FERMI

A White Paper by the FERMI Project

BEHIND THE CURTAIN: CHALLENGES AND GAPS IN THE STRATEGIC ANALYSIS AND COMBAT OF DISINFORMATION CAMPAIGNS

Prepared by the FERMI Consortium &
edited by Netcompany SA

Abstract

This document captures critical insights from the EU-funded **FERMI** project, which focuses on the analysis of disinformation dynamics in the context of crime. It outlines key challenges, lessons learned and identifies gaps across technical developments and legal and ethics requirements, eventually laying out a series of recommendations.

This White Paper serves as a roadmap for policymakers, industry stakeholders, standardisation bodies and technology experts, fostering a more secure and intelligent approach to combating disinformation in general and disinformation-induced crime in particular.

Authors

Spyros Evangelatos (INTRA)
Nikos Dimakopoulos (ITML)
Sven-Eric Fikenscher (BPA) – Coordinator
Joaquín García (ATOS)
Flavia Giglio (KUL)
Michael Lo Giudice (UCSC)
Jenita Rauta (PUCF)
Miia Sainio (PUCF)
Alexia Solomou (IANUS)
Giorgos Stamatis (ITML)
Tim Stuchtey (BIGS)

Foreword: Strategic relevance to EU security and law-enforcement communities

Anecdote suggests that the spread of falsehoods rooted in politically extremist beliefs can easily lead to criminal incidents. The infamous attempt of a group of extremist supporters of US President Donald Trump to storm the US Capitol on 6 January 2021 amidst claims about the presidential election in 2020 allegedly being stolen from their candidate is a case in point. At least seven people died in that effort.[1] Similar events could be witnessed in Brazil in early 2023, where Jair Bolsonaro lost the election to Lula da Silva but questioned the reliability of the most used voting machines and even contested the outcome of the election on that basis.[2]

Anecdotal evidence is soundly supported by empirical studies. An analysis of the National Consortium for the Study of Terrorism and Responses to Terrorism (START) concluded that the number of disinformation- and conspiracy theory-driven terrorist attacks has greatly increased. More specifically, there were just six such attacks "in 2019, versus at least 116 in 2020,"[3] mostly related to false claims about the Covid pandemic. Europol's most recent annual report on terrorism even observes that "[m]ost of the arrested suspects had been consuming and sharing online content involving disinformation, conspiracy theories, antiestablishment and anti-government sentiments […]. These narratives all blend together and foster their path into extremism and eventually to the concrete perpetration of acts of violence."[4] Yet, the nexus between disinformation and politically motivated violence is still gravely under-examined, as the bulk of disinformation research and counter-measures has focused on foreign information manipulation and interference. In the framework of Horizon Europe, the EU's research programme, the FERMI and VIGILANT projects that do address

---

[1] Cameron, C. "These Are the People Who Died in Connection With the Capitol Riot." *The New York Times*. 5 January 2022. https://www.nytimes.com/2022/01/05/us/politics/jan-6-capitol-deaths.html.

[2] Associated Press. "Defeated president contests Brazil election." *POLITICO*. 22 November 2022. https://www.politico.com/news/2022/11/22/brazil-election-contested-bolsonaro-00070589.

[3] Farrell, L. "UMD Report: Conspiracy theories fueled more terror attacks in 2020." *National Consortium for the Study of Terrorism and Responses to Terrorism*. 7 July 2022. https://www.start.umd.edu/news/umd-report-conspiracy-theories-fueled-more-terror-attacks-2020.

[4] Europol. *European Union Terrorism Situation and Trend Report*. Luxembourg: Publications Office of the European Union. 2025, 58.

the impact of false allegations on the crime landscape are the rare exceptions.[5] This White Paper aims to further research and guide policy on the subject matter by discussing the most profound challenges encountered during the FERMI project and drawing policy-relevant conclusions from those.

[5] European Commission, Directorate-General for Migration and Home Affairs. *Research Projects Help Combat Disinformation Ahead of Elections.* 30 May 2024. https://home-affairs.ec.europa.eu/news/research-projects-help-combat-disinformation-ahead-elections-2024-05-30_en.

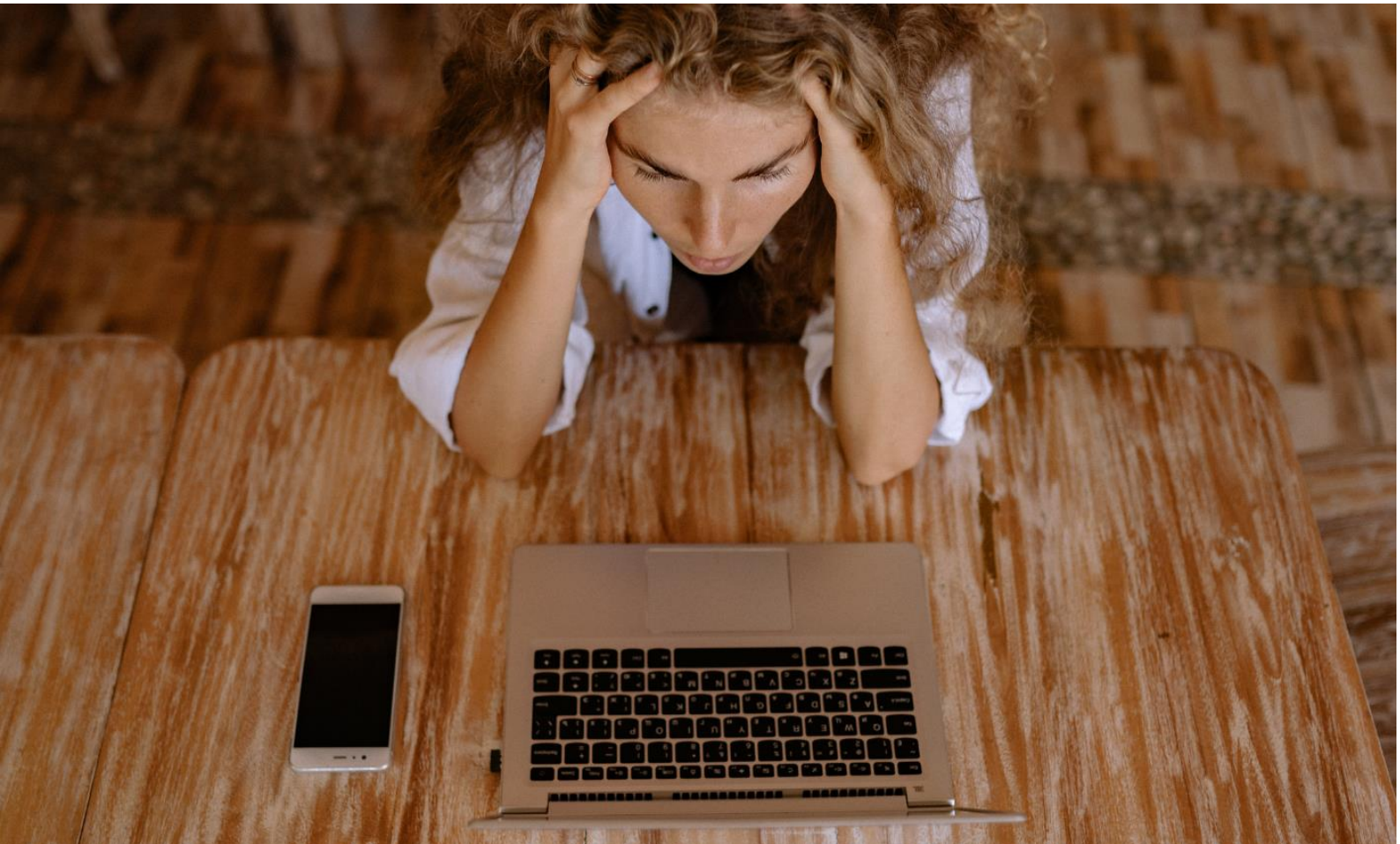
1 Executive summary: Key findings and takeaways

This White Paper presents the FERMI project's key insights into how the study and mitigation of the nexus between disinformation and crime. More specifically, in its research leading to the development of a platform that furthers investigations of social media posts, threat assessments and the study of counter-measures, the FERMI consortium has made the observations and drawn the conclusions as follows.

