Finansuoja Europos Sąjunga
NextGenerationEU

Mykolo Romerio universitetas

NAUJOS KARTOS LIETUVA

Project „Misijomis grįstų mokslo ir inovacijų programų įgyvendinimas“ (Project Nr. 02-002-P-0001) report

Project name: ***AICP-FIMI***

Responsible/Implementation partner (-s): ***MRU***

Action result: ***Regulatory Mapping and Related Analysis***

Disclaimer

The information in this document is provided “as is”, and no guarantee or warranty is given that the information is fit for any particular purpose. The content of this document reflects only the author`s view. The users use the information at their sole risk and liability.

AI-assisted tools were used to support the drafting, structuring and linguistic refinement of this document. The final content was reviewed and approved by the authors, who remain responsible for its accuracy, integrity and suitability for use. The use of AI-assisted tools was limited to preparatory and editorial support and was not intended to infringe, misappropriate or otherwise affect any third-party proprietary, confidential, copyrighted or otherwise protected rights or materials

Executive Summary

Foreign influence and manipulation of information during elections is a regulatory, democratic and fundamental-rights concern. It affects not only cybersecurity and public security, but also the conditions under which voters form opinions, receive information and participate in democratic life. The AICP-FIMI system is designed to support the detection and analysis of foreign information manipulation, coordinated inauthentic behaviour, bot networks and troll-farm activity in online environments connected to electoral processes. Because the system operates in a politically sensitive context and uses AI-enabled analysis of online behaviour, it must be assessed not only as a cybersecurity tool, but also as a technology that may affect fundamental rights.

The need for this regulatory assessment arises from the specific nature of the system. AICP-FIMI is not a conventional IT tool operating in a closed technical environment. It analyses public or platform-derived online information, identifies behavioural patterns, generates risk indicators and may support the work of election monitors, public institutions, cybersecurity bodies, researchers or other stakeholders concerned with democratic resilience. Such activities may involve personal data, profiling, political expression, political opinions, platform governance, cybersecurity risk management and election-related decision-making. The legal framework must therefore be evaluated before deployment, not retrospectively.

The main legal frameworks to be analyse are the General Data Protection Regulation, the Artificial Intelligence Act, the Digital Services Act, the NIS2 Directive, Lithuanian cybersecurity and data-protection law, and EU fundamental rights law. The GDPR is essential because online identifiers, account metadata, posts, interaction histories and behavioural features may constitute personal data when they relate to identifiable natural persons. Automated assessment of account behaviour or coordination patterns may also amount to profiling. In an electoral setting, this raises heightened risks because political discussion may reveal or imply political opinions, which are protected as special-category data under the GDPR.

The AI Act is particularly relevant because it expressly addresses election-related AI systems. AI systems intended to influence the outcome of an election or referendum or the voting behaviour of natural persons are classified as high-risk, except for limited cases where natural persons are not directly exposed to the system output, such as purely administrative campaign tools. AICP-FIMI is not designed to persuade voters or direct political campaigning. However, because it operates in an election-integrity context and may generate outputs that influence institutional responses to online political activity, its AI Act classification must be assessed carefully according to its intended purpose, users, outputs and consequences.

The DSA is relevant but should not be overstated. AICP-FIMI does not itself appear to be an online platform, hosting service or intermediary service merely because it analyses online content. Therefore, the DSA's direct platform obligations do not automatically apply to it. However, the DSA is important because it establishes a legal framework for systemic risks on very large online platforms and search engines, including risks to civic discourse, electoral processes, public security and fundamental rights. AICP-FIMI may support research, risk awareness or institutional understanding of those risks, but it should not be presented as a DSA-regulated content-moderation platform unless its actual service model changes.

NIS2 and Lithuanian cybersecurity law are relevant where the system is developed, hosted, operated or used by entities within the cybersecurity regulatory perimeter, or where it supports the cybersecurity functions of public administration, digital infrastructure, managed services, election-related institutions or other essential or important entities. NIS2 requires appropriate and proportionate cybersecurity risk-management measures and establishes structured reporting obligations for significant incidents. For AICP-FIMI, this makes cybersecurity not merely an internal technical issue, but a legal condition for trustworthiness and operational reliability.

The fundamental-rights dimension is the defining feature of this report. Countering FIMI protects democratic processes, but the tools used for that purpose must not themselves undermine democratic freedoms. The EU Charter protects private life, personal data, freedom of expression and information, freedom of association, non-discrimination, voting rights and effective remedy. AICP-FIMI therefore requires a balanced legal assessment: the system must be capable of supporting electoral integrity while avoiding disproportionate monitoring, misclassification of lawful political speech, discriminatory impact, unjustified restriction of civic participation or opaque automated assessments.

The novelty of this report lies in treating AI-based FIMI detection as a rights-sensitive regulatory category. The analysis does not reduce the system to cybersecurity, content moderation or data protection alone. Instead, it examines the intersection between election integrity, AI governance, personal data protection, platform regulation, cybersecurity and fundamental rights. This integrated approach is necessary because the harm addressed by AICP-FIMI is collective and democratic, while the risks created by the system may be individual, social and institutional.

The conclusion of the report is that AICP-FIMI can be legally and socially justified only if its regulatory model is grounded in legality, necessity, proportionality, transparency, accountability and human oversight. The system should be understood as an analytical and early-warning tool, not as a mechanism for automated censorship, enforcement or legal determination. Consequential decisions must remain with competent human actors acting under applicable law. The main body of this report therefore focuses on applicable legal norms; operational guidance, case studies and recent practices should are addressed separately in its annexes.

Table of Contents

List of Figures

No table of figures entries found.

List of Tables

No table of figures entries found.

Introduction

Foreign information manipulation and interference have become one of the most complex risks facing democratic societies. Election-related disinformation campaigns coordinated inauthentic behaviour, automated amplification, impersonation, troll networks and bot activity can distort the public information environment. These practices may reduce trust in institutions, polarise public debate, suppress participation, mislead voters or create confusion about electoral processes. They also create new challenges for regulators because the same online spaces that enable manipulation are also spaces for legitimate political debate, journalism, activism, satire and civic participation.

AICP-FIMI is designed to respond to this environment by using AI-enabled analytical methods to detect patterns associated with foreign influence and manipulation of information. Its purpose is to provide early warning and analytical support in election-related contexts. The system may process publicly available online data, social-media metadata, account behaviour, content patterns, network relationships and signals of coordinated activity. Because these functions operate close to democratic discourse, the system cannot be assessed only from a technical-efficiency perspective. Its legality depends on the way it respects fundamental rights and fits within EU and Lithuanian regulatory requirements.

This report provides a regulatory assessment of AICP-FIMI. The analysis is non-exhaustive and focuses on the legal frameworks that are most directly relevant to the system's intended purpose, technical characteristics and electoral context. It does not seek to address every legal instrument that may be indirectly connected to disinformation, cybersecurity, online platforms or artificial intelligence. Instead, it prioritises those areas of law that are most likely to shape the lawful development, deployment and use of an analytical system designed to detect foreign information manipulation and coordinated inauthentic behaviour.

The system category assessed in this report is an analytical and early-warning platform. AICP-FIMI is not considered here as a general content-moderation or enforcement system. It does not remove content, suspend accounts, impose sanctions, declare information illegal or make legally binding decisions. Any such consequential action would fall outside the analytical function assessed in this report and would require a separate legal basis, a competent decision-making authority and appropriate procedural safeguards.

