, fosters social division, or seeks to mobilise individuals toward harmful or extremist actions.
- Coordinated Inauthentic Behaviour (CIB): coordinated activity using deceptive identities or accounts to mislead users and manipulate public debate for strategic purposes.
- Digital platforms: online services or systems that facilitate the interaction, exchange, or dissemination of information, goods, or services among users, including social media networks, e-commerce websites, search engines, and other digital communication tools.
- Manipulated or synthetic media: audio, visual, or audiovisual content that has been artificially altered or generated, including through artificial intelligence systems, in a way that is likely to mislead or deceive recipients.
- AI-generated content: text, images, audio, or video created by artificial intelligence systems. While AI can provide useful outputs, it can also generate confident-sounding inaccuracies that may mislead audiences.

Overview of the five Policy frameworks

Policy for the General Public

This policy is written in plain, accessible language, intended for non-specialist users. It prioritises understanding of different types of false and misleading information, practical, everyday verification tools, responsible sharing and reporting behaviours and respectful dialogue and community resilience. The emphasis is on empowerment rather than blame, recognising that ordinary users are often exposed to manipulation without malicious intent.

Policy for Policymakers and Regulatory Bodies

This policy focuses on governance, coordination, and systemic risk mitigation. It deliberately avoids operational content moderation and instead clarifies regulatory roles and limits, oversight mechanisms, transparency and reporting expectations and coordination in crises and election

periods. The aim is to support policymakers in shaping coherent, proportionate, and rights-based responses to disinformation without encroaching on freedom of expression.

Policy for Social Media Platforms

This policy addresses platforms as active socio-technical actors, focusing on integrity of services (including coordinated inauthentic behaviour), algorithmic amplification and monetisation risks, graduated enforcement measures, transparency, user redress, and reporting and crisis and election integrity protocols. The policy reflects current best practices and public commitments of major platforms, while remaining neutral, non-prescriptive and suitable for an EU research context.

Policy for the Information Ecosystem (Journalists, Media, Creators)

This policy addresses journalists, editors, bloggers, influencers, and information providers. It builds on editorial responsibility and professional ethics, verification, correction, and transparency standards, handling of errors and accountability and interaction with user-generated content and automation tools. Rather than imposing moderation logic, it reinforces editorial quality, independence, and public trust.

Policy for Law Enforcement Authorities and the Judiciary

This is a deliberately cautious and institutionally sensitive policy. It avoids any language suggesting discretionary content moderation, fully respects judicial independence, due process, and separation of powers, focuses on institutional communication, risk awareness, and public trust and clarifies appropriate interaction with digital platforms within legal limits. The objective is to address disinformation affecting justice systems, without compromising the rule of law.

AI generated content

AI-generated content, in this document, refers generally to text, images, audio, or video produced wholly or partially by artificial intelligence systems. Such content is increasingly embedded in everyday digital communication and can serve legitimate, creative, educational, and informational purposes. At the same time, AI-generated content introduces specific risks within the information ecosystem, particularly when used to scale, personalise, or obscure misleading narratives.

A key characteristic of AI-generated content is its ability to produce fluent, confident, and contextually plausible outputs. This may result in the dissemination of inaccurate, fabricated, or

misleading information that appears authoritative or credible, even in the absence of malicious intent. When deployed strategically, AI-generated content can amplify existing disinformation narratives, facilitate coordinated inauthentic behaviour, or support online extremism by lowering costs, increasing volume, and enabling rapid adaptation of messages to different audiences.

AI-generated content may contribute to the delegitimisation of institutions, democratic processes, judicial or law-enforcement bodies, scientific knowledge, or social groups. These narratives are sociological and communicative in nature and must not be confused with the criminal-law concept of conspiracy. Their impact lies primarily in the erosion of public trust, the reinforcement of social polarisation, the normalisation of harassment, and, in some cases, the indirect legitimisation of offline harm or violence.

Particular risks arise when AI-generated content is combined with manipulated or synthetic media, such as deepfakes or synthetic audio, or when it is deployed at scale through automated accounts or coordinated networks. In such contexts, AI acts as an enabling technology rather than an independent cause of harm, accelerating dynamics that already exist within the digital information environment.

Across all the following policy frameworks, the response to AI-generated content is grounded in proportionality, transparency, and respect for fundamental rights. The focus is not on banning or suppressing AI-generated expression as such, but on mitigating systemic risks, enhancing user awareness, clarifying responsibilities of relevant actors, and strengthening societal resilience to manipulation.

AI-generated content therefore represents a cross-cutting issue that intersects with disinformation, online extremism, and coordinated inauthentic behaviour. Addressing it requires differentiated approaches tailored to the roles of users, platforms, media professionals, policymakers, and judicial authorities, while preserving freedom of expression, innovation, and democratic debate.

1. General Public

1. Who this Policy is for

This policy is designed for **everyday Internet users,** with particular attention to middle-aged users (45–65 years old)

It is intended for people who engage with digital platforms as part of their daily lives and who:

- use social media, messaging apps, websites, and search engines to stay informed;
- share news or opinions online with friends, family, or community groups;
- may not have professional training in media literacy, digital security, or fact-checking;
- want to better understand how misleading or harmful content works.

This policy does not require technical knowledge.

It is meant to be practical, supportive, and easy to follow.

2. What is misinformation?

Not all false information is the same. Understanding the difference helps you react in a calm, informed and appropriate manner.

Here are some key definitions:

Disinformation: false information shared on purpose to manipulate opinions, create confusion, damage reputations, or influence political or social decisions.

Misinformation: false information sometimes shared by mistake, often by people who believe it is true. Even without bad intentions, it can still cause harm.

Malinformation: true information used in a misleading way, for example by taking it out of context or presenting it selectively to deceive or attack someone.

Conspiracy narratives: stories claiming that powerful groups secretly control events, often without evidence and resistant to facts. These narratives frequently undermine trust in institutions, science, or democracy. They also frequently generate online harrasment and the legitimization of off-line violence as we see in D3.4 and D4.9

Online extremism: content that promotes hatred, violence, or the rejection of democratic values under the guise of protecting an in-group, often by targeting specific out-groups or encouraging division.

3. Why this matters

Misinformation and extremist narratives can have significant negative effects. It is a phenomenon that affects all age groups, but is particularly prevalent among middle-aged users (45–65 years old).

They may:

- foster a spiral of silence, where the public and public officials hold back on engaging in public debates
- foster fear, anger, and mistrust;
- damage or otherwise weaken personal relationships and community cohesion;
- undermine democratic processes;
- cause real-world harm, especially in sensitive areas such as health, migration, elections, and public safety.

At the same time, **freedom of expression is a fundamental right**. This policy does not seek to silence debate, criticism or unpopular opinions. Rather, it aims to support people in:

- making informed choices;
- avoiding the unintentional spread of harmful or misleading content;
- engaging in respectful, responsible and constructive online dialogue.

4. Your role as an Internet user

Every user plays a role in shaping the information environment.

You are encouraged to:

- **pause before sharing** content;
- **correct mistakes** when you realise that information is false;
- **think about real-world consequences**, not only online reactions;
- **respect others**, even when you disagree;
- **avoid amplifying fear, anger, or hatred**.