- Disinformation concerning different forms of political extremism have been found to contribute to violence, but this nexus remains gravely under-examined.
- Based on EU policy documents, disinformation can be defined as claims that are informed by 1) a factual or misleading nature; 2) the intention to obtain economic gain or deceive the public; 3) and lead to public harm. Such a universally accepted definition of disinformation ought to be formally and officially embraced.
- Research greatly depends on access to social media, crime and disinformation datasets that are hard to procure and hugely discrepant. Accordingly,
    - the Digital Services Act (DSA) should be revised to allow private sector companies to request access to social media data, so long as their activities are limited to publicly funded research. Alternatively, researchers should be allowed to pass such data on to private sector companies in the context of joint research,
    - a European crime framework, similar to the American Universal Crime Reporting Program should be established and
    - disinformation datasets should be made available by social media platform providers in a comprehensive and coherent form.
- Social media data, even if available, also vary greatly across platforms, which presents profound technical challenges. The development and availability of standardised high-quality multilingual social media datasets that include coherent information such as timestamps, engagement metrics, identifiers linking content across platforms etc. should be supported.
- The need to train technical tools with sensitive data that are not supposed to be shared can be mitigated by expanding the development and use of federate and swarm learning, keeping users from giving others access to

their datasets whilst fine-tuning the tools' accuracy with their help nonetheless.

- Interoperability has turned out a technical challenge. Tools tailored to law-enforcement agencies (LEA) need to be particularly conscious of interoperability at the technical layer as well as trustworthiness and usability at the human layer.
- Human involvement is also crucial from a legal and ethics standpoint, which requires a strict human-in-the-loop approach and far-reaching risk monitoring in accordance with the AI Act. Ethics and data protection concerns can be mitigated by pre-processing research data, as FERMI did with respect to social media data, in the sense of ensuring anonymisation.
- A lack of digital skills has turned out to increase the public's vulnerability to disinformation. Accordingly, societal resilience can be enhanced through training and awareness-raising initiatives bringing together all crucial stakeholders.



2 Brief introduction of the FERMI project

2.1 Overview of the FERMI project

The FERMI project looks into ***the nexus between disinformation*** (basically defined as factual or misleading claims spread with the intention to obtain economic gain or deceive the public in a way that causes public harm, see section 3.1) ***and politically motivated crime*** in a series of three case studies chosen in view of the relevance of the to-be-examined field of politically motivated violent crime (right-wing, left-wing and Covid-/health-related) and the matching European country. Moreover, the FERMI consortium develops technical and practical solutions for law-enforcement agencies in their fight against disinformation-induced crime. Beyond its technical innovations, FERMI addresses the critical human dimension of counter-disinformation through awareness raising and training activities. These activities focus on increasing understanding of disinformation dynamics and their societal impact, while building societal resilience and digital trust.

2.2 The FERMI use cases

The FERMI project's three use cases concern ***Russian disinformation efforts to stir up tensions in Finland*** by appealing to right-wing extremists, ***disinformation surrounding the Covid virus and EU relief efforts in Belgium***, and ***disinformation about the treatment of a German left-wing extremist*** in a Hungarian jail.

- Russia is sending refugees/asylum seekers/migrants towards the Finnish border and facilitates narratives about their misuse of Scandinavian hospitality, especially in the form of social services being provided.[6] Inflated reports about refugee/asylum seeker/migrant flows (of African and Middle Eastern origins) towards Finland are likely to remind the Fins of the 2015 refugee crisis

[6] Quite tellingly, the EU Disinfo Lab's analysis of the situation in Finland singles out the spread of right-wing falsehoods, especially in an online context. More specifically, the report points out that there are signs "of increasing radicalisation fuelled by far-right communities and networks online." See Moilanen, P., M. Hautala and D. Saari. *Disinformation Landscape in Finland*. Brussels: EU Disinfo Lab. 2023, 3.

where tensions resulting from record immigration led to multiple firebomb attacks on accommodation centres.[7]

- A little-known form of extremism distinct from long-standing radical ideologies, namely health-related extremism, greatly increased amidst the Covid pandemic, as parts of the public were looking for explanations in line with their strict rejection of their governments' Covid policy.[8] As explained above, such allegations caused a rapid increase in non-attributable terrorist attacks.[9] Belgium was a case in point. The Disinfo Lab's country report observed that "[t]he main events were the organisation of demonstrations in Brussels against the health pass and restrictions, followed by violence, and coordinated online through social media. As seen in other countries, this narrative is now very close to technology scepticism, for instance, over the allegedly damaging effects of 5G on health."[10]
- A key contributor to violent left-wing crime in Germany is the Antifa movements.[11] The Antifa East/hammer gang has attacked numerous

---

[7] Often-times such attacks were preceded by anti-refugee sentiments being stirred up "with falsified statistics of immigrants' crimes or claims of specific events witnessed by friends and colleagues, such as incidents of rape or child abduction by refugees." See Koehler, D. "Right-Wing Extremism and Terrorism in Europe. Current Developments and Issues for the Future." *Prism: The Journal of Complex Operations* 6, no. 2. 2016. https://cco.ndu.edu/PRISM/PRISM-Volume-6-no-2/Article/839011/right-wing-extremism-and-terrorism-in-europe-current-developments-and-issues-fo/.

[8] Such conspiracy theories gained traction amidst the shutdowns and further restrictions that were imposed across Europe to stem the tide of Covid's spread. See Lynas, M. "COVID: Top 10 Current Conspiracy Theories." *Alliance for Science*. 20 April 2020. https://allianceforscience.org/blog/2020/04/covid-top-10-current-conspiracy-theories/. The Disinfo Lab's report on Belgium explains that Belgium was sort of a poster child for this trend. In particular, conspiracies centered around Covid-19 tapped into long-standing scepticism over vaccination in general and could benefit from the opposition to Belgium's lockdowns even further. Throughout these proceedings disinformation activities, including on social media, played a prominent role in undermining Belgium's Covid policies. Specifically, the report emphasises that "[t]hough already present before 2020, a growing presence of online communities pushing anti-vax narratives has been seen in Belgium during and after the COVID-19 pandemic. Starting with opposition to mandatory vaccination, the movement rapidly grew." See Alaphilippe, A. *Disinformation Landscape in Belgium*. Brussels: EU Disinfo Lab. 2023, 5.

[9] Farrell, "UMD Report."

[10] The report also notes that "[a]s seen in other countries, this narrative is now very close to technology scepticism, for instance, over the allegedly damaging effects of 5G on health." See Alaphilippe. *Disinformation Landscape*, 5.

[11] The threat of left-wing extremism seems to be lesser in scope across Europe, but it should not be underestimated at all. In Germany, the support of left-wing militancy and the potential for violent left-wing extremism has even grown. Germany's domestic intelligence agency, the Bundesamt für Verfassungsschutz (BfV) reported a rise in the number of "violence-prone" left-

individuals they considered Nazis. Key assaults took place in Budapest in February 2023. Antifa groups and other left-wing extremists have spread disinformation about Maja T's., a leading figure in the hammer gang, trial and imprisonment in Hungary, especially under the hashtag #FreeMaja.[12]

2.3 The FERMI platform

The FERMI platform includes an ***investigation-advancing tool, threat assessment solutions, and impact assessment tools***.

- Investigations are advanced through the development of a Spread Analyser that captures how illicit disinformation posts travel on social media (making it easier for LEAs to detect accounts that were involved in sharing said posts).
- Threat assessments are supported by analysing the posts' underlying sentiments and estimating the evolving crime landscape in the aftermath of a disinformation campaign.
- The broader impact of such campaigns is analysed by estimating the cost of disinformation-induced crime. Based on the scope thereof, counter-measures may be proposed, which LEA end-users are free to consider.