The structure of the report follows a logic of regulatory relevance. The analysis begins with fundamental rights because election-related artificial-intelligence monitoring must be understood both through the democratic interests it seeks to protect and the rights it may affect. It then examines the GDPR, which governs personal data processing and profiling; the Artificial Intelligence Act, which regulates artificial-intelligence systems according to risk and is particularly relevant in electoral contexts; the Digital Services Act, which addresses systemic risks on online platforms and search engines; NIS2 and Lithuanian cybersecurity law, which concern cybersecurity, resilience and incident response; and relevant Lithuanian legal instruments concerning national security and data protection. The Regulation on transparency and targeting of political advertising is also addressed because FIMI-related analysis may overlap with political advertising, targeting practices and electoral influence.

The report deliberately separates binding legal analysis from guidance, practices and case studies. The main body describes the applicable legal framework and explains its relevance to AICP-FIMI. Materials concerning implementation guidance, recent practices, case studies are addressed separately in the annexes. This structure preserves the clarity of the legal analysis while allowing the annexes to provide the practical and contextual material needed to understand how the regulatory framework may operate in real electoral and online-information environments.

1. Methodology and Scope

This report applies a legal-regulatory and fundamental-rights methodology to the assessment of AICP-FIMI. The analysis is non-exhaustive and focuses on the legal frameworks that are most directly relevant to the system's intended purpose, technical characteristics and electoral context. It does not seek to provide technical implementation instructions, software specifications or operational playbooks. Those matters are addressed separately in the annexes where they support the practical understanding of the legal analysis.

The main body of the report identifies and analyses the legal norms that may govern or materially affect the development, deployment and use of AICP-FIMI. The analysis is structured around the system's legally relevant functions. These functions provide the basis for determining which legal frameworks are most relevant.

The report distinguishes between binding legal sources and broader policy or operational context. EU regulations, EU directives, Lithuanian legal acts and fundamental-rights norms are treated as the primary legal sources for the main analysis. Case law, policy documents, strategic communications materials, NATO resources, best-practice documents and institutional guidance are provided separately. Where relevant, they may be addressed in the annexes to explain the broader context, recent practices or operational relevance of the legal framework.

A fundamental-rights perspective is applied throughout the analysis. This is necessary because AICP-FIMI concerns public discourse, online political communication and elections. The report therefore considers not only whether the system may lawfully process data, but also how its use may affect privacy, personal data protection, freedom of expression, media freedom, freedom of association, non-discrimination, voting rights, due process and access to remedy.

The analysis is also deployment sensitive. The legal classification of AICP-FIMI may vary depending on the actor operating the system, the type of data processed, the purpose of the analysis, the use of the system outputs, and whether the system is deployed by public authorities, private service providers, online platforms, researchers or election-related bodies. For this reason, the report avoids treating every possible deployment of AICP-FIMI as legally identical.

The main report is limited to the description and analysis of applicable legal norms and their relevance to AICP-FIMI. Detailed compliance guidance, technical architecture, implementation measures, safeguards, case studies and recent practices are treated as supporting material and are addressed separately in the annexes. This structure preserves the scientific and legal focus of the main report while allowing the annexes to provide the practical context needed to understand the application of the regulatory framework.

2. Fundamental Rights Framework

The fundamental-rights framework is the starting point for the assessment of AICP-FIMI. The system is designed to support the protection of democratic processes against foreign information manipulation and coordinated inauthentic behaviour. At the same time, it operates by analysing online behaviour, political communication and patterns of public discourse. This gives the system a dual legal character: it may contribute to the protection of democratic rights, but it may also interfere with individual and collective rights if used without clear legal limits, adequate safeguards and meaningful human oversight.

The analysis must therefore be situated within three overlapping layers of rights protection: the EU Charter of Fundamental Rights, the European Convention on Human Rights and national constitutional law. These frameworks are not identical, but they protect closely related values: privacy, personal data, expression, association, non-discrimination, electoral participation and access to an effective remedy. In the context of AICP-FIMI, they should be read together because the system operates in a field where EU law, Convention standards and national constitutional protections may all be relevant.

Under the EU Charter, Article 7 protects private and family life, including communications, while Article 8 protects personal data and requires that such data be processed fairly, for specified purposes and on the basis of consent or another legitimate basis laid down by law, subject to independent control. These rights are directly relevant because AICP-FIMI may process account identifiers, metadata, behavioural features, posts, interactions and other online traces linked to identifiable persons.

The European Convention on Human Rights adds a corresponding protection through Article 8, which protects the right to respect for private and family life, home and correspondence. In an online environment, the Convention framework is relevant because the analysis of digital communications, account behaviour and metadata may affect the sphere of private life even where some information is publicly accessible. The

ECHR therefore reinforces the need for legality, necessity and proportionality in any interference with privacy-related interests.

Freedom of expression is central to the assessment. Article 11 of the EU Charter protects the freedom to hold opinions and to receive and impart information and ideas without public-authority interference, and it also protects media pluralism. Similarly, Article 10 ECHR protects freedom of expression, including the right to receive and impart information and ideas regardless of frontiers. This is particularly important for AICP-FIMI because the system operates in the same online space where lawful political speech, journalism, criticism, satire, campaigning and civic mobilisation take place. A false or poorly contextualised classification of online activity as manipulation could chill lawful expression or delegitimise political participation.

The same concern appears at national constitutional level. National constitutions generally protect political expression, freedom of information, participation in public affairs and democratic pluralism. In the Lithuanian context, for example, the Constitution protects the inviolability of private life and communications, the freedom to express beliefs and receive and disseminate information, equality before the law, the right to participate in governing the country, electoral rights, freedom of association and peaceful assembly. These constitutional guarantees are directly relevant where AICP-FIMI is deployed in Lithuania or used in relation to Lithuanian electoral or public-discourse contexts.

Freedom of assembly and association must also be considered. Article 12 of the EU Charter and Article 11 ECHR protect peaceful assembly and association. This matters because coordinated political behaviour is not inherently unlawful or manipulative. Political parties, civil-society organisations, advocacy groups, journalists, campaigners and citizens often coordinate messages, share content and mobilise online. The legal assessment must therefore distinguish coordinated inauthentic manipulation from lawful collective political activity.

Non-discrimination is another core element of the fundamental-rights assessment. Article 21 of the EU Charter and Article 14 ECHR prohibit discrimination, including on grounds such as political or other opinion, language, religion, national or social origin, association with a national minority and other protected characteristics. This is relevant because artificial-intelligence-based behavioural analysis may produce unequal effects across languages, communities, political groups or minority populations. The risk is therefore not limited to privacy. It also includes discriminatory classification, unequal scrutiny or disproportionate impact on lawful expression by particular groups.

Electoral rights give the system its democratic context. Article 39 of the EU Charter protects the right of the EU citizens to vote and stand as candidates in European Parliament elections, while Article 3 of Protocol No. 1 to the ECHR requires free elections at reasonable intervals by secret ballot. National constitutions and election laws provide the primary framework for national elections. In Lithuania, the Constitution protects citizens' right to participate in governing the country and the right to vote once they reach the age established by the Constitution. These guarantees confirm that election-related artificial-intelligence systems must be assessed not only as data-processing tools, but as technologies operating within the constitutional environment of democratic participation.