You do not need to be perfect. What matters is being thoughtful and responsible.

5. Before you share: a simple checklist

Before sharing any content, ask yourself the following questions:

- Do I recognise the source? If not, can I quickly check who is behind it?
- Is the story extremely shocking or emotional? If so, it may be designed to manipulate.
- Can I find the same information reported by **two or three reliable sources with editorial responsibility**?
- Is there a clearly identified author? Can I verify their identity or credibility?
- What do fact-checkers say? Try searching: "[claim] fact check".
- Is the content recent, or is it old or outdated news being reused?
- Are images or videos shown in their original context?
- Does the language seek to make me angry, scared, or outraged?
- Am I reacting emotionally rather than calmly?

If you are unsure, **it is always safer not to comment or share**.

6. Common warning signs

Be **cautious** if you encounter content that includes:

- the use of ALL CAPS and excessive punctuation (!!!) or heavy use of affective emojis;
- claims such as "they don't want you to know this";
- urgent requests to "share before it's deleted";
- anonymous or unnamed sources;

- Known influencers using click bate tactics such as the above examples to push engagement;
- poor spelling or grammar;
- sensational headlines that do not match the substance of the article;
- oversimplified explanations of complex issues;
- websites that imitate the appearance of well-known news outlets.

No single sign is sufficient to conclude that content is false. However, the presence of several of these indicators together should make you raise doubts and prompt increased caution.

7. What to do when you see suspicious content?

✓Report it

You should report content that involves:

- information you know to be false;
- manipulated or synthetic images or videos;
- impersonation accounts or identity misuse;
- coordinated spam or automated (bot) activity;
- harmful health-related misinformation;
- attempts to interfere with elections or civic processes.

When reporting, you should:

- use the reporting tools made available by the platform;
- include links and any relevant contextual details;
- follow up if a reference number is provided.

Reporting suspicious content helps platforms and online communities respond more effectively.

8. Talking to others about false information

✗If someone you know shares misleading content, avoid attacking, shaming or humiliating them.

✓Instead, consider the following approaches:

- ask open and curious questions, such as: “Where did this information come from?”
- express concern in a calm and respectful manner, for example: “I have read that this claim might not be accurate”.
- share reliable sources as alternative references;
- acknowledge emotions, for instance: “I understand why this is upsetting”.Respectful and constructive conversations are generally more effective than confrontation.

9. Your positive impact

By acting responsibly, you can:

- reduce the spread of harmful or misleading content by not engaging with it online;
- help protecting your community and family;
- encourage healthier and more constructive online discussions by leaving room for doubt and uncertainty;
- support reliable and trustworthy information;
- contribute to a safer and more respectful digital environment.
- be critical of the platforms you use. Who benefits from your engagement and how? Should you move your engagement to other platforms?

Small actions, when repeated consistently by many people, can make a meaningful difference.

10. Final message

You do not need to be an expert to help make the internet a better place.
Being curious, cautious, and respectful is enough.

Think before you share.
Check before you trust.
Pause before you react.

2. Policymaker and regulatory bodies

1. Introduction and purpose

This policy is conceived as **regulatory guidance for policymakers and regulatory bodies**, not as an internal content-moderation policy of a single platform or entity.

Its purpose is to provide a coherent governance framework for addressing disinformation, conspiracy theories and online extremism through risk-based oversight, coordination, transparency and accountability, in line with European Union legal and policy instruments.

2. Scope and regulatory role

This policy applies to public authorities, independent regulators and regulatory coordination bodies involved in digital services, media, electoral integrity, consumer protection and fundamental rights.

It does not confer direct content-moderation powers unless explicitly provided by law. Rather, it defines oversight and supervisory expectations; cooperation mechanisms with digital service providers and other stakeholders; systemic-risk mitigation obligations and transparency and accountability standards.

The framework is grounded in existing European instruments, including but not limited to the Digital Services Act (DSA), the Code of Practice on Disinformation and relevant national media and electoral laws.

3. Scope of application

This policy concerns regulatory oversight of content dissemination and amplification through digital services and online intermediaries, including social media platforms, video-sharing platforms, search engines, online advertising systems, messaging services and other services within the scope of applicable European and national law.

While the recommendations outlined in this document are applicable to all age groups, they have been developed with particular attention to middle-aged users (45–65), who have been identified by the SMIDGE project as the age group most vulnerable to misinformation.

4. Definitions

Disinformation: verifiably false or misleading information that is created, presented, and disseminated for economic gain or to intentionally deceive the public and cause public harm.

Misinformation: false or misleading information shared without intent to cause harm.

Malinformation: information that is based on reality but used out of context, manipulated or selectively disclosed to cause harm.

Conspiracy theories / conspiratorial narratives: narrative frameworks alleging secret plots by powerful actors, typically resistant to falsification and often used to delegitimise institutions, democratic processes, or scientific knowledge. This definition is sociological and communicative in nature and must not be confused with the criminal-law concept of conspiracy.

Online extremism: the dissemination online of ideological content that legitimises or promotes violence, hatred, or the systematic denial of fundamental rights, including but not limited to violent extremism, terrorist propaganda and radicalisation narratives.

Manipulated or synthetic media: audio, visual, or audiovisual content that has been artificially altered or generated (including through AI systems) in a manner that is likely to mislead recipients.

5. Responsibility and governance

Regulatory bodies should ensure clear internal governance arrangements that guarantee independence, expertise, transparency, and accountability in the exercise of their supervisory and coordination functions. This includes:

- clear allocation of regulatory and supervisory responsibilities;
- mechanisms to manage conflicts of interest;
- access to technical, legal, and scientific expertise;
- periodic public reporting on regulatory activities and outcomes.

6. Guiding principles

6.1 Fundamental rights and proportionality

All regulatory action must respect freedom of expression, due process and the principle of proportionality. Misinformation, as such, is not necessarily illegal; regulatory responses must therefore focus on risk mitigation and transparency, rather than blanket content removal.

6.2 Risk-based and evidence-driven governance

Measures shall be calibrated to the nature, scale, and impact of identified risks, taking into account empirical evidence and contextual factors such as elections, public health emergencies, or security crises.

6.3 Shared responsibility

Addressing disinformation and conspiratorial narratives requires coordinated action among regulators, platforms, media actors, civil society, and research institutions.

Regulatory bodies should promote and, where legally appropriate, require:

- accuracy and reliability of information processes;
- transparency regarding content-related interventions;

- impartial and non-discriminatory practices;
- user empowerment and access to redress;
- evidence-based decision-making.

7. Detection, monitoring and risk assessment

Regulatory bodies should establish or require mechanisms for:

- continuous monitoring of systemic information risks;
- periodic systemic risk assessments, particularly in relation to electoral processes, public health, and social cohesion;
- post-incident reviews to identify lessons learned and structural vulnerabilities.

Monitoring activities must be transparent, privacy-preserving, and proportionate.