---

wing extremists from around 9600 to 11200 between 2020 and 2024 (which is still a somewhat smaller number than the number of violence-prone right-wing extremists, but the trend is surely worrisome). The BfV also observed a "high level of radicalisation" among violence-prone left-wing extremists. See Bundesministerium des Innern und für Heimat: Bundesamt für Verfassungsschutz. *Verfassungsschutzbericht 2024*. Berlin. 2025, 142. Moreover, in 2024 left-wing motivated crimes went up by 37.9 per cent. Property damage incidents increased in particular. On the upside, incidence of assault decreased. Interestingly, assaults mostly targeted alleged Nazis (as opposed to LEAs, as in the past). Usually, such acts of violence are carried out by small groups in a planned and targeted manner with Antifa movements being the dominant players amongst the attackers. See Bundesministerium des Innern. *Verfassungsschutzbericht*, 142-143.

[12] Ibidem, 146-147.

Moreover, a Swarm Learning framework enables LEA end-users to train global machine learning models whilst ensuring data-privacy[13] by avoiding the need to have a central server agent through which all data flows are going.

[13] No agent should be able to make any inference about the data of any other, except for the output produced through the aggregation of all the provided data.

3 Placing disinformation on the map of counter-extremism and -terrorism

3.1 Delineating disinformation

In 2018, the European Commission tasked a High-Level Expert Group (HLEG) to provide policy suggestions on how to tackle disinformation. The result was the report "A multi-dimensional approach to disinformation: report of the independent high-level group on fake news and online disinformation", published in the same year.[14] The report defines disinformation as ***"false, inaccurate or misleading information designed, presented or promoted to intentionally cause public harm or for profit."***[15]

After the publication of the report, the European Commission adopted the communication "Tackling Online Disinformation: a European approach", the first policy document addressing the phenomenon at the EU level. The Commission adopted a definition of disinformation very similar to that of the HLEG. According to the communication, disinformation is any ***"verifiable false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and may cause public harm"***.[16]

Another contribution to the challenge of specifying a definition of disinformation came from the report "Information disorder: Toward an interdisciplinary framework for research and policy making", commissioned in 2017 by the Council of Europe. In the report, the authors Wardle and Derakhsan defined disinformation as any ***"information that is false and deliberately created to harm a person, social group, organization or country"***.[17]

While there are some differences among the definitions, they present ***three core elements: 1) a factual or misleading nature of the information; 2) the intention of the actors to obtain economic gain or deceive the public; 3) public harm.***

---

[14] European Commission. Directorate-General for Communication Networks, Content and Technology. *A Multi-Dimensional Approach to Disinformation: Report of the Independent High Level Group on Fake News and Online Disinformation.* Luxembourg: Publications Office of the European Union. 2018.

[15] *Ibidem*.

[16] European Commission. "Tackling Online Disinformation: A European Approach." *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, COM/2018/236.* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0236.

[17] Wardle, C., and H. Derakhshan. *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making*. Strasbourg: Council of Europe. 2017.

3.2 The often-overlooked role of disinformation in radicalisation and terrorism

As explained above, the nexus between disinformation and politically motivated violence is still ***gravely under-examined***. In Horizon Europe's Cluster 3 (Civil Security for Society) research on the subject matter basically boils down to the FERMI and VIGILANT projects,[18] whereas there are numerous Cluster 2 projects (Culture, Creativity and Inclusive Society) on how to safeguard democracy from the ramifications of disinformation. This is in line with the political trend to largely (or even exclusively) focus anti-disinformation activities on mitigating foreign information manipulation and interference.

This is all the more surprising in consideration of the fact that extremist belief systems can make individuals and groups particularly susceptible to buying into the logic of false allegations, if those corroborate long-standing beliefs. More specifically, numerous ***extremist beliefs are rooted in founding myths*** such as anti-Semitic resentments advocating conspiracy theories of Jewish world domination, often-times even citing fake documents such as the infamous Protocols of the Elders of Zion.[19] The available research on the subject, as preliminary as it is, has revealed that political extremists are particularly prone to "the belief that out-groups are engaged in secret actions to control in-group outcomes."[20] It has even been argued that violent extremists' "greatest tool is the mis-, dis-, and mal-information […] that feeds extremist movements and ideologies."[21]

3.3 The key narratives of extremist disinformation campaigns

A recent study of the Radicalisation Awareness Network (RAN) that was authored by Francesco Farinelli on behalf of the European Commission identifies anti-immigrant conspiracy theories, anti-Semitic conspiracy theories, anti-establishment and anti-elite conspiracy theories and conspiracy theories in the COVID-19 context as the most profound popular beliefs that are promoted by political extremists relying on false allegations.[22]

---

[18] European Commission. Directorate-General for Migration and Home Affairs. *Research Projects.*

[19] Farinelli, F. *Conspiracy Theories and Right-Wing Extremism – Insights and Recommendations for P/CVE. Radicalisation Awareness Network (RAN)*. Luxembourg: Publications Office of the European Union.
2021, 5.

[20] Berger, J. M., *Extremism*. Cambridge, MA: The MIT Press Essential Knowledge Series. 2018, 66.

[21] Mines, A. "The Fractured Threat Landscape." *Police Chief Magazine*. 2022, 36. https://www.policechiefmagazine.org/fractured-threat-landscape/.

[22] Farinelli. *Conspiracy Theories and Right-Wing Extremism*.

Further research has found out that certain ***right-wing "[p]arties*** such as the National Democratic Party of Germany[23] and The Third Way ***have been involved in organizing protest groups online (typically via Facebook) and stirring up anti-refugee sentiments*** with falsified statistics of immigrants' crimes or claims of specific events witnessed by friends and colleagues, such as incidents of rape or child abduction by refugees."[24]

The importance of violent right-wing extremism notwithstanding, anti-establishment propaganda is also at the heart of violent left-wing extremism. This is corroborated by further in-depth research that has identified "a link between [left- and right-wing] political extremism and a general susceptibility to conspiracy beliefs. Although the ***extreme left may sometimes endorse different conspiracy theories*** (e.g. about capitalism) than the extreme right (e.g. about science or immigration), both extremes share a conspiratorial mindset, as reflected in a deep-rooted distrust of societal leaders, institutions, and other groups, allied with a corresponding tendency to explain unexpected, important events through conspiracy theories."[25]

The spread of ***health-related disinformation*** by political extremists ***is a more recent trend that was hugely shaped by Covid-related disinformation***.[26] There are countless false allegations about Covid, the most crucial ones are centred around the role of "5G and other wireless technologies – which range from causing cancer and killing animals and plants to causing the coronavirus outbreak."[27]

3.4 From words to action: How extremist disinformation campaign can lead to violence

The above-mentioned disinformation narratives might easily translate into organised violence, as right-wing groups "like The Third Way have also published guidebooks on how to organize large-scale protests, and have officially

---

[23] The aforementioned developments in Germany will also be taken into consideration in the sense that BPA, albeit not an LEA partner in a narrow sense, will provide a further data set on Bavaria, Germany to ensure that such proceedings can be examined in-depth.

[24] Koehler. "Right-Wing Extremism and Terrorism."

[25] van Prooijen, J.-W. "Voters on the Extreme Left and Right Are Far More Likely to Believe in Conspiracy Theories." *EUROPP – European Politics and Policy at LSE Blog*. 2 March 2015. http://bit.ly/1zS8hW3.

[26] Lynas. "COVID: Top 10 Current Conspiracy Theories."

[27] Farrell. "UMD Report."

registered ***demonstrations that***, in the majority of cases, ***devolved into violent action*** or took place shortly before arson attacks."[28]

On the extreme left, some groups such as the German-based "Organisierte Autonomie" ("Organised Autonomy") and the "Antifaschistischer Aufbau Muenchen" ("Anti-fascist Base Munich") have placed the blame for the war in Ukraine squarely on NATO's shoulders,[29] which implies a certain susceptibility to Russian disinformation. In ensuing Antifa gatherings like ***during the annual Labour Day marches violent activities did unfold***.[30] The above-mentioned myths surrounding the arrest of a leading member of the Antifa hammer gang is another case in point.