The right to an effective remedy is also relevant. Article 47 of the EU Charter and Article 13 ECHR protect access to an effective remedy where rights are affected. At national level, constitutional guarantees commonly provide judicial protection for violated rights; in Lithuania, the Constitution expressly provides that a person whose constitutional rights or freedoms are violated has the right to apply to the court. This is important where AICP-FIMI outputs are used in a way that affects individuals, organisations, accounts, media actors or political participants. If an analytical output contributes to a consequential decision, affected persons should not be left without meaningful explanation, review or remedy.

Finally, limitations on rights must satisfy a strict legal standard. Article 52 of the EU Charter requires any limitation on Charter rights to be provided for by law, respect the essence of those rights and freedoms, and comply with necessity and proportionality in pursuit of a recognised general interest or the protection of the rights and freedoms of others. The ECHR applies a comparable structure to qualified rights such as Articles 8, 10 and 11, which permit restrictions only where they are prescribed by law, pursue a legitimate aim and are necessary in a democratic society. National constitutional law also generally requires that

restrictions on fundamental rights be grounded in law and justified by constitutionally recognised interests. This limitation framework should guide the entire assessment of AICP-FIMI.

3. GDPR

The GDPR is one of the central legal instruments applicable to AICP-FIMI. It applies where the system processes personal data. Public online data remains personal data where it relates to an identified or identifiable natural person. GDPR Article 4 defines personal data as any information relating to an identified or identifiable natural person, including identification through online identifiers. It also defines processing broadly and defines profiling as automated processing used to evaluate personal aspects such as behaviour, interests, reliability, location or movements.

AICP-FIMI may process several categories of personal data: social-media handles, account identifiers, post histories, interaction patterns, network links, timestamps, linguistic features, inferred behavioural characteristics and technical metadata. Even if the system does not seek to identify individuals by name, the data may still relate to identifiable users or account operators.

The GDPR principles in Article 5 are directly relevant. Personal data must be processed lawfully, fairly and transparently; collected for specified, explicit and legitimate purposes; limited to what is necessary; accurate; retained no longer than necessary; and secured against unauthorised or unlawful processing. The controller must also be able to demonstrate compliance. These principles are particularly important in election-related monitoring because inaccurate or excessive processing may affect political expression and participation.

The lawful basis for processing depends on the operator and deployment model. If the system is operated by a public authority under a statutory election-integrity, cybersecurity or public-interest mandate, Article 6(1)(e) or Article 6(1)(c) may be relevant. If operated by a private research or analytical entity, Article 6(1)(f) may be relevant, subject to a balancing assessment. Public authorities cannot rely on legitimate interests for processing carried out in the performance of their tasks. The applicable lawful basis must therefore be determined by the role and legal mandate of the operator.

Political opinions receive special protection under the GDPR. Article 9(1) prohibits the processing of personal data revealing political opinions unless one of the exceptions in Article 9(2) applies. In the AICP-FIMI context, this is particularly relevant because election-related online discourse may reveal or imply political opinions even where the system is not designed to identify them directly. Posts, hashtags, account interactions, campaign-related activity and network associations may all create a risk that political opinions are processed directly or indirectly. Where special-category data are involved, the controller must identify both a lawful basis under Article 6 GDPR and a separate Article 9(2) condition. There is no single Article 9(2) basis that will apply to all AICP-FIMI deployments. Article 9(2)(e) may be relevant where the data have been manifestly made public by the data subject, but this should not be treated as a general basis for large-scale behavioural or network analysis, since derived inferences may go beyond what the individual manifestly made public. Where AICP-FIMI is used by a public authority or under a clear statutory mandate connected to election integrity or protection against foreign information manipulation, Article 9(2)(g) may be relevant. This requires a substantial public interest basis in the EU or Member State law, proportionality, respect for the essence of the right to data protection and suitable safeguards. For research-oriented uses, Article 9(2)(j) may also be relevant, provided the processing meets the conditions for research or statistical purposes and includes the safeguards required by Article 89(1). Accordingly, Article 9 GDPR must be assessed on a deployment-specific basis. In practice, Article 9(2)(g) may be the most relevant basis for public-interest electoral integrity uses, Article 9(2)(j) for research-oriented uses and Article 9(2)(e) only for limited cases involving data clearly made public by the individual.

Article 22 GDPR is relevant where automated processing produces decisions with legal or similarly significant effects. It gives individuals the right not to be subject to decisions based solely on automated processing, including profiling, that produce such effects, subject to limited exceptions and safeguards. AICP-FIMI is described as an analytical and early-warning system rather than a system that makes binding decisions. This distinction is important because the risk under Article 22 increases if outputs are used

automatically for sanctions, account restrictions, investigations, public labelling or other consequential actions.

Article 35 GDPR is highly relevant because the system may involve new technologies, large-scale monitoring, profiling and potentially special-category data. A DPIA is required where processing is likely to result in a high risk to rights and freedoms, including systematic and extensive evaluation based on automated processing, large-scale processing of special-category data, or systematic monitoring of publicly accessible areas on a large scale. In an electoral FIMI context, these triggers are likely to be important.

4. AI Act

The AI Act is a core legal instrument for AICP-FIMI because the system uses AI-enabled methods to analyse behaviour, detect patterns and generate risk indicators. The Act is based on a risk-based framework and imposes stronger obligations on high-risk AI systems.

The most significant election-related provision is the classification of certain AI systems as high-risk. The AI Act states that AI systems intended to influence the outcome of an election or referendum or the voting behaviour of natural persons in the exercise of their vote are high-risk, except where natural persons are not directly exposed to the output, such as tools used for purely administrative or logistical political-campaign purposes.

AICP-FIMI is not described as a persuasion or campaigning tool. Its stated function is detection and analysis of manipulation. However, because its outputs may influence institutional understanding of electoral risks, platform responses or public communication about information manipulation, the AI Act assessment must focus on intended purpose, deployment context, users, exposure of natural persons to outputs, and consequences of the system's classifications.

The AI Act also recognises that an AI system listed in Annex III may not be high-risk where it does not pose a significant risk of harm to health, safety or fundamental rights, including where it does not materially influence the outcome of decision-making. This makes the distinction between analytical indicators and consequential decisions legally important.

For AICP-FIMI, the AI Act is not only a technical compliance framework. It is also a fundamental-rights framework because election-related AI systems may affect political participation, expression, equality and democratic legitimacy. The AI Act analysis should therefore be integrated with the Charter and GDPR analysis.

5. Digital Services Act

The DSA is relevant to AICP-FIMI, but its role must be described accurately. The DSA regulates intermediary services and imposes particular obligations on online platforms, with additional systemic-risk obligations for very large online platforms and very large online search engines. AICP-FIMI does not appear to host user content, provide an online platform, operate a search engine or moderate content directly. On that basis, it should not be characterised as directly subject to DSA platform obligations solely because it analyses online information.

The DSA remains relevant because it provides the EU's legal framework for systemic platform risks, including risks connected to disinformation and elections. Article 34 requires VLOPs and VLOSEs to assess systemic risks, including actual or foreseeable negative effects on fundamental rights and on civic discourse, electoral processes and public security. The Regulation also identifies risks arising from coordinated disinformation campaigns, fake accounts, bots and automated or partially automated behaviour.

AICP-FIMI may therefore be relevant as a tool that contributes to research, detection, identification or understanding of systemic risks. The DSA also provides for access to publicly available platform data by qualifying researchers for research that contributes to detecting, identifying and understanding systemic risks. This is relevant where AICP-FIMI is used in a research or public-interest analytical context.