8. Regulatory intervention framework

Regulatory bodies should apply a graduated and procedurally fair intervention ladder, which may include:

- Requests for information or clarification;
- Structured dialogue and voluntary mitigation commitments;
- Enhanced monitoring and reporting obligations;
- Formal compliance actions or investigations, where legally grounded;
- Sanctions or penalties strictly within the applicable legal framework.

All interventions must be accompanied by clear statements of reasons and accessible redress mechanisms, in line with due-process guarantees.

9. Whistleblowing and external reporting

Regulatory bodies should ensure the availability of secure and accessible channels for reporting systemic disinformation risks, coordinated manipulation, or regulatory

non-compliance. Whistleblowers acting in good faith must be protected in accordance with applicable whistleblower-protection legislation.

10. Coordination and crisis response

Regulatory bodies should define coordination protocols with other national regulators (media, elections, cybersecurity); Digital Services Coordinators and EU-level bodies, where applicable and trusted research and fact-checking networks.
Special crisis protocols should be activated during election periods or emergencies, including rapid information exchange and coordinated public communication strategies designed to avoid amplification of false narratives.

11. Transparency and Reporting

Regulatory bodies should require regular public reporting, including aggregated data on:

- number and categories of reports received;
- response times and types of measures adopted;
- appeal and reversal rates;
- recurring disinformation narratives (in anonymised form);
- actions undertaken during high-risk periods.

Transparency reports should enable public scrutiny while safeguarding personal data and trade secrets.

12. Education, awareness and societal resilience

Policy measures should support:

- media and digital literacy programmes across age groups;
- prebunking and inoculation strategies addressing common manipulation techniques;
- independent journalism and pluralistic information ecosystems;
- research and evaluation of the effectiveness of awareness campaigns.

13. Sanctions

Sanctions may only be applied by competent authorities within the limits and procedures established by applicable law.

14. Review and adaptation

This policy framework should be reviewed periodically to reflect technological developments, emerging risks, and empirical evidence regarding the effectiveness of adopted measures.

3. Social media platform

1. Introduction and purpose

This policy establishes the principles, governance mechanisms, and operational measures adopted by the **platform** to prevent, mitigate and address the spread of disinformation, misinformation, malinformation, conspiracy narratives and online extremism.

The policy applies to user-generated content and platform functionalities, including content dissemination, amplification, monetisation, and recommendation systems. An active age group online is that of middle-aged users (45–65 years old), whose specific vulnerabilities require particular consideration. This policy takes these risks into account and aims to safeguard the integrity of public discourse while respecting freedom of expression, due process, and applicable legal obligations.

2. Scope of application

This policy applies to:

- user-generated content (UGC), including text, images, audio, video, and live content;
- accounts, pages, groups, channels, and networks of accounts;
- platform systems for ranking, recommendation, and amplification of content;

- advertising and sponsored content, including political and issue-based advertising where applicable;
- sharing, re-sharing, and linking functionalities that significantly contribute to content dissemination.

3. Definitions

Disinformation: verifiably false or misleading information that is created, presented, and disseminated intentionally to deceive the public or to obtain economic or strategic advantage, and which may cause public harm.

Misinformation: false or misleading information shared without the intent to cause harm.

Malinformation: information based on real facts that is used out of context, selectively disclosed, or manipulated to cause harm.

Conspiracy theories / conspiratorial narratives: narrative frameworks alleging secret plots by powerful actors, typically resistant to falsification and often used to delegitimise institutions, democratic processes, scientific knowledge, or social groups.

Online extremism: the online dissemination of ideological content that legitimises or promotes violence, hatred, or the systematic denial of fundamental rights, including violent extremism, terrorist propaganda, and radicalisation narratives.

Manipulated or synthetic media: audio, visual, or audiovisual content that has been artificially altered or generated, including through AI systems, in a manner likely to mislead recipients.

Coordinated Inauthentic Behaviour (CIB): coordinated activity using deceptive identities or accounts to mislead users and manipulate public debate for strategic purposes.

4. Objectives and guiding principles

The platform is guided by the following principles:

- **accuracy and reliability**: promoting a healthy information ecosystem and reducing the dissemination of content that is demonstrably false or misleading and that may cause harm to individuals, communities, or public interests;

- **freedom of expression and proportionality**: applying proportionate measures, while recognising the legitimacy of public interest discourse, satire, criticism, and the contextual use of information;
- **transparency and accountability**: providing clear rules, reasoned and explainable decisions, and regular public reporting;
- **non-discrimination**: enforcing policies impartially and mitigating risks of bias;
- **user empowerment**: enabling users to understand the context of the contest encountered, the sources behind it, and the remedies available in the event of enforcement actions;
- **service integrity**: preventing manipulation of the platform through deceptive practices, automated amplification, and coordinated abuse.

5. Governance, roles, and accountability

The platform maintains internal governance structures to ensure effective implementation of this policy, including:

- policy and integrity functions responsible for defining and updating rules;
- trust and Safety operations for content review, escalation, and quality assurance;
- data science and engineering teams responsible for safety-by-design in recommendation and amplification systems;
- legal and compliance functions ensuring adherence to applicable laws and cooperation with competent authorities;
- transparency and reporting functions responsible for public accountability.

6. Policy domains

6.1 Harm-Based Misinformation

Enhanced measures are applied where false or misleading information may cause real-world harm, including risks to public health, democratic processes, public safety, or social cohesion.

6.2 Integrity of Services

The platform prohibits and enforces against:

- coordinated inauthentic behaviour and influence operations;
- impersonation and deceptive identities, except for clearly labelled parody or satire;
- artificial or automated amplification designed to manipulate engagement metrics.

6.3 Manipulated and Synthetic Media

Manipulated or synthetic media that may mislead users is subject to labelling, reduced distribution, or removal, depending on the level of risk and potential harm.

6.4 Online Extremism

This policy operates in coordination with existing rules addressing terrorism, violent extremism, hate speech, incitement to violence, and related abuses.

7. Detection and monitoring

The platform employs a combination of:

- user reporting mechanisms with clear categorisation and feedback;
- human review with escalation for high-risk or sensitive cases;
- automated detection tools to identify coordinated activity and anomalous patterns;
- cooperation with independent fact-checking organisations as qualified sources of verification and context.

Automation is used as an assistive tool and is subject to quality controls and periodic evaluation.

8. Enforcement measures

Enforcement actions are applied based on severity, potential harm, intent, coordination, and recurrence, and may include:

- friction and prompts to encourage informed sharing;
- labelling and contextual information;
- downranking and reduced algorithmic amplification;
- limits on sharing, forwarding, or engagement;
- restrictions on monetisation and advertising;

- removal of content in cases of illegality or serious harm;
- account-level measures, including warnings, feature limitations, suspension, or termination;
- network-level action against coordinated inauthentic behavior.

9. Appeals and User Redress

Users receive clear notification of enforcement actions, including the reasons for the decision and available remedies. The Platform provides accessible appeal mechanisms, human review for high-impact cases, and periodic publication of aggregate appeal and reversal statistics.

10. Transparency and Reporting

The Platform publishes regular transparency reports with aggregated data on:

- enforcement actions by category;
- response times and handling volumes;
- appeals and reversal rates;
- measures addressing coordinated manipulation and integrity threats;
- advertising-related enforcement where applicable.