The evidence of the nexus between numerous Covid-related disinformation campaigns and violence is particularly overwhelming. Interestingly, "the first year of the COVID-19 pandemic did not dramatically alter the number of terrorist attacks around the world […] but individual conspiracy theory extremists were involved in ***an increasing number of incidents*** […] [namely] six in 2019, versus at least 116 in 2020, in countries ranging from Australia and New Zealand to the United States, Canada, United Kingdom and Germany. Nearly all were non-lethal, and a surprising 96% were aimed at damaging telecom targets […]," [31] which is the obvious result of the 5G myth mentioned above.

Accordingly, the development of the end-user tools FERMI has produced is more urgent than ever. However, said development comes with numerous technical, as well as legal and ethics challenges that are laid out as follows, including the FERMI consortium's efforts to overcome those and the lessons learned in that effort.

---

[28] Koehler. "Right-Wing Extremism and Terrorism."

[29] Bayerisches Staatsministerium des Innern, für Sport und Integration. *Verfassungsschutzbericht 2023*. Munich. 2024, 265–66.

[30] See, for example, Verfassungsschutz Baden-Württemberg. *Linksextremistische Ausschreitungen bei "Revolutionärer 1. Mai Demonstration" in Stuttgart*. 17 May 2024. https://www.verfassungsschutz-bw.de/,Lde/Startseite/Meldungen+und+Archiv/Ausschreitungen+am+1_+Mai+in+Stuttgart.

[31] Farrell. "UMD Report."

4 Technical Challenges and Lessons Learned

4.1 Data collection and filtering at scale

Throughout the implementation of FERMI, several technical challenges emerged that are critical to understanding the limitations and opportunities in the use of AI for disinformation analysis. These challenges span the entire data pipeline, from the collection and filtering of heterogeneous datasets to the integration of analytical outputs into operational intelligence workflows. Key issues include dealing with the volume and variability of social media and crime-related data, handling multilingual and context-specific narratives, ensuring consistency across time and platforms and maintaining interoperability within secure environments. This section presents a structured analysis of these challenges, along with practical lessons learned from system deployment and real-world testing with security stakeholders.

4.1.1 Social media data

A major technical barrier encountered in FERMI was the ***limited availability of high-quality, structured social media data*** suitable for analysing the evolution and mitigation of disinformation campaigns.[32] Effective analysis requires not just isolated messages or posts but network-level snapshots that capture user interactions, propagation dynamics and the evolution of (false) narratives over time along with the evaluation of the effectiveness of mitigation measures such as the deployment of fact-checkers,[33] content removal, account suspensions, coordinated counter narratives, etc. This level of granularity is critical for understanding how disinformation spreads, how it is amplified or countered, and which actors are central to the campaign. However, obtaining such complete datasets proved exceptionally difficult due to platform restrictions and the lack of publicly available, time-sensitive network data.

This problem was particularly acute in the case of X (formerly Twitter), which has historically been a key platform for the spread of both legitimate and malicious narratives. In recent years, X has shifted toward a closed API model, significantly restricting access for research and public-interest analysis. Where previously

[32] Evangelatos, S., E. Veroni, V. Efthymiou and C. Nikolopoulos. "Modeling Disinformation Spread in Social Networks: Phase Transitions and Mean-Field Analysis." *ACM Transactions on the Web* 19, no.4. 2025, 1-24.

[33] Evangelatos, S., M. Konidi, E. Veroni, S. Karagiorgou and C. Nikolopoulos. "A Perturbation-Theoretic Model for Fact-Checker Deployment in Dynamic Disinformation Networks." Accepted for publication in *Companion Publication of the 17th ACM Web Science Conference (WebSci)*. Dublin: ACM. 2025.

researchers could retrieve structured data including follower networks, retweets and timestamps, ***current X API offerings are either prohibitively expensive or technically limited***. This undermines the ability of research projects to develop and validate models that require longitudinal datasets to detect disinformation campaigns and track the effectiveness of mitigation measures across different phases.

While the Digital Services Act introduces new transparency requirements for very large online platforms operating in the EU, including obligations to provide access to vetted researchers,[34] the practical implementation of these provisions remains fragmented. Notably, the current framework explicitly restricts access to academic or non-for-profit institutions, effectively ***excluding for-profit organisations even when they are conducting research in the public interest,*** including in EU-funded projects. This limitation significantly undermines the ability of multidisciplinary consortia, like FERMI, to fully leverage the expertise and technological capabilities of private-sector partners involved in disinformation analysis and counterterrorism intelligence.

Additionally, social media data were also required for the training of the models which analysed them. Training data is a critical component in the development of machine learning models, particularly for supervised approaches such as sentiment analysis. Since the goal of sentiment analysis was to classify social media posts as negative, neutral, or positive, the training data needed to consist of posts paired with sentiment labels. This ensured that the model could learn from representative examples and generalize effectively to unseen data.

Obtaining high-quality training data, however, is a persistent challenge. Two main strategies are commonly used: leveraging existing open-source datasets, or collecting and manually annotating new data. The latter, while offering greater control over data quality and domain relevance, is highly resource-intensive and falls outside the primary scope of the project. Consequently, the project's approach relied on identifying and reusing open-source datasets that were both relevant to social media sentiment analysis and sufficiently large to support robust training.

Within FERMI, a comprehensive review of publicly available resources was conducted to maximize data quality while ensuring sufficient volume. Most labelled datasets originated from X, given its wide adoption in sentiment analysis research. Nonetheless, datasets from other platforms such as Reddit were also integrated, which proved valuable for evaluating the model across different

[34] For the specifics of the DSA, see Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act).

social media contexts. This combination allowed us to balance quality, diversity, and scalability in the data collection process

4.1.2 Crime data

Relative to other jurisdictions, the European public security landscape could be described as accessible – in terms of research data. Indeed, a great deal of statistical information is accessible from public entities at a range of geographical granularities. While this degree of accessibility is worth praising, the data demands of innovative statistical methods, which are experiencing a shift towards AI-powered approaches, goes beyond what is currently available.

In the development of FERMI's technological offerings, the need for incident level crime data emerged, wherein criminal events (the commission of crimes) are reported at a singular level. That is, where one row is one crime. Widely accessible in the United States, since implementation of the Universal Crime Reporting Program in 1930, American police jurisdictions provide incident-level crime often with the day and hour in which the incident occurred. Within the framework of FERMI, LEA partners provided incident-level data, formatted with privacy- and security-by-design, ensuring the AI-driven innovations of FERMI were achieved whilst acting as a proof-of-concept for the potential production and release of crime-data at an incident-level.

The importance of incident-level crime data to research will certainly grow in the years to come, as AI-driven methods' capacity to provide the extraordinary insights they have become known for is premised on them being provided a granularity of data not previously foreseen or required. As researchers and LEAs expand in their use of predictive methods to understand and react to the

time and space of crime, the *when and where*, cooperation between the two will be largely dependent on the sharing of data. A European framework, similar to that of the American Universal Crime Reporting Program but with the full integration of the General Data Protection Regulation's (GDPR) and the Findability, Accessibility, Interoperability, and Reuse of digital assets' principals as an objective, is a natural next step.

A truly effective European strategy for harmonizing crime data should also include the distinct typologies of politically-motivated crime — left-wing, right-wing, religious, and foreign-influenced forms. Standardised politically motivated crime would be accessible to law enforcement agencies for effective prosecution, but also to the academic community for rigorous analysis and policy evaluation. Sharing such data amplifies its value, as knowledge grows when it is disseminated rather than hoarded.

4.1.3 Disinformation data

Differently from crime data, disinformation data, in both the European and extra-European jurisdictions, is characterised by small batch sizes and the divergent features between dataset makes aggregation or the utilization of several contemporarily difficult and even in certain instances, as was the case in FERMI, impossible. With academia and academic researchers being the primary users and producers of disinformation datasets, the data is biased, understandably so, towards the needs of academics.

Hence, disinformation data is overwhelmingly English and American-centric, with the former the international language of academic journals and the latter being home to the majority of prominent research institutions (especially in the field of political science). While originating from past research endeavours needs, this American focus also potentially canalizes future research, especially with the aforementioned movement of social media platform's APIs towards pay-per-access models. Likewise, the researchers who compiled datasets, with notable exception where the objective was the production of a dataset, aimed to develop said datasets with the features (i.e., the variables within the dataset, that describe elements of the disinformation) as needed for their experimentation (i.e., the concept they were studying).