The DSA's fundamental-rights approach is particularly important. It requires systemic-risk mitigation by VLOPs and VLOSEs to take account of fundamental rights and to give particular consideration to freedom of expression. This reinforces the conclusion that FIMI detection must avoid unnecessary interference with lawful political speech

6. NIS2 Directive and Lithuanian Cybersecurity Law

The NIS2 Directive is relevant where AICP-FIMI is developed, hosted, operated or used by entities within the Directive's sectoral and organisational scope, or where it supports cybersecurity functions for such entities. This may include public administration, digital infrastructure, managed services, security-relevant services or election-related institutional environments depending on the deployment model.

Article 21 requires essential and important entities to take appropriate and proportionate technical, operational and organisational measures to manage cybersecurity risks and prevent or minimise the impact of incidents. These measures include risk analysis, incident handling, business continuity, supply-chain security, secure development and maintenance, vulnerability handling, testing, cyber hygiene, cryptography, access control and authentication.

Article 23 establishes reporting obligations for significant incidents. Affected entities must provide an early warning within 24 hours, an incident notification within 72 hours, and a final report within one month after the incident notification. These obligations are relevant to AICP-FIMI because the system may itself be part of a regulated entity's cybersecurity environment, may process sensitive information, and may support the detection of malicious foreign information operations or cyber-enabled manipulation.

Lithuanian cybersecurity law is relevant as the national framework transposing and operationalising cybersecurity obligations in Lithuania. In the context of AICP-FIMI, Lithuanian law is particularly important where the system is operated by Lithuanian public bodies, election-related institutions, cybersecurity bodies, research organisations or service providers falling within national cybersecurity rules. The analysis should treat Lithuanian law as part of the binding domestic framework, not merely as strategic background.

7. Lithuanian National Security Strategy

Lithuania's National Security Strategy is relevant because foreign information manipulation and hostile influence operations are part of the national-security environment in which AICP-FIMI is designed to operate. However, the strategy should be treated differently from directly applicable regulatory duties. It provides strategic context and national-security priorities, but it does not replace the need for a specific legal basis for personal-data processing, AI deployment or institutional action affecting individuals.

The legal relevance of the strategy is therefore contextual. It explains why detection of foreign influence and information manipulation may serve a public-interest objective. It does not, by itself, authorise unrestricted monitoring of online political communication. Where AICP-FIMI is used by public authorities, the operational mandate, data-processing powers and limits must be found in applicable national or EU law.

8. Lithuanian Law on Legal Protection of Personal Data

Lithuania's Law on Legal Protection of Personal Data supplements the GDPR at national level. Its relevance lies in institutional, procedural and national-law aspects of data protection, including supervision and national implementation matters. The GDPR remains the main substantive framework for the processing of personal data in the AICP-FIMI context.

Where AICP-FIMI is deployed in Lithuania, Lithuanian data-protection law must be read together with the GDPR. This is particularly relevant where processing is carried out by public authorities, where national law provides the basis for public-interest processing, or where Lithuanian supervisory procedures and remedies apply.

9. Regulation on Transparency and Targeting of Political Advertising

Regulation (EU) 2024/900 on transparency and targeting of political advertising is relevant where AICP-FIMI analyses political advertisements, advertising delivery, targeting practices or campaign-related manipulation. The Regulation establishes harmonised rules for political advertising and addresses concerns relating to information manipulation, foreign interference, targeting and transparency.

AICP-FIMI is not itself necessarily a provider of political advertising services. However, the Regulation is relevant if the system's data sources or outputs concern political advertising, ad targeting, ad

delivery, sponsor transparency or cross-border political influence. It should therefore be included as an applicable legal instrument in the electoral-context analysis, but only to the extent the system interacts with political advertising data or supports assessment of political-advertising transparency.

10. Law Enforcement Data Protection Framework

Where AICP-FIMI is used for general election monitoring, research, cybersecurity or public-interest analysis, the GDPR will normally be the primary data-protection framework. However, if the system is used by competent authorities specifically for the prevention, investigation, detection or prosecution of criminal offences, the Law Enforcement Directive and its national implementing framework may become relevant. The GDPR itself states that it does not apply to processing by competent authorities for law-enforcement purposes, which is governed by a specific Union legal act, Directive (EU) 2016/680.

This distinction matters because the same technical system may fall under different data-protection regimes depending on the operator and purpose. AICP-FIMI used for civil electoral resilience analysis is not legally identical to AICP-FIMI used for criminal investigation.

11. Conclusion

AICP-FIMI addresses a genuine democratic and security concern: the manipulation of information environments during elections by foreign or coordinated actors. The system's purpose is legitimate in principle, because democratic societies have a strong interest in protecting electoral integrity, public security and the conditions for informed political participation. However, the legal analysis shows that the legitimacy of the objective does not remove the need for strict fundamental-rights and regulatory safeguards.

The central conclusion is that AICP-FIMI must be assessed as a rights-sensitive AI system operating in an electoral context. It is not sufficient to classify it simply as a cybersecurity platform, a disinformation tool, or a cloud analytics system. Its legal significance comes from the combination of AI-enabled behavioural analysis, public online data, political discourse, possible profiling, election-related outputs and institutional use.

The GDPR is essential because the system may process personal data and conduct profiling, including in contexts where political opinions may be revealed or inferred. Public availability of data does not eliminate GDPR protection. Where election-related analysis involves systematic and large-scale monitoring, automated evaluation, or special-category data, the risk to rights and freedoms becomes substantial.

The AI Act is also central because EU law specifically recognises election-related AI as a high-risk area where systems are intended to influence election outcomes or voting behaviour. AICP-FIMI is not designed to influence voters, but its proximity to electoral processes means that its intended purpose, output exposure and effect on decision-making must be assessed carefully. The distinction between analytical indicators and consequential decisions is legally decisive.

The DSA is relevant but should be framed correctly. AICP-FIMI is not automatically subject to DSA platform obligations unless it provides regulated intermediary or platform services. Its stronger connection to the DSA lies in the Regulation's systemic-risk framework for very large online platforms and search engines, especially risks affecting civic discourse, electoral processes, public security and fundamental rights.

NIS2 and Lithuanian cybersecurity law are relevant because the system may operate within cybersecurity-sensitive institutional environments and may support the detection of hostile information operations. Cybersecurity obligations are therefore part of the legal trust model for AICP-FIMI, especially where the system is hosted, operated or used by regulated entities.

Counter-FIMI technologies protect democracy only if they do not undermine the rights that make democratic participation meaningful. The rights to privacy, personal data protection, freedom of expression, media pluralism, freedom of association, non-discrimination, voting and effective remedy must shape the interpretation of every applicable legal framework.

The final regulatory position is therefore that AICP-FIMI may be justified as an analytical and early-warning system, but not as an automated enforcement or censorship mechanism. Its outputs should be understood as risk indicators requiring human assessment and lawful institutional action. The main report

should remain focused on applicable legal norms, while practical guidance, case studies, safeguards, technical controls and recent practices should be addressed separately in annexes.