Reports are designed to enable public scrutiny while protecting personal data and legitimate commercial interests.

11. Crisis and Election Integrity Protocols

During periods of heightened risk, such as elections or public emergencies, the Platform activates enhanced measures, including rapid escalation, temporary limitations on virality, strengthened labelling and context, and coordination with competent authorities and trusted partners.

12. Education and User Empowerment

The platform supports media literacy and resilience through:

- educational campaigns on manipulation techniques;

- contextual tools providing information about sources and accounts;
- pre-bunking initiatives addressing common disinformation strategies;
- evaluation of effectiveness through appropriate metrics.

13. Third-Party Partners and Advertisers

Partners, advertisers, and third-party service providers are required to comply with applicable integrity and transparency standards. Violations may result in restrictions, suspension, or termination of partnerships, in accordance with contractual and legal frameworks.

14. Review and Adaptation

This Policy is reviewed periodically to reflect technological developments, emerging risks, regulatory changes, and evidence regarding the effectiveness of adopted measures.

4. Journalists, Media Organisations, Content Creators and Information Providers

1. Purpose and Scope

This policy establishes common principles and professional standards for individuals and organisations involved in the **production, publication and dissemination of information**, including journalists, editors, bloggers, influencers, digital media outlets, and other content creators.

Its purpose is to prevent and mitigate disinformation, conspiracy narratives and online extremism, while safeguarding freedom of expression, editorial independence, and the public's right to be informed. While the recommendations outlined in this document are applicable to all age groups, they have been developed with particular attention to middle-aged users (45–65), who have been identified by the SMIDGE project as the age group most vulnerable to misinformation.

This policy does not replace existing journalistic codes of ethics or legal obligations. It complements them by addressing risks specific to the contemporary digital information environment.

2. Scope of Application

This policy applies to:

- editorial content, articles, videos, podcasts, and multimedia products;
- press releases and official statements produced for public dissemination;
- content published on websites, newsletters, and verified social media accounts;
- collaborations with freelancers, contributors, influencers, and third-party content providers;
- sponsored, branded, or partnership content, where editorial integrity may be at risk.

3. Definitions

Disinformation: False or misleading information created and disseminated intentionally to deceive the public, manipulate opinions, influence political or social processes, or cause reputational or societal harm.

Misinformation: False or misleading information shared without intent to cause harm, which may nonetheless distort public understanding or contribute to harmful outcomes.

Malinformation: Information based on real facts that is used out of context, selectively framed, or presented in a misleading manner to cause harm, confusion, or delegitimisation.

Conspiracy theories / conspiratorial narratives: Narratives alleging secret plots by powerful actors, typically resistant to verification or falsification, and often used to undermine trust in institutions, science, journalism, or democratic processes. This concept is distinct from the criminal-law offence of conspiracy.

Online extremism: The dissemination of content that promotes hatred, violence, or the systematic rejection of democratic values and fundamental rights, including content linked to radicalisation or terrorist ideologies.

Manipulated or synthetic media: Audio, visual, or audiovisual content that has been altered or generated, including through artificial intelligence systems, in a manner likely to mislead audiences.

4. Guiding Principles

Information producers commit to the following principles:

- **accuracy and verification**: ensuring that information is verified, contextualised, and based on reliable sources;
- **editorial independence**: preserving autonomy from political, economic, or ideological pressures;
- **freedom of expression**: protecting pluralism, debate, and criticism, while avoiding the dissemination of demonstrably false or harmful content;
- **transparency**: clearly distinguishing facts, opinions, and sponsored content;
- **impartiality and fairness**: avoiding discriminatory or misleading framing;
- **accountability**: acknowledging and correcting errors promptly and visibly.

5. Content Production and Verification Standards

Before publication, content producers should:

- verify information using primary and independent sources;
- corroborate sensitive claims through multiple reliable sources;
- consult subject-matter experts where appropriate;
- assess the context, timing, and potential impact of publication;
- clearly label archival, illustrative, or reconstructed materials;
- ensure that images, videos, and audio are not presented in a misleading manner.

Special caution should be exercised when reporting on elections, public health, emergencies, judicial proceedings, or vulnerable groups. Special attention should be used when reporting on conspiracy theories and understand that the online environments around conspiracy theories often have a dimension of online harassment and legitimization of violence (see D4.7 and D4.9), which reporting may inadvertently enhance by making an

online environment or online actors visible to a larger audience, while lending them the authority of legacy media with editorial responsibility.

6. Handling Errors and Corrections

When false or misleading information is identified:

- corrections or updates should be issued promptly and clearly;
- the correction should explain the nature of the error and provide accurate information;
- where appropriate, the correction should be disseminated through the same channels as the original content;
- internal records of corrections should be maintained to support learning and improvement.

Corrections should be transparent and proportional, without attempting to conceal or minimise mistakes.

7. Detection of Disinformation and Risk Awareness

Information producers should develop awareness mechanisms that enable them to identify coordinated campaigns aimed at manipulating public debate, the reuse of recycled or out-of-context content presented as current, the dissemination of manipulated or synthetic media, and the construction of narratives designed to provoke fear, outrage, or polarisation.

Where appropriate, collaboration with independent fact-checking organisations and research institutions is encouraged.

8. Use of Automated Tools and Technology

Automated tools, including artificial intelligence systems, may be used to support verification, monitoring, or analysis, provided that:

- editorial judgment remains central to decision-making;
- automated outputs are reviewed critically by human editors;
- limitations, biases, and error rates are understood and mitigated;
- transparency is ensured regarding the role of automation.

9. Interaction with User-Generated Content

Where information producers host or curate user-generated content:

- clear community standards should be published;
- reporting mechanisms should be accessible and understandable;
- moderation decisions should be consistent and explainable;
- lawful and legitimate expression should not be suppressed solely because it is controversial.

10. Whistleblowing and Internal Reporting

Secure channels should be available for employees, collaborators, and contributors to report, in good faith, concerns related to disinformation risks or breaches of editorial standards.

Whistleblowers shall be protected from retaliation in accordance with applicable whistleblower-protection laws and ethical standards.

11. Education, Training, and Public Engagement

Information producers should invest in:

- continuous training on verification techniques and emerging manipulation tactics;
- media literacy initiatives aimed at audiences and communities;
- partnerships with educational institutions, fact-checking organizations, and civil society;
- transparency about editorial practices and decision-making processes.

The effectiveness of such initiatives should be periodically evaluated.

12. Governance and Responsibility

Media organisations and content producers should define clear internal responsibilities that ensure effective editorial oversight and quality control, compliance with legal and ethical standards, and transparency and accountability toward the public.

Registered professional journalists remain subject to their professional codes of ethics and disciplinary regimes.

13. Review and Adaptation

This Policy shall be reviewed periodically to reflect technological developments, evolving information threats, changes in professional standards, and empirical evidence on the impact of disinformation.

4.1. Information Area

1. Definitions

This policy defines the guidelines for identifying, preventing, and countering disinformation on the official channels of the **company** and the **digital platforms** managed by it.