As such, there is a great deal of variance between datasets, in terms of features and also in terms of inclusion rules; what content was considered to be disinformation. This presents a challenge when research, as in the case of FERMI, requires large datasets with representative sample of the universe of disinformation. Given FERMI's need to study the relation between offline crime and online disinformation, a sample of disinformation was required that covered the

same time-span as the crime sampled and was at least representative of whole body of disinformation online.

Aggregating small batch datasets, compiled with different inclusion rules – to the point of often having differing definitions of disinformation – and different variables - with some omitting vital information such as date and time – is not possible while maintaining representativeness and balance. Thus, a clear need exists for comprehensive disinformation datasets, though given the present direction of social media platforms, with both Twitter and Meta closing public access to their APIs, this need will become increasingly difficult to solve.

4.2 Language, context, and narrative complexity

The quality and representativeness of training data strongly influence model performance. The closer the training dataset aligns with the intended application, the more reliable the model's predictions are likely to be. However, this creates a limitation when applying models across different social media platforms and domains, where language use, style, and context can vary significantly.

Social media channels introduce particular challenges for natural language processing. Posts are typically brief and constrained by platform-specific formats, encouraging ***the use of slang, acronyms, and evolving forms of expression***. Each platform develops its own linguistic conventions, resulting in diverse and rapidly changing vocabularies that complicate text analysis. These characteristics underscore the inherent difficulty of building a generalised sentiment analysis model that performs consistently well across all platforms and domains.

Another key challenge lies in balancing general versus domain-specific training data. While domain-general datasets offer scale, they often include high variability and noise, leading to reduced precision. In contrast, domain-specific datasets enable models to specialise and achieve higher accuracy in targeted contexts but risk limiting flexibility if they are overly narrow. Language coverage adds a further layer of complexity, as ***high-quality multilingual resources remain scarce in domains such as health and politics***.

Within FERMI, these challenges were addressed through a balanced data strategy. The research team curated and evaluated datasets that reflected the project's use cases while avoiding over-specialisation. To improve platform robustness, training data included posts from sources beyond X, enabling evaluation across multiple social media contexts. Additionally, integrated translation mechanisms ensured that sentiment analysis could extend to non-English content, supporting broader applicability and inclusiveness.

4.3 Temporal and cross-platform inconsistencies

One of the central technical challenges in the analysis of disinformation campaigns is the lack of temporal coherence across datasets, particularly when data is collected from different sources or social media platforms. Disinformation narratives often evolve rapidly, exploiting emerging events or societal tensions, and are frequently reshaped as they move across time and audiences.[35] Without time-stamped, synchronised datasets, it becomes difficult to reconstruct the sequence of events, identify the original sources of a campaign, or understand how specific messages were amplified. In the context of FERMI, this limited the ability to generate a continuous and reliable timeline of campaign activity, which is crucial for linking online disinformation to offline radicalisation or coordinated threat behaviour.

Compounding this issue is the fragmentation of data across the various social media platforms, each with distinct formats, access policies, user behaviours and moderation practices. A disinformation campaign may begin on fringe platforms or encrypted channels, then migrate to mainstream social media in order to reach a wider audience. However, due to ***inconsistent data access*** and ***varying degrees of openness across platforms***, FERMI's technical tools faced difficulty in capturing the full lifecycle and spread of a given narrative. This siloed view hinders the ability to understand how narratives mutate across ecosystems, which actors are driving the transition between platforms, and when mitigation measures (such as content takedowns or fact-checking labels) are applied and whether they are effective.

Finally, the ***lack of standardised metadata*** (e.g., timestamps, engagement metrics, identifiers linking content across platforms, etc.) posed a major barrier to inter-platform correlation and campaign mapping. Even when partial data was available, aligning content temporally and thematically across sources requires significant manual effort or complex inference. For example, a video spreading disinformation might first appear on TikTok, get screen-recorded and reshared on X with a different caption and later be referenced in a Reddit thread. Without consistent timestamps, user identifiers or cross-platform metadata, it is extremely difficult to determine whether these are independent posts or part of a coordinated disinformation campaign.

These inconsistencies reduce the effectiveness of AI-driven approaches that depend on temporal and semantic coherence to detect coordinated inauthentic behaviour or narrative shifts and turned out too complex and time-consuming to

[35] Denniss, E., and R. Lindberg. "Social media and the spread of misinformation: infectious and a threat to public health." *Health Promotion International* 40, no. 2. 2025. https://doi.org/10.1093/heapro/daaf023.

be mitigated during the lifetime of the FERMI project. Addressing this challenge will require not only technical advances but also ***improved cooperation from social media platforms and policy frameworks*** that enable consistent, cross-platform access to relevant disinformation data for trusted research and security applications.

4.4 Interoperability and integration with intelligence workflows

One of the key technical challenges identified in FERMI was the difficulty of integrating disinformation analysis tools into existing intelligence workflows used by LEAs and national security bodies. These workflows are often rigid, highly regulated and rely on legacy systems which are not designed for the processing of complex, high-volume, open-source data such as the ones found on social media platforms. Integrating AI-driven insights into such environments requires careful ***alignment with operational timelines and institutional trust frameworks***, none of which can be achieved through stand-alone or siloed systems.

Interoperability issues also stemmed from the heterogeneity of tools and data formats used across different institutions. Disinformation analysis typically involves a mixture of structured and unstructured data, multiple languages and evolving taxonomies of threat indicators and campaign patterns. Without a ***common data model or standardised interfaces***, exchanging and reusing insights across systems becomes highly inefficient. Within FERMI, efforts to develop flexible, modular architectures and APIs revealed that technical compatibility alone is insufficient. Semantic alignment, shared ontologies and contextual interpretation mechanisms are equally necessary to ensure that analytical outputs are actionable to downstream users.

Furthermore, operational integration was hindered by the limited information about many AI components. LEA personnel often require a clear understanding of how risk indicators are generated or how certain conclusions are drawn, particularly when these feed into policy decisions. Black-box models, even if accurate, are unlikely to gain institutional acceptance unless ***accompanied by robust interpretability features, traceability mechanisms and auditability***. The FERMI platform has been revised accordingly in view of LEA end-user feedback. The lessons from FERMI strongly indicate that future systems must prioritise not only interoperability at the technical layer but also trustworthiness and usability at the human layer, in order to ensure meaningful adoption within real-world LEA environments.

4.5 Lessons learned from deployment and testing

The FERMI platform consists of multiple components, developed by different organizations, which interact with each other to produce their final results. To facilitate the deployment of these components, after the finalization of the architecture, deployment pipelines were set up for all components so that each partner could deploy the new version of their component on their own. This enabled each partner, following the agreed communication interfaces, both to be able to work on their component following their own timeline, but also ***to deploy without dependence on the others***. Therefore, after resolving initial issues, deployment became the responsibility of each partner.

To further streamline the process while maintaining the continuous operational availability of the platform, tests became necessary. With these tests, it was necessary to cover both the functionality but also verify that the agreed communication interfaces were followed. For tests like these, which were component specific, it was natural to become the responsibility of each component developer and have them run automatically as part of the deployment process. Including the automated tests in the deployment process as expected, ***reduced the bugs in the deployed versions*** significantly.

Despite the automations in testing and deployment, manual testing by humans is important as well. Human testing can provide significant information about the way the platform is actually used, since as is common in software users do not always follow the flow the designer envisioned. Within FERMI this feedback was very important as it allowed ***uncovering software issues*** and enabled significant improvements in the user interface as well, in order to achieve ***better user experience***.

5 Legal and Ethics Challenges and Lessons Learned

5.1 Adopting a common legal framework/standardisation

The common thread amongst the EU policy documents on disinformation notwithstanding, there still is no universally accepted definition. Different countries adopt different definitions, and as such the definition is ultimately left to the subjective discretion of platforms themselves that have to moderate content. It is ***challenging to find common ground among stakeholders for reaching consensus for a universally accepted definition on disinformation***. Moreover, the vast majority of standards produced currently is done at the European or international level. No initiative was identified at the domestic level, from the bilateral meetings held with various national standardisation bodies, in terms of disinformation. Furthermore, there is currently no specifically-tailored adopted standard on disinformation, either at the EU or the international level.