Bibliography

- Charter of Fundamental Rights of the European Union. Available at: https://eur-lex.europa.eu/eli/treaty/char_2012/oj/eng
- European Convention on Human Rights. Available at: https://www.echr.coe.int/documents/d/echr/convention_eng
- Regulation (EU) 2016/679 - General Data Protection Regulation (GDPR). Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng.
- Regulation (EU) 2024/1689 - Artificial Intelligence Act. Available at: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
- Regulation (EU) 2022/2065 - Digital Services Act (DSA). Available at: https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng
- Directive (EU) 2022/2555 - NIS2 Directive. Available at: https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng
- Regulation (EU) 2024/900 - Transparency and Targeting of Political Advertising Regulations. Available at: https://eur-lex.europa.eu/eli/reg/2024/900/oj/eng
- Directive (EU) 2016/680 - Law Enforcement Data Protection Directive. Available at: https://eur-lex.europa.eu/eli/dir/2016/680/oj/eng
- Republic of Lithuania National Security Strategy. Available at: https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/TAIS.167925/asr
- Republic of Lithuania Law on Cyber Security. Available at: https://e-seimas.lrs.lt/portal/legalAct/lt/TAD/f6958c2085dd11e495dc9901227533ee/asr
- Republic of Lithuania Law on Legal Protection of Personal Data. Available at: https://www.e-tar.lt/portal/lt/legalAct/TAR.5368B592234C/asr
- EU Action Plan Against Disinformation. Available at: https://www.eeas.europa.eu/sites/default/files/action_plan_against_disinformation.pdf
- NATO - NATO's approach to counter-information threats. Available at: https://www.nato.int/en/what-we-do/wider-activities/natos-approach-to-counter-information-threats
- NATO Strategic Communications Centre of Excellence. Available at: https://stratcomcoe.org

ANNEX 1
RECENT PRACTICES

1. Scope of sources

1.1. European Union policy documents and reports

1. Coordinated Plan on AI (2021 Review)[1]

The European Commission Coordinated Plan on Artificial Intelligence establishes EU AI principles prior to the entry into force of the AI Act, which clearly clarifies the context in which the AI Act operates, it emphasizes "AI for the Public Sector" (including law enforcement).

It outlines a comprehensive EU framework for developing human-centric and trustworthy AI, emphasizing that compliance with the GDPR and data protection legislation is a prerequisite for AI development and data sharing. It highlights the Commission's proposal for the Artificial Intelligence Act, a horizontal regulatory framework that adopts a risk-based approach to safety and fundamental rights, distinguishing between high-risk and lower-risk AI systems.

Relevance:

- The document outlines a strategy to "accelerate investments in AI technologies" to drive resilient economic and social recovery. For a project developing an AI-driven platform, this signals a favorable environment for securing support, as the EU aims to "act on AI strategies and programmes" to ensure the region benefits from "first-mover adopter advantages".
- The strategy explicitly aims to ensure that "AI works for people and is a force for good in society". The AICP-FIMI project's focus on protecting democratic processes fits the broader mandate to utilize AI for societal benefit. Furthermore, the plan emphasizes "building strategic leadership in high-impact sectors," which serves as a relevant context for election security platforms.
- The Coordinated Plan functions alongside the AI Act, which establishes the first legal framework to address AI risks. Implementing the project will require navigating these upcoming rules, specifically designed to "address the potential high risks AI poses to safety and fundamental rights". This is critical for the project, as identifying bots and analyzing information flows likely intersects with the "harmonised rules" intended to position Europe as a global leader in safe and ethical AI.

2. Report of the High Representative of the Union for Foreign Affairs and Security Policy to the Council - "Annual Progress Report on the Implementation of the Strategic Compass for Security and Defence (2025)[2]

Report focuses on strengthening cybersecurity through the implementation of the Cyber Resilience Act and the Cyber Solidarity Act, aiming to improve detection, preparedness, and response to large-scale threats, alongside the Network and Information Security (NIS) Directive for critical infrastructure.

Relevance:

- The document explicitly identifies Foreign Information Manipulation and Interference (FIMI) and hybrid strategies as critical security threats that "undermine our democracies, security, societies and lives". It confirms the active operational use of the EU FIMI Toolbox and the EU Hybrid Toolbox to protect democratic integrity, specifically citing successful interventions during the June 2024 European elections and the Moldovan referendum and elections in late 2024 against intensifying Russian destabilization efforts.

- The report underscores that hybrid activities have expanded to include sabotage, cyberattacks, and the weaponization of information, necessitating a coordinated response.
- The document also highlights the importance of the FIMI Information Sharing and Analysis Centre (FIMI ISAC), which serves as a platform to reinforce the analytical capabilities of civil society organizations in countering disinformation. Furthermore, the text emphasizes a "whole-of-society approach" to preparedness, linking civilian and military efforts to build resilience against these multifaceted risks, which directly validates the need for platforms capable of identifying bots and manipulation at scale.
- The document mentions other new regulations that are relevant to the project:
    - Cyber Resilience Act: Entered into force in December 2024. It sets out common cybersecurity requirements for products with digital elements (hardware and software). Since the project aims to build a cloud platform, it will have to meet these security standards to protect itself from risks.
    - Cyber Solidarity Act: Entered into force in February 2025. It includes a European cybersecurity alert system. Therefore, successfully launched platform that identifies bot farms could potentially be integrated into this ecosystem or act as a source of information about incidents.

3. European Commission. European approach to artificial intelligence, policy review (2025)[3].

The European approach to artificial intelligence - European Commission relevant policy review - outlines a comprehensive strategy focused on excellence and trust, aiming to make the EU a global leader in AI while safeguarding democratic values and fundamental rights.

Relevance:

- The policy review emphasizes the core goal that AI systems used in the Union are safe, transparent, and human-centric. Adherence to these goals is an important factor for the further success of the project.
- The document underscores that building high-performance, robust systems requires high-quality data, which is supported by broader initiatives like the EU Cybersecurity Strategy and the Data Union Strategy.

4. AI Continent Action Plan (April 2025)[4]

The provisions of "AI Continent Action Plan" (COM(2025) 165) focus on the implementation of the EU AI Act to create a single market for trustworthy and human-centric AI, establishing support structures like the AI Office and an AI Act Service Desk to assist with regulatory compliance. The plan emphasizes that the AI Act will function alongside existing legislation such as the GDPR (referenced regarding the interplay of laws) and underscores strict safeguards for data security, confidentiality, and integrity within the upcoming Data Union Strategy.

Relevance:

- The plan underscores that "trustworthy and human centric AI" is important not just for the economy, but for "preserving the fundamental rights and principles that underpin our societies". This suggests that EU AI should serve as a protective layer for democratic institutions rather than a destabilizing force.

5. Apply AI Strategy (October 2025)[5]

The Commission Communication COM(2025) 723 sets out a sectoral framework to accelerate the adoption of trustworthy Artificial Intelligence (AI) across 10 key industrial and public sectors, including healthcare, energy, and notably, defence, security, and space. Building on the EU AI Act, the strategy emphasizes the creation of "sectoral flagships" to boost competitiveness while upholding principles of non-discrimination and human-centric design.

Relevance:

- The document explicitly recognizes that AI is reshaping the security landscape, noting that "cybercrime, sabotage and terrorism are blended into hybrid attacks, where AI is often exploited by malicious actors". It highlights that organized criminal groups use AI to "accelerate, upscale and broaden the reach of their illicit activities". To counter these threats, the strategy mandates the "swift delivery of AI-based solutions for internal security," specifically calling for tools that can "detect anomalies," "analyse and respond to incidents more effectively," and fight against the "malicious use of AI".
- The document notes concerns that AI may negatively impact "media plurality" and "cultural diversity". The strategy supports the development of platforms that ensure "trustworthy and human centric AI", aiming to prevent the erosion of democratic values. In the defense sector, the document emphasizes the need for "situational awareness" and "threat surveillance" which aligns with platform's goal of monitoring bot networks and information flows.
- The document provides specific details on how the AI Act will influence the implementation of projects like AICP-FIMI. The Commission is prioritizing the release of "guidelines on the classification of AI systems as high-risk". Since the platform will operate in the sensitive area of elections and democratic processes, it will likely be subject to these high-risk obligations.
- The document stresses that "General-Purpose AI Code of Practice" and prohibitions on unacceptable risks are already applicable, which might be relevant for the implementation of the project.
- One should be aware that ill adopt standardisation requests for reporting "AI systems'... impact on energy consumption," which may apply to the cloud platforms.
- Moreover, the document emphasizes "lawful and effective access to data" for security purposes. It also stresses the need for "secure, efficient, and reliable data sharing" and ensuring "safety is embedded by design", which implies adherence to data protection principles without citing specific GDPR articles.
- The document introduces the "Security Action for Europe (SAFE)", which allows Member States to invest in "AI-powered equipment and cybersecurity".Additionally, the strategy aims to build an "AI toolbox dedicated to public administrations".