This policy applies to:

- the intentional spread of false or misleading information aimed at manipulating public opinion, influencing political decisions, or damaging the reputation of individuals, institutions, or organizations (**disinformation**);
- the spread of false content, even if unintentional, which may distort public perception (**misinformation**). This includes certain types of misinformation that can cause real-world harm, certain types of technically manipulated content, or content interfering with democratic processes;
- the deliberate use of accurate information in a misleading or harmful context to cause reputational damage, political instability, or public confusion (**malinformation**).

Other relevant definitions for the application of these policies are:

- **conspiracy**: a criminal agreement between two or more parties to commit an illegal act or to accomplish a lawful act through unlawful means. The agreement itself constitutes the crime, regardless of whether the intended act is carried out;

- **digital platforms**: online services or systems that facilitate the interaction, exchange, or dissemination of information, goods, or services among users. These platforms may include social media networks, e-commerce websites, search engines, and other digital communication tools;
- **fake news**: false or misleading information presented as legitimate news, often intended to mislead or influence public opinion.

2. Objectives and Guiding Principles

The company commits to:

- **accuracy and reliability**: ensuring the accuracy and reliability of shared and produced content, based on verified and authoritative sources;
- **freedom of expression**: protecting freedom of expression by guaranteeing the right to debate and criticism, while respecting truth and accuracy. The company will ensure that content moderation measures are proportionate and necessary;
- **transparency**: operating with transparency by making public the reasons behind content-related interventions;
- **impartiality**: guaranteeing impartiality by avoiding discrimination based on political, ideological, religious, or cultural grounds;
- **user empowerment**: empowering users by promoting a responsible use of platforms.

3. Scope of Application

This policy applies to all content distributed through:

- press releases and official statements;
- official websites;
- social media channels and, in general, content shared through third-party platforms where the company has an official presence;
- promotional and informational materials;
- content produced internally or acquired from third parties;

4. Content Production Guidelines

The following verification procedure shall be followed:

- **editorial review process**: all content produced by internal staff must undergo fact-checking before publication. Senior editorial approval must be obtained for sensitive topics.
- **social media management and visual content**: staff responsible for social media must receive dedicated training on fact-checking techniques, monitor shared content in real time, and follow rapid response protocols for corrections. All visual content must be clearly labelled when using stock images, illustrations, or recreations, must not be manipulated in ways that alter meaning, and must provide proper attribution for third-party material.

5. Detection and Monitoring Strategies

To identify and counter disinformation, the following measures will be adopted:

- Automated monitoring
    - AI-powered anomaly detection systems;
    - Pattern recognition for suspicious content distribution;
    - Real-time threat assessment algorithms.
- Independent fact-checking
    - Collaboration with certified fact-checking organizations;
    - Cross-verification with authoritative sources;
    - International fact-checking partnerships.
- Citizen reporting
    - Accessible reporting mechanisms for public;
    - Transparent review processes;
    - Feedback loops to reporters.

If necessary, the following additional measures may be implemented:

- collaboration with independent fact-checkers: Verification of reports through qualified external partners;
- automated and manual monitoring: Use of artificial intelligence tools to detect anomalies in content and suspicious reports.

Control is organized on three levels, based on the origin of the content:

- **Internal Editorial Control**
    - Source verification: content produced internally will be subject to an editorial review process to verify the accuracy of the information;
    - Guidelines for social media managers and editorial staff: editorial staff will receive ongoing training on fact-checking techniques and critical evaluation of sources;
    - Legal supervision: the legal team will ensure that the content complies with current regulations on information and communication.
- **Control of Third-Party Content**
    - Licenses and partnerships: content acquired from third parties will be evaluated based on criteria of reliability and compliance with the company's policy;
    - Termination of agreements: the spread of false or misleading content by external partners may result in the termination of commercial agreements.
- **Monitoring of User-Generated Content (UGC)**
    - Reporting and review: users will be able to report suspicious content through a dedicated function;
    - Automated and manual moderation: user-generated content will be examined through automated systems and manual review by a dedicated team.

6. Whistleblowing

Employees and collaborators are encouraged to report any suspicious activities that may involve the spread of disinformation, including the dissemination of false or misleading information in internal communications, the sharing of unverified information on external platforms that could harm the organisation's reputation, and the intentional or negligent creation of misleading content.

Reports can be made through the following channels:

- a dedicated whistleblowing hotline;
- an encrypted online reporting platform;
- direct communication with the designated Ethics and Compliance Officer.

All reports of disinformation will be treated confidentially. Whistle-blowers will be protected from any form of retaliation, including termination of employment, harassment, or discrimination, in accordance with applicable whistleblower protection laws. The organization is committed to ensuring that any individual who raises a concern in good faith is supported and protected throughout the process.

Whistleblower protection will be extended to external parties who report in good faith on potential disinformation practices involving the company.

7. Intervention Measures

When disinformation is identified, content managers should act promptly and transparently.

For own content, false information must be removed or updated within two hours of verification, a clear correction explaining the error should be published, the audience should be notified through the same channels as the original content, and an internal record should be maintained for learning purposes.

For third-party content, warning labels with links to verified sources should be added, visibility in recommendations or feeds should be reduced, clearly false or harmful content should be removed, and the reliability of the source should be reviewed for future collaborations.

8. Education and Awareness

In order to increase awareness among employees and collaborators on the topics of information verification, the following actions may be implemented:

- **Awareness and media literacy campaigns**: organizing information campaigns on the dangers of disinformation and fact-checking techniques;
- **Internal training**: training courses for employees and collaborators on source verification and information management. Training should include simulations that allow participants to understand the real risk of exposure to disinformation;
- **Partnerships with educational institutions**: promoting school and university projects to develop critical thinking and the ability to assess sources;
- **Transparency on editorial practices**: publishing editorial guidelines and regular reports on fact-checking and moderation activities.

Educational campaigns will be conducted in partnership with fact-checking organizations and NGOs specializing in media literacy.

The effectiveness of media literacy campaigns will be periodically evaluated through user surveys and performance indicators.

Users will be informed of these activities and encouraged to follow the forum moderation guidelines.

9. Appeals and Review Mechanisms

Appeal and review procedures should take into account the following steps:

- **appeal against removal**: users may appeal the removal of content through a dedicated form;
- **review by an independent committee**: disputes will be evaluated by a review committee operating under impartiality and transparency criteria;
- **periodic policy review**: this policy will be reviewed annually to adapt to new challenges and changes in the digital environment.

Users will receive a detailed explanation of the decision, including references to the violated policy provisions.

10. Responsibility and Governance

The main levels of responsibility can be identified in:

- **Editorial responsibility**: the Editorial Director and the Quality Control Committee will be responsible for enforcing this policy;
- **Legal supervision**: the legal team will ensure that the policy complies with national and international regulations on communication and broadcasting;
- **Public reporting**: the Company will publish an annual report on removed content, received reports, and actions taken.

All staff involved in content moderation will undergo periodic training on new disinformation trends and technological developments.

11. Sanctions

Violations of this policy by employees, collaborators, or partners may result in disciplinary measures such as an official warning, the temporary or permanent suspension of access to company systems, the termination of contractual agreements, or legal action in cases of serious violations or reputational damage, with any criminal acts being reported to the competent authorities. In all cases, registered professional journalists retain their ethical responsibility.