5.2 The AI Act and the changing legal framework

The Artificial Intelligence Act (*hereafter*, AI Act), adopted in 2024, established rules regarding certain AI systems, in order to foster the development of human-centric and trustworthy AI across the EU.[36] It adopts a risk-based approach, with AI systems being categorised based on the risk they pose to health, safety and fundamental rights. In particular, it defines four levels of risk for AI systems: unacceptable risk, high risk, limited risk and minimal risk. Subsequently, different obligations are imposed, in particular, on providers and deployers of AI systems, according to this risk-based categorisation.[37]

The AI Act does not generally apply to AI-related research, testing and development activities.[38] However, certain provisions are relevant in case AI-driven technologies developed for the law enforcement context are placed on the market or put into service. For example, the AI Act prohibits the provision or deployment of AI systems making risk assessments of individuals in order to assess or predict the likelihood of them committing a crime, based solely on the

[36] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (*hereafter*, AI Act), art. 1.

[37] AI Act, art. 8-17.

[38] AI Act, art. 2.8.

profiling[39] or the assessment of their personality traits and characteristics.[40] The prohibition, however, does not apply in case the AI systems are used to simply support a human assessment already based on objective and verifiable facts linked to a crime. Additionally, the AI Act categorises AI systems as high-risk if they are used in the law enforcement context to assess the risk of an individual committing a crime, or to assess personality traits and characteristics or past criminal behaviours of individuals or groups.[41] AI systems used to profile individuals in the law enforcement context are also considered as high risk.[42]
Accordingly, two general observations of relevance to FERMI can be made. First of all, sentiment analysis can imply an assessment by law enforcement authorities of social media users' characteristics or groups therein.[43] This aspect should be considered when providing or deploying tools which analyse the spread and emotional tone of disinformation content. Secondly, whilst the FERMI platform is not designed to provide crime predictions on specific individuals or groups, , any crime predictions concerning high-risk areas based on historical data or geographical information carry some ethical concerns that should not be underestimated.[44] Ultimately, the applicability of the AI Act to AI tools developed in FERMI, or in similar projects, should be subject to further analysis and adequately considered in future the provision and deployment.

---

[39] The definition of 'profiling' is enshrined in the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (hereafter: General Data Protection Regulation), art. 4(4). Profiling is defined as 'any form of automated processing of personal data consisting of the use of personal data to evaluate certain personal aspects relating to a natural person, in particular to analyse or predict aspects concerning that natural person's performance at work, economic situation, health, personal preferences, interests, reliability, behaviour, location or movements'.

[40] AI Act, art. 5.1(d).

[41] AI Act, art. 6.2, Annex III.6(d).

[42] AI Act, art. 6.2 Annext III.6(e).

[43] Gottschalk, T., and F. Pichierri. "About Migration Flows and Sentiment Analysis on Twitter Data: Building the Bridge Between Technical and Legal Approaches to Data Protection." *The Legal and Ethical Issues Workshop @LREC2022*. Marseille, France. 2022, 27-37.

[44] Anastasopoulou, M. "Exploring algorithmic governance: The AI Act and new realities for criminal justice, and fundamental rights." *New Journal of European Criminal Law* 16, no. 2. 2025, 176-196; Cuypers, A. "Minority Report in the EU? The AI Act's Weak Spot on Crime Prediction." *CiTiP Blog*. 2025. https://www.law.kuleuven.be/citip/blog/minority-report-in-the-eu-the-ai-acts-weak-spot-on-crime-prediction/?utm_source=chatgpt.com.

5.3 Protecting the data and identities of data subjects

Whenever possible data subjects are asked to provide informed consent to having their data processed in accordance with the GDPR article 6 1. (a). However, this is completely infeasible as far as the processing of social media data is concerned. In this case, the FERMI project invokes article 6 1. (f) of the GDPR, which introduces legitimate interest as a lawful basis for processing. More specifically, it is explained that processing is legal if it is "necessary for the purpose of the legitimate interests pursued by the controller or by a third party except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data."[45]

To ensure those rights are safeguarded to the greatest extent possible, ***anonymisation and aggregation measures*** are carried out. Pre-processing of social media data includes anonymising the user IDs, usernames (user IDs and usernames will be turned into i.e. userA, userB, userX etc.), other identifiers, posts and even the deletion of links that may reveal potentially sensitive characteristics about X users (up to the point of replacing emoji characters with corresponding texts/keywords that will not uncover any of the users' characteristics). Of course, the pre-processed data set would not be available on the FERMI platform. Moreover, the sentiment analysis tool would immediately aggregate social media content without storing any posts or Tweets or personal data, which adds another protective layer to the effort to maintain the "the interests or fundamental rights and freedoms of the data subject which require protection of personal data."

These restrictive measures notwithstanding, all of the tools' functions (grasping the spread of the to-be-examined social media messages, figuring out whether the accounts are operated by real persons or bots, which are crucial steps to collect evidence in the event of illegal activities, and analysing the sentiment of the social media messages) can still be fully implemented and evaluated. Neither of these steps requires that X users are identifiable, instead the message's spread and content need to be properly analysed.

Upon the project's conclusion the tools will be further developed and, once they are fully exploitable, LEAs will have the chance to acquire them and apply them without pre-processing and anonymisation measures. In other words, in real-life use the FERMI platform does not inhibit the identification of social media requiring LEA activities such as investigations.

[45] General Data Protection Regulation, art.6 1. (f).

5.4 The responsible use of digital sources

Disinformation does not exist in a vacuum, strategies often leverage underlying societal vulnerabilities by echoing and exacerbating pre-existing biases and inequalities. The interplay between offline and online vulnerabilities underlines that combating efforts ought to address both realms. Accordingly, the ***responsible use of digital sources*** (which does not have to be synonymous with programming skills, instead the knowledge of the media environment, including the trustworthiness of sources is required) is absolutely crucial to rein in the spread of disinformation. This includes, amongst other things, being able to assess the credibility of websites and online sources, and distinguish between fact, opinion, and disinformation. For example, knowing how to spot signs of a website spreading false allegations (e.g., poor design, grammatical errors, sensationalised headlines) or being able to verify information by cross-referencing it with reputable sources.[46] Accordingly, training activities were conducted and supplementing materials were produced[47] to increase the understanding of disinformation dynamics, build digital trust and enhance critical thinking.

[46] The FERMI consortium. *Digital Trust: A Practical Path to Combating Disinformation and Fostering Resilience*. 2025. https://fighting-fake-news.eu/materials/training-materials.

[47] The FERMI training materials are available at: https://fighting-fake-news.eu/materials/training-materials.

6 Recommendations

6.1 For data management, storage and security

FERMI adopted a privacy-by-design approach to data management, storage and security, ensuring that sensitive datasets remain protected throughout the entire lifecycle of AI model development and deployment. Particular attention was given to the handling of crime occurrence data provided by LEAs, with safeguards in place to ensure secure storage, controlled processing and full compliance with applicable legal and ethical standards.

To address the challenge of training AI models on highly sensitive and non-transferable datasets, ***the project employed Swarm Learning*** – a decentralised machine learning paradigm that ***enables models to be trained locally*** within the secure infrastructure of each participating data provider. Instead of transferring raw data, only model parameters are shared across the swarm network. This significantly reduces the risk of data leakage, unauthorised access or potential re-identification of individuals, as no personal or operational data ever leaves the premises of the LEAs.

The integration of Swarm Learning with GDPR-compliant data governance mechanisms, including secure communication protocols and strict access controls, ensured that the AI development process is both secure and ethically sound. This approach safeguarded individual privacy and the operational integrity of LEAs while simultaneously establishing a replicable and scalable framework for future AI applications requiring the processing of sensitive or regulated data across distributed environments. Drawing from the project's results, ***decentralised privacy-preserving AI architectures***, such as Swarm Learning, should be prioritised and promoted, especially for projects involving the processing of sensitive or regulated datasets such as the ones originating from security stakeholders. Thus, the use of ***privacy-by-design approaches*** in future EU-funded security research projects should be encouraged.