6. Progress Report on the Digital Rulebook Implementation. Commission Policy. European Commission. September 2025.[6]

This progress report highlights the main achievements and ongoing efforts to simplify and enhance the efficiency of Europe's digital rulebook, ensure its implementation and maintain its enforcement effectively. The Commission is conducting a comprehensive stress-test of EU digital rules, to ensure that they remain effective and efficient, and that their application is optimised in the real world.

With this in mind, the Commission will put forward a Digital Simplification Package to be adopted in the near future, including a Digital Omnibus to simplify and optimise rules on data, cyber and artificial intelligence, but also an EU Business Wallets to support the regulatory compliance and interactions between businesses and administrations. The Package will also launch a comprehensive digital fitness check.

Relevance:

- The document addresses information and societal threats primarily through the implementation of the Digital Services Act (DSA) and the Artificial Intelligence Act (AI Act), with a strong focus on safeguarding democratic integrity and fundamental rights. To counter information threats, the "Code of Conduct on Disinformation" has been integrated into the DSA framework, utilizing a "Rapid Response System" to report time-sensitive content that threatens the integrity of elections, as demonstrated during the 2024 European and 2025 national elections.
- Regarding societal threats, the document emphasizes protecting "democracy, equality, the rule of law and human rights," ensuring that AI remains "safe and trustworthy," and mitigating systemic risks related to illegal content and public health on large platforms.
- While the document does not explicitly use the term "hybrid threats," it covers the substance of this area through extensive cybersecurity provisions, including the "Cyber Solidarity Act" for managing large-scale incidents and the NIS2 Directive to enhance the resilience of critical sectors and digital infrastructure.
- Psychological threats are addressed implicitly through the priority placed on the "safety and security of minors," with specific enforcement actions taken against platforms hosting pornographic content to restrict access to services deemed a "high risk to minors" and to protect their well-being online.

Quote:

"A hallmark of the Code is the implementation by signatories of the Rapid Response System for elections, which allows non-platform signatories to swift reporting of time-sensitive content, accounts or trends that they view as threats to the **integrity of elections**. This system was used for example for the 2024 European elections, the German elections in February 2025, the 2025 Polish elections."[7]

7. "European Democracy Shield and EU Strategy for Civil Society pave the way for stronger and more resilient democracies". Press Release. European Commission. November 2025.[8]

The Commission has presented the European Democracy Shield, setting out a series of concrete measures to empower, protect, and promote strong and resilient democracies across the EU. According to Commission, an open civic space is at the core of our democracies, and this is why the Commission has also put forward an EU Strategy for Civil Society, for stronger engagement, protection and support to civil society organisations who play essential roles in our societies. Both initiatives had been outlined in the political guidelines and this year's State of the Union address by President von der Leyen.

Relevance:

- The European Democracy Shield and the EU Strategy for Civil Society present measures to protect the key pillars of our democratic systems: free people, free and fair elections, free and independent media, a vibrant civil society and strong democratic institutions.
- The document characterizes Foreign Information Manipulation and Interference (FIMI) as a sophisticated "hybrid attack" used by authoritarian regimes to "erode citizens' trust in democratic institutions," widen societal divisions, and discredit democratic actors. It specifically highlights the technological evolution of these threats, noting that malicious actors utilize "inauthentic use of social media," "fake accounts," and "bot-driven amplification" to distort public debate.
- These operations are increasingly decentralized and operate across multiple platforms to evade detection, often using AI-generated content (deepfakes) to manipulate information flows.
- To counter these psychological and societal threats, the EU is establishing a European Centre for Democratic Resilience to improve "situational awareness" and the collective capacity to "detect and anticipate" disinformation campaigns.

- The document explicitly calls for the development of tools to trace "coordinated inauthentic behaviour" including cross-platform coordination and the use of bots or algorithmic amplification. This confirms a direct operational need for platforms like AICP-FIMI that can identify malicious automated networks. Furthermore, the strategy promotes a "whole-of-society approach" encouraging collaboration between the new Centre, civil society, and researchers to build societal resilience against these disruptive narratives.
- The document provides specific details regarding the AI Act, it explicitly states that the AI Act establishes "transparency obligations" for providers of certain AI systems.
- The AI Act requires the "duty to mark and enable detection" of artificially generated or manipulated content. For this project, this implies that a platform may need to not only identify bots but also verify whether content circulating on social media complies with these mandatory labeling standards.
- The Commission is preparing specific "guidance on the fair, transparent, human-centred and responsible use of AI in electoral processes".
- The document references "personal data protection" in the context of safety recommendations for politicians. It also discusses the EU Digital Identity Wallets to enable "secure identification," which relies on privacy-preserving technologies. While not explicitly detailing GDPR, the text emphasizes "lawful and effective access to data" for researchers under the Digital Services Act (DSA), which is the primary mechanism mentioned for accessing the platform data this project would likely need to analyze bot networks.

8. European Parliament Motion for a Resolution on Hybrid Provocations. European Parliament. October 2025.[9]

October 2025 motion strongly condemns Russia's military and hybrid provocations (e.g., drone incursions). This illustrates the evolving nature of hybrid threats into the physical and military domains.

Among other, EP calls on the Commission and the Member States to urgently increase investments in security and defence in order to close all capability gaps, establish strong deterrence and provide adequate support to Ukraine, ensuring that the EU and its Member States are equipped to address all threats, from hybrid and cyber threats to conventional military challenges, and that planning, research, development, procurement and management of capabilities are all done jointly and through a European lens; stresses, therefore, the importance of using these investments to stimulate joint action at European level, including with Ukraine, in order to improve the efficiency of public spending, enhance interoperability and boost the EU's strategic autonomy, instead of perpetuating the present state of market fragmentation, divergent and incompatible capabilities, wasteful investments and external dependencies.

Relevance:

- The resolution explicitly addresses hybrid and information threats by condemning Russia's systematic use of cyberattacks, disinformation campaigns, electoral interference, and the sabotage of critical infrastructure. It highlights that drones are increasingly used as tools for hybrid warfare by both state and non-state actors and notes the specific danger of Global Satellite Navigation Systems (GNSS) jamming and spoofing.
- To counter these evolving risks, the document urges the EU to adopt a comprehensive defense strategy that utilizes the Hybrid Threats Centre of Excellence and the European Cybersecurity Competence Centre to enhance cybersecurity, situational awareness, and the resilience of critical underwater and digital infrastructure.
- Regarding societal and psychological dimensions, the document emphasizes that security extends beyond military assets to include the protection of democratic values and social stability. It warns that vulnerabilities undermining citizens' trust in institutions could render the EU unstable despite

its military strength, urging Member States to design defense investment plans that also strengthen social cohesion.