5. Law Enforcement Authorities and the Judiciary

1. Purpose and Institutional Context

This policy establishes a common framework for addressing disinformation, conspiracy narratives, and online extremism within the activities of law enforcement authorities and judicial bodies.

It is not a content-moderation policy and does not introduce new investigative or adjudicatory powers. Its purpose is to support **institutional integrity, public trust, and the proper administration of justice**, while fully respecting judicial independence, due process, and fundamental rights.

2. Scope of Application

This policy applies to:

- institutional communications issued by law enforcement authorities and judicial bodies;
- official digital channels, including websites and verified social media accounts used for public information purposes;
- interactions with digital platforms and other stakeholders in the context of investigations, judicial proceedings, or public information activities;
- internal awareness, training, and risk-prevention measures related to disinformation affecting justice and public security.

This policy does not apply to judicial decision-making on the merits of cases, evidentiary assessments, or prosecutorial discretion, which remain governed exclusively by the applicable law.

3. Definitions

Disinformation: verifiably false or misleading information disseminated intentionally to deceive the public or cause harm, including by undermining trust in justice systems, public institutions, or democratic processes.

Misinformation: false or misleading information shared without intent to cause harm, which may nonetheless distort public understanding or interfere with judicial or law enforcement activities.

Malinformation: information based on real facts that is used out of context, selectively disclosed, or framed in a misleading manner to cause reputational damage, obstruct justice, or generate public confusion.

Conspiracy theories / Conspiratorial narratives: narrative constructs alleging secret plots by powerful actors, typically resistant to falsification, and frequently used to delegitimise judicial institutions, law enforcement agencies, democratic governance, or scientific knowledge. This definition is sociological and communicative in nature and must not be confused with the criminal-law concept of conspiracy.

Online extremism: the online dissemination of ideological content that legitimises or promotes violence, hatred, or the systematic denial of fundamental rights, including content associated with violent extremism, terrorism, or radicalisation.

Manipulated or synthetic media: audio, visual, or audiovisual content that has been altered or generated, including through artificial intelligence systems, in a manner likely to mislead recipients.

4. Objectives and Guiding Principles

Law enforcement authorities and judicial bodies act in accordance with the following principles:

- **institutional accuracy and reliability**: ensuring that official communications are factually accurate, contextualised, and based on verified information;
- **judicial independence and neutrality**: preserving the impartiality of judicial authorities and avoiding any action that could compromise the perception or reality of independence;
- **freedom of expression and legality**: respecting freedom of expression and the right to information, within the limits established by law;
- **proportionality and necessity**: adopting only those measures that are strictly necessary and proportionate to legitimate institutional objectives;
- **avoidance of institutional amplification**: refraining from actions that may unintentionally amplify false, misleading, or manipulative narratives;
- **public trust and transparency**: communicating in a manner that strengthens confidence in justice and public institutions without prejudicing proceedings.

5. Institutional Communications and Public Information

Institutional communications by law enforcement and judicial bodies shall be carefully framed and strictly limited to the information necessary to ensure public understanding, safety, and institutional transparency.

They shall avoid commentary of any kind on ongoing proceedings that could – directly or indirectly – prejudice investigations or trials or affect the rights of the parties involved. Communications should rely on neutral, precise, and non-sensational language, avoiding speculation, value judgements, or premature conclusions.

Whenever information is shared, it should clearly distinguish between verified facts, procedural or contextual information, and legal outcomes.

Special care shall be exercised in high-profile cases, crises, or matters involving heightened public sensitivity, where media attention and public reaction may amplify the impact of institutional statements. In such contexts, communications should be guided by principles of proportionality and careful respect for the presumption of innocence until a final and binding judgement.

6. Risk Awareness and Monitoring

Law enforcement and judicial authorities should develop internal awareness mechanisms to identify and assess risks posed by disinformation.
Those risks include:

- coordinated campaigns aimed at delegitimising judicial institutions or interfering with proceedings;
- manipulated or synthetic media targeting judges, prosecutors, investigators, or witnesses;
- narratives that encourage harassment, intimidation, or violence against institutional actors.

Any monitoring activity shall be conducted in accordance with applicable law and shall not involve mass surveillance or disproportionate data collection.

7. Interaction with Digital Platforms and External Actors

Where legally appropriate, law enforcement and judicial authorities may engage with digital platforms and other relevant stakeholders to address risks related to disinformation and harmful online content. Such engagement may include reporting potentially illegal content, or content that poses serious risks to public safety or to the proper administration of justice.

Authorities may also request the preservation of digital evidence, in accordance with applicable procedural law, where this is necessary to safeguard investigations or judicial proceedings. In addition, cooperation with platforms and other stakeholders may extend to transparency and contextualisation initiatives, provided that such cooperation does not result in the delegation of judicial or investigative responsibilities.

All interactions of this kind should be conducted in full respect of due process, applicable jurisdictional limits, and the principle of separation of powers.

8. Response Measures

Responses to disinformation affecting law enforcement or judicial activities may take different forms, depending on the nature and impact of the content involved.
Such responses may include:

- institutional clarification or correction issued through official channels;
- public information initiatives aimed at restoring factual accuracy and providing appropriate context;
- procedural measures provided by law, where disinformation constitutes a criminal offence or otherwise interferes with the proper administration of justice.

Judicial or law enforcement authorities should refrain from engaging in discretionary content removal or imposing account-level sanctions outside the legal frameworks that govern such actions, to preserve due process and the separation of institutional roles.

9. Whistleblowing and Internal Reporting

Secure and confidential channels shall be made available for the good faith reporting of concerns related to disinformation risks that may affect institutional integrity or public trust.

Whistleblowers shall be protected in accordance with whistleblower-protection legislation in force, including anonymisation where applicable.

Safeguards against retaliation shall apply to both internal and, where relevant, external reporting procedures.

10. Training and Capacity Building

Authorities should promote continuous training for judges, magistrates, prosecutors, investigators, and staff on:

- information disorders and manipulation techniques;
- risks posed by synthetic media and emerging technologies;
- ethical and legal constraints governing institutional communication;
- protection of judicial independence in digital environments.

Training activities should be evidence-based and tailored to specific institutional roles.

11. Transparency and Accountability

Without prejudice to confidentiality obligations and procedural secrecy, authorities should actively promote transparency as means of strengthening institutional accountability and public trust. This may be achieved through the publication of general guidelines on institutional communication, as well as through periodic reporting on awareness-raising and training activities carried out within the institution.

Where appropriate, transparency efforts may also include cooperation with independent research bodies and monitoring initiatives, with a view to supporting informed public debate and improving institutional practices, while fully respecting legal constraints and institutional independence.

12. Review and Adaptation

This Policy shall be reviewed periodically to reflect technological developments, evolving disinformation tactics, changes in the legal framework, and empirical evidence concerning risks to justice systems and public trust.

