BRIEFING

STOA/ESMH Options Brief

Automated tackling of disinformation

Major challenges ahead

The study maps and analyses current and future threats from online misinformation, reports the currently adopted socio-technical and legal approaches to tackling disinformation and discusses the challenges of evaluating their effectiveness and practical adoption. Drawing on and complementing existing literature, the study considers the findings of relevant journalistic and scientific studies and policy reports about detecting, containing and countering online disinformation and propaganda campaigns. It traces recent development and trends and identifies significant new or emerging challenges. It also addresses potential policy implications of current socio-technical solutions for the EU.

Policy options

The study provides options that could be followed and highlights the stakeholders best placed to act upon these at national and European level. The solutions proposed to counter disinformation require very different expertise and rely on the intersection of technological innovations and civil society involvement. Therefore the policy options are similarly organised: some focus on the solutions and tools that could be implemented and others look at how to address the needs of the plethora of stakeholders involved.

Option 1: Enable research and innovation on technological responses

Many of the technological responses described in the report rely on the new capabilities that machine learning and artificial intelligence are poised to bring. A common data framework around disinformation should be established to enable such research and innovation.

- Set up a public repository of online content where researchers and civil society can have access to sponsored advertisement campaign data with a facilitated use of machine-learning systems on these dataset;
- Enable cross-platform research by imposing data standardisation in order to map and analyse a broader disinformation scope.

The development of non-transparent monitoring systems within the large corporations that own and run social media platforms raises questions about their willingness and capacity to tackle disinformation monitoring when acting on their own. Data is essential for training machine-learning and AI algorithms. In a data-driven environment, the main question is not only if the data is collected but whether if it is accessible, how and when. As disinformation techniques are evolving very quickly, researchers need to have uninterrupted access to large amounts of data to evaluate, and re-design methodologies.

Governments and policy makers are thus in a position to help establish this much needed cooperation between social platforms and scientists, to promote the definition of policies for ethical, privacy-preserving research and data analytics, and to ensure the archiving and preservation of social media content of key historical value.

For instance, the collection of all publications and advertisements by political parties and official candidates in a public depository is proposed that would be accessible for researchers and trusted third-parties. Smart use of blockchain technology would also make content falsification almost impossible by proving the origin of content. This technology would be particularly useful for videos or audio content, which can be falsified through deepfake technology. For example, a crowdsourced rating system could enable verified content to prevail over unverified news.

There are emerging technologies for veracity checking and verification of social media content. These include tools developed in several EU funded projects (e.g. PHEME, WeVerify, InVID), tools assisting crowdsourced verification (e.g. CheckDesk, Veri.ly), citizen journalism (e.g. Citizen Desk), and repositories of checked facts/rumours (e.g. FactCheck). However, many of those tools are prototypes resulting from research outcomes and require further improvements. Whether disinformation is genuine, algorithmic or paid, the corresponding detection algorithms need to be further developed, to achieve accuracy comparable to the email spam filter technology.

The outcomes of these research activities should be made open source to enable open science and verifiable algorithms as well as to enable companies and platforms to experiment easily with the new technology. A shared approach is required to address human rights, journalistic and technological challenges caused by this phenomenon. Collaborative research would enable the creation of tools, such as reverse video search, and foster readiness to face upcoming risks and scenarios.

Option 2: Improve the legal framework for transparency and accountability of platforms and political actors for content shared online

In order to build a coherent legal framework and avoid what could be considered as a fragmentation of the digital space, with a different set of rules applicable at the national and regional level, a coherent global framework of regulation should be put in place. When it comes to content regulation, many advocate for holding social media companies accountable for moderating the information shared on their online platforms. Moderation, however, should not compromise freedom of speech, privacy, transparency and accountability.

- Provide information and social platforms with a new legal status of 'information fiduciary' that would grant such companies specific role and obligations towards their users as facilitators of free-speech and collectors of personal data;
- Foster user-centric regulation with a strong moderation process including an independent appeal procedure and independent control of measures implemented by platforms and political actors.

To fully establish their liability, some advocate for social platforms to be considered publishers and be held accountable for the content shared on their online apps. As they operate more as communication channels than as media agencies, such qualification can seem inadequate, with a risk of giving technology companies the possibility to limit freedom of speech.

Digital media and social media companies could become "information fiduciaries" - a new legal category - granting such companies specific role and obligations as facilitators of free-speech and collectors of personal data. The idea (similar to the obligations of doctors and lawyers) is to make

user care and confidentiality the primary responsibility of social platforms, with loyalty towards end users not being compromised for business reasons.