- The resolution calls for strategic communication to provide citizens with reliable information channels, ensuring that disinformation campaigns seeking to exploit security incidents and manipulate public perception are effectively countered.

9. Eighth progress report on the implementation of the 2016 Joint framework on countering hybrid threats and the 2018 Joint communication on increasing resilience and bolstering capabilities to address hybrid threats. Joint staff working document. European Commission. October 2024.[10]

This document is the Eighth Progress Report (SWD(2024) 233 final) from the European Commission and the High Representative, summarizing the EU's actions to counter hybrid threats and bolster resilience from July 2023 to June 2024. The report highlights efforts to protect democratic processes (especially the 2024 European **elections**) and critical infrastructure (energy, maritime, and space) against threats like FIMI (Foreign Information Manipulation and Interference), cyberattacks, and the instrumentalization of migration.

Key outcomes include validating the framework for EU Hybrid Rapid Response Teams, strengthening cybersecurity among EU institutions, advancing the legal frameworks of NIS 2 and CER Directives, and integrating the security of AI technologies into the economic security risk assessment.

***Relevance*:**

- The document outlines a comprehensive EU response to hybrid threats, which are increasingly driven by Russia's war of aggression against Ukraine and conflicts in the Middle East, utilizing tactics such as cyberattacks, the sabotage of critical infrastructure, and the instrumentalization of migration.
- To counter these evolving dangers, the EU has advanced its Hybrid Toolbox, notably validating the framework for EU Hybrid Rapid Response Teams to provide tailored assistance to Member States and partners. Significant emphasis is placed on information threats, specifically Foreign Information Manipulation and Interference (FIMI) and disinformation; the report details measures taken to protect the integrity of the 2024 European elections, including the Defence of Democracy Package, the Digital Services Act, and the expansion of the Code of Practice on Disinformation to mitigate systemic online risks.
- Regarding societal and psychological security, the report highlights efforts to build resilience against radicalization, violent extremism, and hate speech, viewing these as vulnerabilities that can be exploited to destabilize society. The EU Internet Forum plays a key role in moderating terrorist content and addressing the emerging risks of generative AI mixed with disinformation.
- Furthermore, the document stresses a "whole-of-society" approach to psychological defense, promoting digital literacy and education to help citizens critically engage with the online world, thereby strengthening trust in democratic institutions against foreign interference.

10. Union Rolling Work Programme for European cybersecurity certification. Policy and Legislation. February 2024.[11]

The Union Rolling Work Programme for European cybersecurity certification identifies strategic priorities for future European cybersecurity certification schemes. This first URWP points to areas where European cybersecurity certification schemes are envisaged due to legislative developments as well as to areas for future reflection regarding cybersecurity certification, which might eventually lead to requests for new schemes where necessary and appropriate. Furthermore, it outlines the strategic priorities to be considered when preparing any European cybersecurity certification scheme.

Relevance:

- The document focuses on the certification of ICT products and services to ensure they are secure against cyberattacks.
- Regarding relevance of AI Act and regulatory influence, the document explicitly links cybersecurity certification to the AI Act.
- The document notes that the AI Act mentions cybersecurity certification as a method for demonstrating conformity with cybersecurity requirements.
- The framework emphasizes a "security-by-design" and "security-by-default" approach throughout the lifecycle of development. It means that AI platform must integrate security measures (like vulnerability handling) from the earliest design phase, rather than adding them later.
- It highlights the intersection of certification and data protection. The document states that standards used for certification can "sustain requirements... e.g. on personal data". This implies that if the platform seeks certification, it will likely need to adhere to technical standards that ensure GDPR compliance.

11. Council Conclusions on the Future of Cybersecurity. Council of the EU. May 2024.[12]

Adopted in May 2024, these conclusions emphasize the need for effective, non-fragmented implementation of NIS2 and the Cyber Resilience Act (CRA). They call for greater support for SMEs in cybersecurity compliance and explicitly acknowledge the dual benefits and challenges of AI and quantum computing for security.

The Council conclusions recall the importance to focus on implementation, strengthen coordination and collaboration, and avoid fragmentation of cybersecurity rules in sectorial legislation. They also call to further clarify roles and responsibilities in the cyber domain, to strengthen the cooperation in the fight against cybercrime, and to work on a revised blueprint of the cyber crisis management framework.

The support to micro, small and medium size enterprises, and the need to respond to the challenges presented by the new technologies are also highlighted. A multistakeholder approach, including cooperation with the private sector and academia is encouraged to close the skills gap. In light of the changed and rising threat level, the Council finally invites the European Commission and the High Representative to present a revised cybersecurity strategy.

***Relevance*:**

- The document explicitly addresses hybrid and information threats by calling for a coherent approach to risk assessment that takes into account "ongoing efforts to counter hybrid threats, such as physical sabotage and foreign information manipulation and interference".
- It emphasizes that the revised EU cybersecurity crisis management framework ("Blueprint") must be compatible with existing instruments like the "EU Hybrid Toolbox" and the "EU Cyber Diplomacy Toolbox" to effectively manage the full crisis lifecycle across civilian and military domains.
- Regarding psychological and societal threats, the text highlights the dangers of social engineering and the "misuse of emerging and disruptive technologies," specifically citing deepfakes created by AI as a threat to digital trust.
- The Council underscores that cybersecurity is the "cornerstone of a successful digital society," necessary to maintain "public trust" in the systems that underpin the functioning of the economy and society. Additionally, it notes the need to protect against "spill-over risk" where cyber threats impact the broader public and cross-border environments.

12. ENISA NIS360 2024 report: A comprehensive look at cybersecurity maturity and criticality of NIS2 sectors. March 2025.[13]

The European Union Agency for Cybersecurity's first NIS360 report identifies areas for improvement and tracking of progress across NIS2 Directive sectors. The NIS360 is a new product by the EU Agency for Cybersecurity, ENISA, that assesses the maturity and criticality of NIS2 sectors, providing both a comparative and a more in-depth analysis. The goal of the NIS360 is to help national authorities and cybersecurity agencies in the Member States tasked with the implementation of the NIS2, (1) to understand the overall picture, (2) to help them with prioritisation, (3) to highlight areas for improvement, and (4) to facilitate monitoring of sectors' progress. The NIS360 also aims to support policy makers at national and EU level, to give input on policy and strategy development, and initiatives to build up cyber resilience.

Relevance:

- The document focuses primarily on technical cybersecurity maturity and the implementation of the NIS2 Directive. However, regarding societal and hybrid-related concerns, the report identifies the Public Administrations sector as a "prime target for hacktivism and state-nexus operations," highlighting the risk of politically motivated actions and state-sponsored hybrid interference.

Quote:

- "Public administrations: [...] Being a prime target for hacktivism and state-nexus operations, the sector should aim to strengthen its cybersecurity capabilities leveraging the EU Cyber Solidarity Act and exploring shared service models among sector entities on common areas e.g., digital wallets."14

13. ENISA Threat Landscape (ETL). October 2025.[15]

ENISA Threat Landscape (ETL) introduces a revised and concise format designed to deliver insights through a threat-centric approach and enhanced contextualisation. This edition integrates additional analysis of adversary behaviours, vulnerabilities and geopolitical drivers, aimed at both strategic and operational audiences, offering an actionable perspective on trends shaping the EU's cyber threat environment. The ETL 2025 provides an overview of the European cyber threat ecosystem from July 2024 to June 2025, drawing on nearly 4 900 selected and curated incidents. The reporting period highlights a maturing threat environment characterised by rapid exploitation of vulnerabilities and growing complexity in tracking adversaries. This flagship report highlights the shift toward diversified and convergent threat campaigns, which include several findings, e.g:

- AI-Enabled Threats: By early 2025, AI-supported phishing campaigns reportedly accounted for over 80% of observed social engineering, leveragingsynthetic media and jailbroken models.
- Ransomware & State Actors: Ransomware remains the core threat, but state-aligned threat groups have intensified long-term cyberespionage against critical sectors like telecommunications and logistics.