Annex

ANNEX I

POLICY GUIDANCE FOR MIDDLE-AGED INTERNET, SOCIAL MEDIA AND AI USERS (45–65)

Navigating Disinformation, Conspiracy Narratives and Emerging Technologies

1. Purpose of this Annex

This Annex provides practical guidance for adults aged approximately **45 to 65** who use:

- social media platforms and messaging apps;
- online news websites and search engines;
- video platforms and podcasts;
- AI-based tools such as chatbots, image generators, voice assistants, or automated writing tools.

It aims to support informed, confident, and responsible digital participation, without requiring technical expertise or advanced digital skills.

This document does not assign blame. It recognises that digital environments have changed rapidly and that adapting to them is a shared societal challenge.

2. Why this Age Group is particularly Exposed

Users in this age range often:

- rely on the internet (and social media in particular) as a primary source of news and information;
- combine traditional media habits with social media and messaging platforms;
- participate in family, community, or professional online groups;

- increasingly encounter AI-generated content without clear labels.

At the same time, they may face:

- information overload;
- highly emotional or polarising content;
- content shared by trusted contacts rather than verified sources;
- uncertainty about how AI systems generate information.

These factors can increase exposure to misleading, manipulative, or false content, even without active searching.

3. Understanding Today's Information Environment

The online information environment is different from traditional media because:

- anyone can publish and reach large audiences;
- algorithms prioritise engagement, not accuracy;
- content may be personalised to reinforce existing views;
- AI systems can generate realistic text, images, audio, and video at scale.

This does not mean that everything online is false, but it does mean that extra attention and care are required.

4. Key Concepts Explained Simply

- **Disinformation:** false information shared intentionally to mislead or manipulate.

- **Misinformation:** False information shared by mistake, often by people who believe it is true.
- **Malinformation:** True information used out of context or in a misleading way to cause harm.
- **Conspiracy narratives:** Stories claiming that hidden groups secretly control events, often rejecting evidence and undermining trust in institutions.
- **AI-generated content:** Text, images, audio, or video created by artificial intelligence systems. AI can be useful, but it can also generate confident-sounding inaccuracies.

5. Practical Guidance for Everyday Use

5.1 Before Believing or Sharing Content

Pause and ask yourself:

- Who created this content?
- Where does it come from?
- Is it confirmed by more than one reliable source?

- Does it appeal mainly to fear, anger, or outrage?
- Is it asking me to share urgently?

If something feels rushed or emotionally charged, slowing down is often the best response.

5.2 Recognising Common Warning Signs

Be cautious if content includes:

- dramatic or alarmist language;
- claims that "mainstream media is hiding the truth";
- anonymous or unclear sources;
- old events presented as current;
- images or videos without context;
- promises of simple answers to complex problems.

One warning sign alone is not proof—but several together deserve attention.

6. Using Social Media and Messaging Apps Responsibly

When using platforms such as Facebook, WhatsApp, Telegram, X, or similar services:

- avoid forwarding messages without verification;
- remember that content shared by friends or relatives can still be incorrect;
- use platform reporting tools when you encounter harmful or misleading content;
- do not feel obligated to respond or engage with provocative posts.

Silence and non-sharing are sometimes the most effective responses.

7. Understanding and Using AI Tools Safely

AI tools can be helpful, but they are not sources of truth.

When using AI systems:

- treat outputs as suggestions, not facts;
- verify important information independently;
- be cautious with health, legal, financial, or political advice;
- remember that AI may sound confident even when it is wrong.

AI is a tool, not an authority.

8. Talking with Family, Friends and Colleagues

Disagreements about information are common.

When discussing sensitive topics:

- listen before responding;
- avoid ridicule or confrontation;
- share reliable sources calmly;
- acknowledge uncertainty where it exists.

Respectful dialogue preserves relationships and reduces polarisation.

9. Your Role in a Healthy Digital Society

By acting thoughtfully, you contribute to:

- reducing the spread of harmful content;
- protecting vulnerable people;
- strengthening trust within communities;
- setting a positive example for younger and older users alike.

Digital responsibility is not about perfection; it is about awareness and care.

10. Final Message

You do not need to be a digital expert to navigate today's information world.

Curiosity, patience, and critical thinking are enough.

Pause before sharing.
Check before trusting.
Ask before reacting.

ANNEX II

POLICY GUIDANCE FOR INFLUENCERS AND DIGITAL CONTENT CREATORS

Addressing Disinformation, Conspiracy Narratives and Online Extremism

1. Purpose of This Annex

This Annex provides sector-specific guidance for influencers, content creators, streamers, and public figures who produce and share content online, particularly on social media and video-sharing platforms.

It complements the project's general policy framework by addressing the unique role and impact of influencers in the contemporary information ecosystem.

This document is not a legal or enforcement instrument. Its purpose is to promote responsible practices, transparency, and awareness, while respecting freedom of expression and creative autonomy.

2. Why Influencers Matter in the Information Ecosystem

Influencers play a distinct role because they:

- reach large audiences through personal trust relationships rather than institutional authority;
- communicate in informal, emotional, and persuasive ways;
- often mix opinions, entertainment, personal experience, and factual claims;
- can rapidly amplify content across platforms.

As a result, influencers can unintentionally contribute to the spread of:

- false or misleading information;
- conspiracy narratives presented as "alternative truths";
- polarising or extremist framings;
- manipulated or decontextualised media.

The impact of influencer content may be significant even when there is no intent to deceive.

3. Scope of Application

This Annex applies to:

- influencers and creators with a significant or recurring audience;
- bloggers, vloggers, streamers, podcasters, and social media personalities;
- creators producing informational, commentary, lifestyle, political, health-related, or news-adjacent content;
- monetised and non-monetised content alike.

The guidance applies across platforms and formats, including posts, stories, videos, livestreams, and podcasts.

4. Key Risks Associated with Influencer Content

Influencers should be particularly aware of the following risks:

- Blurring facts and opinions, making it difficult for audiences to distinguish verified information from personal beliefs;
- Emotional amplification, especially fear, outrage, or distrust;
- Unverified claims, especially in health, science, elections, security, or crises;
- Re-sharing content without context, including old, manipulated, or misleading material;
- Monetisation incentives that reward sensational or polarising narratives;
- Algorithmic amplification, which can increase the reach of problematic content beyond the creator's intent.

5. Core Principles for Responsible Influencing

Influencers are encouraged to adopt the following principles:

5.1 Transparency

- Clearly distinguish facts, opinions, and personal experiences.
- Disclose sponsorships, partnerships, and paid promotions.
- Indicate uncertainty where information is incomplete or evolving.

5.2 Accuracy and Care

- Avoid presenting unverified claims as established facts.
- Exercise extra caution when discussing sensitive topics such as health, elections, public safety, or judicial matters.
- Refrain from oversimplifying complex issues in ways that may mislead.

5.3 Proportionality and Context

- Avoid sensational framing designed primarily to provoke fear, anger, or mistrust.
- Provide context when discussing controversial or emotionally charged issues.
- Avoid language that delegitimises entire institutions or groups without evidence.

5.4 Accountability

- Be willing to correct mistakes openly and promptly.
- Engage responsibly with feedback and criticism.
- Recognise the influence your content may have on audiences.