6.2 For technical development

The EU-funded FERMI project explored how disinformation campaigns are used to fuel radicalisation and support terrorism, with a strong emphasis on technical innovation for counterterrorism intelligence support. Through its work, the FERMI consortium has identified key limitations in current analytical tools and infrastructures, highlighting the need for more advanced, automated and interoperable systems.

To effectively understand and counter increasingly sophisticated disinformation campaigns, future efforts should prioritise the development of ***advanced***

analytical models capable of capturing narrative evolution,[48] cross-platform propagation and contextual relevance.[49] This includes the integration of natural language processing (NLP) models adapted to multilingual, low-resource and domain-specific datasets, as well as tools that can assess sentiment,[50] intent and actor influence in real time. These next-generation capabilities should be designed to support intelligence analysts with transparent, explainable outputs that enhance situational awareness rather than overwhelm it with raw data.

In addition, disinformation campaigns operate at scale and speed, often exploiting socio-political moments to destabilise public trust. Addressing this requires robust ***automation of key analytical*** processes - from ingestion and annotation of online data to anomaly detection and campaign attribution. Automation should be guided by ***human-in-the-loop principles*** to ensure interpretability and strategic relevance while reducing cognitive overload. Thus, research and deployment of modular, task-specific AI agents that can operate autonomously but feed into human-supervised workflows for tactical and strategic decision-making should be supported.

6.3 For preparing operational uptake

The above-mentioned steps should also lay the ground for operational uptake. Moreover, operational uptake crucially depends on ***interoperability and privacy-preserving*** mechanisms. This is due to the huge sensitivity and diversity of data involved in counterterrorism intelligence, especially when shared across EU Member States. Future systems should incorporate decentralised learning architectures (e.g. federated or swarm learning) that enable cross-institutional collaboration without requiring centralised access to sensitive datasets. At the same time, ***adherence to common data models, metadata standards and secure APIs*** are essential to ensure seamless integration into national and EU-level platforms. The EC should invest in the standardisation and certification of such infrastructures to support trustworthy, scalable adoption across the security ecosystem.

---

[48] Evangelatos, S., E. Veroni, V. Efthymiou and C. Nikolopoulos. "Modeling Disinformation Spread in Social Networks."

[49] Giudice, M. V. L., et al. "Informative (Dis)information: Exploring the Correlation Between Social Media Disinformation Campaigns and Real-World Criminal Activity." *2024 IEEE International Conference in Electronic Engineering, Information Technology & Education (EEITE)*. Chania, Greece. 2024, 1–6.

[50] Evangelatos, S., E. Veroni, V. Efthymiou and C. Nikolopoulos. "The Nexus Between Big Data Analytics and the Proliferation of Fake News as a Precursor to Online and Offline Criminal Activities." *2023 IEEE International Conference on Big Data. BigData.* Sorrento, Italy. 2023, 4056–64.

6.4 For policy and governance (e.g. guidelines, data-sharing frameworks, cross-border cooperation)

A revision of the DSA to expand the scope of players that can seek access to social media would be advisable.

As explained above, one of the key obstacles all research projects on disinformation need to overcome is the lack of data access, in particular to social media, crime and disinformation datasets. Some access issues could be greatly reduced by adjusting the legal landscape. Social media data is a case in point. As explained above, the Digital Services Act enables experts interested in the subject matter to seek access to comprehensive API datasets of social media providers. However, per the terms and conditions of the DSA, access can only be granted to researchers/research institutions, leaving out partners who do not fall under these categories. Accordingly, a revision of the DSA to ***expand the scope of players that can seek access to social media*** would be advisable. Whilst it makes sense to place clear restrictions on the access to social media data on the part of private sector companies, which should not be easily upended, it would advance future research activities, if companies involved in publicly funded projects were exempted from the data access ban under the DSA.

As far as crime data is concerned, one needs to distinguish between access issues and methodological challenges, namely huge discrepancy between country-specific crime data making it extremely challenging, if not impossible, to feed the same models/train the same tools with them. Adhering to the fundamental management principle that one cannot govern what one does not measure, the creation of a unified data framework would empower both national and EU authorities to identify trends, allocate resources proportionally, and devise evidence-based interventions that reflect the full complexity and diversity of extremist threats across Europe. In that regard, the existing American standards for publishing crime data can serve as a role model. A special focus may be placed on politically motivated crime data, which is crucial for the study of extremist ramifications of disinformation campaigns but is currently measured vastly differently across Europe. The German system for recording politically motivated crime, characterised by its comprehensive categorisation and methodological clarity, offers an exemplary benchmark for Europe-wide efforts. Nevertheless, enhanced accessibility and streamlined procedures at the EU level are essential, so that scholars and practitioners throughout the Union can utilise reliable information to inform preventive and investigative strategies, thereby

fostering an integrated and knowledge-driven response to extremism. Obviously, such standardisation efforts would need to go hand-in-hand with strengthening accessibility. Albeit sensitivities with respect to classification proceedings are fully understandable, EU legislation could set clear standards in terms of crime reporting, which may include at least some basic transparency requirements. That being said, support for federated or swarm learning, as suggested above, may facilitate the training of tools with sensitive data without sharing them.

6.5 For broader societal resilience

Beyond the technical and operational readiness of LEAs mentioned above, the FERMI project also recognised that combating disinformation requires a ***comprehensive approach to digital literacy and public awareness***. Effective measures must operate at the intersection of digital and societal resilience by investing in initiatives that support media literacy, critical thinking skills and fact-checking to build informed digital citizenry. Addressing the complexity of disinformation requires a multi-faceted whole-of-society approach with inclusive efforts stemming from both the public and private sphere. Building digital trust and combating disinformation is crucial for furthering an informed and critical public capable of discerning 'truth' in the age of information overload.

7 Conclusion

7.1 Summary of strategic insights

Whilst there are numerous competing definitions, it seems reasonable to consider a claim a piece of disinformation, if it is based on three elements: 1) a factual or misleading nature; 2) the intention to obtain economic gain or deceive the public and 3) aspired ramifications in the form of public harm. Future disinformation research and policy efforts should embrace this definition and distinguish between disinformation (that meet the three above-mentioned criteria) and other forms of false and misleading allegations.

The understanding of public harm, however, needs to include radicalisation activities, especially crime. So far, the EU's research and policy efforts have placed a strong emphasis on the impact of Foreign Information Manipulation and Interference, whereas the nexus of disinformation and extremist crime has been largely overlooked, with some notable exceptions like the funding of FERMI and VIGILANT.

To advance future research activities access to social media, crime, and disinformation data needs to be facilitated. FERMI's Swarm Learning framework (or similar decentralised privacy-preserving AI architectures) can be used to train tools with sensitive data without sharing the latter, providing a solution for the reluctance to make crime data available. The further development and standardisation of such decentralised privacy-preserving AI architectures should be supported by the EU.

Interoperability and the involvement of humans should feature highly prominently in future research. The latter is also of huge importance from a legal and ethics perspective. Amidst the AI Act, sensitivity to data protection and the role of AI has been steadily growing but might be somewhat eased by anonymising social media and possibly further datasets before processing them any further.

Access to social media data should be expanded by exempting private sector players from the DSA's ban on requesting API data, so long as said data is only used in the framework of publicly funded projects. Alternatively, researchers should be allowed to pass such data on to private sector companies in the context of joint research. Given the huge variance between social media datasets, disinformation datasets should be made available by social media platform providers in a comprehensive and coherent form. More specifically, this should include the development of standardised high-quality multilingual datasets with coherent timestamps, engagement metrics, identifiers linking content across platforms etc. Crime reporting across the EU should comply with

some basic standards, transparency-wise and in the form of creating very specific European disinformation datasets, preferably in the form of a European crime framework. Moreover, societal resilience against disinformation needs to be strengthened.