There is also an argument that fiduciary obligations exist irrespective of the contractual information in the platforms' terms of service. As the European General Data Protection Regulation (GDPR) framework sets strong privacy protection, it could be investigated how such obligations could also be applied to information shared online.

This legal status and regulation on the minimum set of obligations on dataset access for researchers on disinformation and the public data repository put forward in option 1 would be more effective if applied globally. A unified European framework supported by a coherent input from Member States in international fora such as the OECD and the G7 certainly should be prioritised.

Finally, a binding regulatory framework should set the necessary safeguards to the risks posed by tech-only solutions, in particular regarding abusive moderation. By putting the user back at the centre of the moderation process, companies should provide citizens and regulators with transparent lists of the content removed and suspended accounts, as well as notice and justifications for moderation. Ultimately, the user should be able to appeal the moderation decision.

Option 3: Strengthen media and improve journalism and political campaigning standards

In order to safeguard public trust in media, strong press standards should be defined and followed, supplemented by transparency on political advertising and political campaign standards. National fact-checking initiatives should be promoted, as a collaboration between different media organisations, journalists and independent fact-checkers.

- Promote strong press standards and support media literacy activities;
- Promote fact-checking efforts with the perspective of having at least one independent fact-checking network per member state;
- Encourage collaborative projects between journalists, media and researchers focused on content verification and fact checking, for instance, fact-checking efforts could rely on automated methods for checking against shared database of already debunked misinformation.

Many major media are already carrying out key fact-checking and media literacy activities. However, these should be widened and adopted by all media. Therefore, media should commit to respect high ethical standards, for example refraining from the use of clickbait headlines on social media. In addition, as disinformation can sometimes spread locally in a regional dialect, local reporting should also be supported with government subsidies.

To become an efficient source of reliable information for citizens, fact-checking could rely on automated methods for checking against statistical sources or a shared database of already debunked misinformation. Hence, the outcomes of new research (as described in Option 1) need to be made understandable by non-specialists and disseminated to journalists and public organisations, in order to inform and strengthen their ability to detect and debunk disinformation. There is also ample scope for collaborative projects between journalists, media and researchers focused on content verification and fact-checking.

Option 4: Support of interdisciplinary approaches and localised involvement from civil society

Multiple stakeholders with different backgrounds are working on the digital misinformation and the disinformation ecosystem. Taken in isolation, these different approaches can address particular aspects of the issue, but, as also argued by the EU High Level Expert Group (HLEG report, 2018), its complex, multi-dimensional nature can only be tackled fully through a multi-stakeholder approach.

- Structure the involvement of civil society around the creation of independent consultative bodies such as CNNUM (Conseil National du Numérique, 'Digital National Council') in France;
- Invest in civil society resilience actions similar to those created by the MSB (Civil contingencies agency) in Sweden in order to build a stronger awareness of the population;
- Prioritise investment in multidisciplinary research teams that could form research action networks.

Analysing the impact of social media on society and democracy involves a variety of actors from different backgrounds, including data scientists, artificial intelligence researchers, political and social scientists as well as journalists. They need to work together alongside social media platforms and policy makers to gain a full understanding of the mechanisms behind viral disinformation and the most effective ways to contain and prevent it.

Journalists and scientists need access to public social media posts for research and experimentation purposes. In order to prevent potential bias and conflict of interest, exclusive partnerships between social platforms and designated science labs need to be avoided. Instead, data and collaborations need to be made available to all stakeholders.

Thus, the study puts forward the option to support, both at European and national levels, the creation of independent bodies that facilitate collaborative, evidence-based research and help promote best practice in the detection and prevention of online disinformation. These bodies could also act as initiative incubators. They could be designed as independent consultative bodies, similar to the Digital National Council in France, combined with national or European research institutes that bring together scientists from multiple organisations and disciplines.

This document is based on a study on 'Automated tackling of disinformation' (PE 624.279) financed under the European Science-Media Hub budget and published in March 2019. The study was carried out by EU DisinfoLab in response to a request from the Panel for the Future or Science and Technology (STOA) and managed by the Scientific Foresight Unit within the Directorate-General for Parliamentary Research Services (DG EPRS) of the European Parliament. Authors: A. Alaphilippe, A. Gizikis, C. Hanot of EU DisinfoLab and K. Bontcheva of The University of Sheffield. STOA administrator responsible: Mlhalis Kritikos.

DISCLAIMER AND COPYRIGHT

esmh@ep.europa.eu (contact)

http://www.europarl.europa.eu/stoa/ (STOA website)

https://sciencemediahub.eu/ (ESMH website)

http://epthinktank.eu (EPRS website)