6. Practical Guidance: Before Posting or Sharing

Before sharing content that includes factual claims, influencers are encouraged to ask:

- Is this information verified by reliable sources?
- Am I sharing this because it is accurate, or because it is shocking or viral?
- Could this content cause real-world harm or misunderstanding?
- Am I clearly signalling what is opinion and what is fact?
- Would I be comfortable explaining and defending this content publicly later?

If there is doubt, delaying or refraining from posting is often the most responsible choice.

7. Handling Corrections and Errors

Mistakes can happen. Responsible behaviour includes:

- acknowledging errors clearly and without defensiveness;
- correcting the information using the same channels and reach as the original content, where possible;
- avoiding silent deletions that leave audiences confused;
- learning from errors to improve future content.

Transparent corrections help build long-term credibility.

8. Interaction with Audiences and Communities

Influencers are encouraged to:

- avoid encouraging harassment, targeting, or hostility;
- discourage followers from attacking individuals or institutions;
- moderate comments responsibly where feasible;
- promote respectful dialogue, even in disagreement.

Influence carries not only visibility, but also norm-setting power.

9. Monetisation and Ethical Considerations

When monetising content, influencers should:

- avoid promoting false or misleading claims for financial gain;
- be particularly cautious with sponsored content related to health, finance, or politics;

- ensure that commercial incentives do not override accuracy or responsibility.

Ethical monetisation supports sustainable and trustworthy creator ecosystems.

10. Cooperation and Reporting

Influencers are encouraged to:

- use platform reporting tools when encountering harmful or deceptive content;
- cooperate with fact-checking initiatives where appropriate;
- participate in media literacy and awareness initiatives.

These actions contribute to a healthier digital environment without limiting creative freedom.

11. Final Message

Being an influencer does not require being an expert in everything.
It does require awareness, care, and responsibility.

Trust is your most valuable asset.
Protecting it protects your audience, and your own credibility.

ANNEX III

POLICY GUIDANCE FOR JOURNALISTS AND BLOGGERS

Reporting on Health, Political and Geopolitical Issues in Crisis and Wartime Contexts

1. Purpose of This Annex

This Annex provides specialised guidance for journalists, bloggers, and independent information producers who report on:

- public health and medical issues;
- political developments;
- geopolitical conflicts, wars, and international crises.

It complements the project's general policy framework by addressing high-risk information environments, where inaccurate, incomplete, or misleading reporting may cause serious real-world harm.

This document does not limit editorial independence or freedom of expression. Its purpose is to promote accuracy, proportionality, and ethical responsibility in particularly sensitive contexts.

2. Why Health and War Reporting Require Special Care

Health crises and armed conflicts share key characteristics:

- high levels of fear, uncertainty, and emotional stress;
- rapidly evolving and incomplete information;
- strategic use of propaganda, psychological operations, and information warfare;
- strong public reliance on media reporting for personal and collective decision-making.

In these contexts, even unintentional errors may:

- undermine public trust in institutions and science;
- influence life-or-death decisions;
- escalate tensions or hatred;
- expose civilians, medical personnel, or journalists to risk.

3. Scope of Application

This Annex applies to:

- professional journalists and editors;
- bloggers and independent reporters;
- digital media outlets and newsletters;
- commentators producing news-adjacent content on health or geopolitical issues;
- freelancers and contributors operating in crisis environments.

It applies across formats, including articles, videos, podcasts, social media posts, live updates, and newsletters.

4. Core Principles for High-Risk Reporting

4.1 Accuracy Over Speed

- Prioritise verification over immediacy.
- Avoid publishing unconfirmed claims simply because competitors have done so.
- Clearly indicate when information is preliminary or evolving.

4.2 Proportionality and Harm Awareness

- Assess potential consequences before publication.
- Avoid unnecessary sensationalism, graphic imagery, or alarmist framing.
- Consider whether publication could expose individuals or groups to danger.

4.3 Independence and Critical Distance

- Maintain critical distance from official sources, governments, armed actors, and advocacy groups.
- Avoid becoming a vector for propaganda, even unintentionally.
- Treat leaks, claims, and exclusive information with heightened scrutiny.

4.4 Transparency

- Clearly distinguish facts, analysis, hypotheses, and opinions.
- Disclose limitations, uncertainties, and conflicts of interest.
- Explain sourcing choices where relevant.

5. Health Reporting: Specific Guidance

When reporting on health, medicine, or public health emergencies:

- rely on authoritative and peer-reviewed sources where available;
- consult qualified medical or scientific experts;
- avoid presenting anecdotal evidence as scientific proof;
- clearly distinguish correlation from causation;
- avoid amplifying unproven treatments, cures, or preventive measures;
- contextualise risks, statistics, and probabilities accurately.

Special caution is required when:

- reporting on vaccines, epidemics, mental health, or mortality;
- covering experimental treatments or early research findings;
- discussing health measures that may affect individual behaviour.

6. Political and Geopolitical Reporting in Wartime

When covering armed conflicts, wars, or international crises:

- clearly identify the source of claims (official, military, intelligence, independent, eyewitness);
- avoid presenting one-sided narratives as established fact;
- verify images, videos, and satellite material to prevent reuse or manipulation;
- be cautious with casualty figures, territorial claims, and operational details;
- avoid publishing information that could compromise civilian safety or humanitarian operations.

Journalists and bloggers should be aware that information itself may be used as a weapon in conflict environments.

7. Visual Content, Images and Videos

For both health and war reporting:

- verify the origin, date, and location of images and videos;
- clearly label archival or illustrative material;
- avoid using shocking visuals solely to attract attention;
- explain when visuals cannot be independently verified.

Manipulated or synthetic media should be explicitly identified and contextualised.

8. Handling Errors, Updates and Corrections

In crisis reporting, errors may occur. Responsible practice includes:

- issuing corrections promptly and clearly;
- updating articles transparently as new information emerges;
- avoiding silent edits that obscure changes;
- explaining why earlier information was incorrect or incomplete.

Corrections are a sign of professionalism, not weakness.

9. Interaction with Audiences

Journalists and bloggers are encouraged to:

- moderate comment sections where feasible to prevent harassment or incitement;
- avoid engaging in inflammatory exchanges;
- correct false claims circulating among audiences when appropriate;
- foster informed and respectful discussion.

Audience trust depends not only on content, but also on tone and engagement.

10. Use of Technology and Automation

When using AI tools, data analytics, or automated systems:

- ensure human editorial oversight at all stages;
- understand limitations and biases of automated tools;
- avoid delegating editorial judgment to algorithms;
- be transparent about the use of automation where relevant.

Technology should support, not replace, editorial responsibility.

11. Training and Professional Development

Media organisations and independent reporters are encouraged to invest in:

- training on disinformation tactics and information warfare;
- digital verification and open-source intelligence techniques;
- trauma-informed reporting;
- ethical standards for crisis journalism.

Continuous learning is essential in rapidly evolving environments.

12. Final Considerations

Reporting on health crises and wars is among the most demanding tasks in journalism.

Accuracy, restraint, and responsibility are not constraints on freedom of expression, they are its foundation.

In high-risk contexts, what is published matters not only for public debate, but for real lives.