7.2 The way forward for European counter-disinformation capabilities

These insights should be translated into clear policy activities. These should include the following steps:

1) The threefold definition of disinformation should guide the EU's further policies and research activities. More specifically, the European Council may pass a policy document on the urgency of addressing the disinformation threat and express clear support for a common delineation thereof along the lines of the three focal points mentioned above. Such a policy document should place a special emphasis on the impact of disinformation on crime.
2) Such a universally accepted definition of disinformation ought to be adopted either at the International Organization for Standardization or the European Committee for Standardization - European Electrotechnical Committee for Standardization (CEN-CENELEC) levels, as well as the domestic level, at least in EU Member States, in order to ensure the uniformity of approach by platforms and LEAs tackling disinformation.[51]
3) Based on such a policy document (and the available insights from recent EU-funded projects such as FERMI and VIGILANT), future EU research funding should prioritise efforts to mitigate the impact of disinformation on crime. In that regard, interoperability, tool independence, pre-processing anonymisation of research data, and federated or swarm learning should feature prominently.
4) DG Home and Europol are encouraged to support and subsidise the use of federated or swarm learning tools by LEAs across the EU.
5) The DSA should be revised to
    a. enable private sector companies to get access to social media data for their work on publicly funded research projects and to
    b. enforce standardisation across social media platforms to ensure a sufficient level of coherence.

[51] The FERMI project produced the two recommendations for operational standards on disinformation and responsible use of AI that might form the basis of initiatives for standardisation on disinformation. - Another way these drafts could be adopted, or form the initial basis of future standards, is via a CEN Workshop Agreement. The appropriate funding and substantive expertise ought to be allocated to initiating such adoption.

6) A European crime transparency framework, preferably based on the American Universal Crime Reporting Program, should be established.
7) Building societal resilience through tailored training and awareness initiatives should be an integral part of counter-disinformation strategies. These efforts must bring together all stakeholders, including LEAs, policymakers, tech companies, civil society, the media and the general public, and should adapt continuously to evolving disinformation tactics, emerging technologies, distinct national contexts and the diverse populations both targeted by and instrumentalised within disinformation campaigns.

8 References

Alaphilippe, A. *Disinformation Landscape in Belgium*. Brussels: EU Disinfo Lab. 2023.

Anastasopoulou, M. "Exploring Algorithmic Governance: The AI Act and New Realities for Criminal Justice and Fundamental Rights." *New Journal of European Criminal Law* 16, no. 2. 2025, 176–96.

Associated Press. "Defeated president contests Brazil election." *POLITICO.* 22 November 2022. https://www.politico.com/news/2022/11/22/brazil-election-contested-bolsonaro-00070589.

Bayerisches Staatsministerium des Innern, für Sport und Integration. *Verfassungsschutzbericht 2023*. Munich. 2024.

Berger, J. M. *Extremism*. Cambridge, MA: The MIT Press Essential Knowledge Series. 2018.

Bundesministerium des Innern und für Heimat: Bundesamt für Verfassungsschutz. *Verfassungsschutzbericht 2024*. Berlin. 2025.

Cameron, C. "These Are the People Who Died in Connection With the Capitol Riot." *The New York Times*. 5 January 2022. https://www.nytimes.com/2022/01/05/us/politics/jan-6-capitol-deaths.html.

Cuypers, A. "Minority Report in the EU? The AI Act's Weak Spot on Crime Prediction." *CiTiP Blog*. 3 June 2025. https://www.law.kuleuven.be/citip/blog/minority-report-in-the-eu-the-ai-acts-weak-spot-on-crime-prediction/.

Denniss, E., and R. Lindberg. "Social media and the spread of misinformation: infectious and a threat to public health." *Health Promotion International* 40, no. 2. 2025. https://doi.org/10.1093/heapro/daaf023.

European Commission. *Tackling Online Disinformation: A European Approach. Communication COM/2018/236*. 26 April 2018. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0236.

European Commission. Directorate-General for Communication Networks, Content and Technology. *A Multi-Dimensional Approach to Disinformation: Report of the Independent High Level Group on Fake News and Online Disinformation*. Luxembourg: Publications Office of the European Union. 2018.

European Commission. Directorate-General for Migration and Home Affairs. *Research Projects Help Combat Disinformation Ahead of Elections.* 30 May 2024. https://home-affairs.ec.europa.eu/news/research-projects-help-combat-disinformation-ahead-elections-2024-05-30_en.

Europol. *European Union Terrorism Situation and Trend Report*. Luxembourg: Publications Office of the European Union. 2025.

Evangelatos, S., M. Konidi, E. Veroni, S. Karagiorgou and C. Nikolopoulos. "A Perturbation-Theoretic Model for Fact-Checker Deployment in Dynamic Disinformation Networks." Accepted for publication in *Companion Publication of the 17th ACM Web Science Conference (WebSci).* Dublin: ACM. 2025.

Evangelatos, S., E. Veroni, V. Efthymiou and C. Nikolopoulos. "Modeling Disinformation Spread in Social Networks: Phase Transitions and Mean-Field Analysis." *ACM Transactions on the Web* 19, no.4. 2025, 1-24.

Evangelatos, S., E. Veroni, V. Efthymiou and C. Nikolopoulos. "The Nexus Between Big Data Analytics and the Proliferation of Fake News as a Precursor to Online and Offline Criminal Activities." *2023 IEEE International Conference on Big Data (BigData)*. Sorrento, Italy. 2023, 4056–64.

Farinelli, F. *Conspiracy Theories and Right-Wing Extremism – Insights and Recommendations for P/CVE*. *Radicalisation Awareness Network (RAN)*. Luxembourg: Publications Office of the European Union. 2021.

Farrell, L. "UMD Report: Conspiracy theories fueled more terror attacks in 2020." *National Consortium for the Study of Terrorism and Responses to Terrorism*. 7 July 2022. https://www.start.umd.edu/news/umd-report-conspiracy-theories-fueled-more-terror-attacks-2020.

Gottschalk, T., and F. Pichierri. "About Migration Flows and Sentiment Analysis on Twitter Data: Building the Bridge Between Technical and Legal Approaches to Data Protection." *The Legal and Ethical Issues Workshop @LREC2022*. Marseille, France. 2022, 27–37.

Koehler, D. "Right-Wing Extremism and Terrorism in Europe. Current Developments and Issues for the Future." *Prism*: *The Journal of Complex Operations* 6, no.2. 2016. https://cco.ndu.edu/PRISM/PRISM-Volume-6-no-2/Article/839011/right-wing-extremism-and-terrorism-in-europe-current-developments-and-issues-fo/.

Lo Giudice, M. V., et al. “Informative (Dis)information: Exploring the Correlation Between Social Media Disinformation Campaigns and Real-World Criminal Activity.” *2024 IEEE International Conference in Electronic Engineering, Information Technology & Education (EEITE)*. Chania, Greece, 2024.

Lynas, M. “COVID: Top 10 Current Conspiracy Theories.” *Alliance for Science*. 20 April 2020. https://allianceforscience.org/blog/2020/04/covid-top-10-current-conspiracy-theories/.

Mines, A. “The Fractured Threat Landscape.” *Police Chief Magazine*. 2022, 36-41. https://www.policechiefmagazine.org/fractured-threat-landscape/.

Moilanen, P., M. Hautala and D. Saari. *Disinformation Landscape in Finland.* Brussels: EU DisinfoLab. 2023.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC.

Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act).

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

The FERMI consortium. *Digital Trust: A Practical Path to Combating Disinformation and Fostering Resilience*. 2025. https://fighting-fake-news.eu/materials/training-materials.

van Prooijen, J.-W. “Voters on the Extreme Left and Right Are Far More Likely to Believe in Conspiracy Theories.” *EUROPP – European Politics and Policy at LSE Blog*. 2 March 2015. http://bit.ly/1zS8hW3.

Verfassungsschutz Baden-Württemberg. *Linksextremistische Ausschreitungen bei "Revolutionärer 1. Mai Demonstration" in Stuttgart*. 17 May 2024. https://www.verfassungsschutz-

bw.de/,Lde/Startseite/Meldungen+und+Archiv/Ausschreitungen+am+1_+Mai+in+Stuttgart.

Wardle, C., and H. Derakhshan. *Information Disorder: Toward an Interdisciplinary Framework for Research and Policy Making.* Strasbourg: Council of Europe. 2017.