***Relevance*:**

- The document extensively details hybrid and information threats, particularly highlighting the convergence of cyber operations with Foreign Information Manipulation and Interference (FIMI). It describes "hybrid campaigns" where adversaries, notably Russia-aligned actors, use digital platforms like Telegram to recruit individuals for physical sabotage, vandalism, and arson across NATO countries, thereby extending conflict beyond cyberspace.
- A dedicated section on FIMI outlines how state-aligned actors (Russia and China) employ "inauthentic news articles," "fabricated investigations," and AI-driven "deepfakes" to interfere in EU elections, discredit public institutions, and manipulate public perception.

- Regarding psychological and societal threats, the report warns of the "erosion of resilience" caused by campaigns designed to undermine democratic processes and social cohesion. Hacktivist and state-aligned groups exploit polarizing societal topics-such as migration and LGBTQ+ rights-to create division and target public administration and essential services.
- The text also highlights psychological manipulation through "social engineering" and "pig-butchering scams" that build false trust to defraud victims,, as well as "pressure tactics" by ransomware groups aimed at intimidation. Furthermore, the document notes tangible societal harms, ranging from the disruption of critical sectors like healthcare and transport to physical threats such as the kidnapping of crypto-asset holders.

Quotes:

- „Hybrid campaigns […], especially with activities aligned with Russian objectives continuing to impact EU MSs beyond cyberspace. [...] investigative reporting detailed pro-Russia groups using Telegram to recruit EU-based individuals for sabotage, vandalism, arson and influence operations across NATO countries.“16
- „Approximately a quarter of the documented FIMI content focused on degrading the Union through negative narratives. [...] France, Germany and Poland are frequently targeted with narratives aimed at discrediting their government, military and intelligence services, often accusing them of destabilisationefforts abroad or failing in their fundamental duties...“17 (ENISA Threat Landscape, p. 43)

14. 2024 Report on the State of Cybersecurity in the Union. December 2024.[18]

This document marks the first report on the state of cybersecurity in the Union, adopted by ENISA in cooperation with the NIS Cooperation Group and the European Commission, in accordance with Article 18 of the Directive (EU) 2022/2555 (NIS2). The report aims at providing policy makers at EU level with an evidence-based overview of the state of play of the cybersecurity landscape and capabilities at the EU, national and societal levels, as well as with policy recommendations to address identified shortcomings and increase the level of cybersecurity across the Union.

This report provides policy recommendations to strengthen the EU's cybersecurity framework. It Recommends strengthening support for the implementation of NIS2, and calls for the development of an EU horizontal policy framework to address supply chain security challenges faced by both public and private sectors.

Relevance:

- The document identifies malicious cyber activity as a clear component of wider hybrid threats aimed at destabilizing EU society, democracy, and values, particularly through Foreign Information Manipulation and Interference (FIMI) and disinformation campaigns.
- State-aligned actors and non-state groups are increasingly leveraging AI to create fake content and conduct influence operations, with the geopolitical landscape-such as the Russian war of aggression against Ukraine-heavily influencing tactics designed to manipulate civilian populations and impact elections. These advanced hybrid threats, linked to interference and the dissemination of disinformation, are projected to remain top-ranking risks for the EU through 2030.
- Furthermore, the report highlights the psychological and societal dimensions of the threat landscape, noting that hacktivists utilize "Fear, Uncertainty, and Doubt" to amplify the disruptive impact of their operations. Incidents targeting civil society and the general public represented a notable share of observed events, often involving social engineering and data breaches.

- Consequently, the report emphasizes the critical need to strengthen "societal capabilities," recommending harmonized national efforts to improve cybersecurity awareness and cyber hygiene among citizens to ensure the population is resilient against these evolving threats.

Quotes:

- „Malicious cyber activity has become a clear component of wider hybrid threats, such as disinformation and physical acts of sabotage and violence, seeking to undermine and destabilise EU society, democracy and values“19.
- „Hacktivists use common tactics, such as DDoS attacks and website defacements, but also “Fear, Uncertainty, and Doubt” to amplify the impact of their operations“20.

1.2. Recommendations, guidelines, codes of practice

1. Commission Guidelines on Prohibited AI Practices (Feb 2025)[21]

Non-binding but critical guidance on interpreting the "unacceptable risk" categories (e.g., remote biometric identification). These interpretations are vital for dual-use vendors (e.g., camera/drone manufacturers) to know when a police use-case crosses the line into a ban.

These guidelines provide an overview of AI practices that are deemed unacceptable due to their potential risks to European values and fundamental rights. The guidelines specifically address practices such as harmful manipulation, social scoring, and real-time remote biometric identification, among others.

The guidelines are designed to ensure the consistent, effective, and uniform application of the AI Act across the European Union. While they offer valuable insights into the Commission's interpretation of the prohibitions, they are non-binding, with authoritative interpretations reserved for the Court of Justice of the European Union (CJEU). The guidelines provide legal explanations and practical examples to help stakeholders understand and comply with the AI Act's requirements. This initiative underscores the EU's commitment to fostering a safe and ethical AI landscape.

Relevance:

- The document focuses on regulatory prohibitions, it provides definitions of manipulative practices that directly describe the FIMI tactics.
- Harmful Manipulation and Deception: The Guidelines detail prohibited AI practices that deploy "subliminal," "purposefully manipulative," or "deceptive techniques". These are defined as techniques capable of "materially distorting the behaviour of a person" by impairing their ability to make an informed decision. This includes AI systems (like bots or deepfakes) that present false or misleading information to deceive individuals and undermine democratic processes.
- Exploitation of Vulnerabilities: The document highlights the threat of AI systems that exploit the vulnerabilities of specific groups due to their age, disability, or social/economic situation. In the context of FIMI, this is relevant as malicious actors often target disenfranchised or vulnerable populations to amplify divisiveness.
- Societal Harm: The rationale for these prohibitions is to prevent practices that contradict Union values, specifically "respect for human dignity, freedom, equality, democracy, and the rule of law". The text explicitly links these manipulations to "societal harms" and the erosion of trust in democratic institutions.
- The Guidelines introduce strict constraints on how potential platforms as AICP-FIMI can operate, particularly regarding data collection and user analysis, namely:
    - Prohibition on Scraping for Facial Recognition (Article 5(1)(e)). Such platforms must not use AI to scrape facial images from social media to build databases of "known trolls" or bots if those databases are used for facial recognition.

- The Guidelines prohibit the "untargeted scraping of facial images from the internet or CCTV footage" to create or expand facial recognition databases.
- "Untargeted" is defined as scraping without a specific focus on a given individual (e.g., indiscriminately harvesting profile pictures).
- This prohibition applies even if the images are publicly available on social media, as publishing a photo does not constitute consent for inclusion in a database.
- Prohibition on Biometric Categorization of Political Opinions (Article 5(1)(g)). Such platforms cannot use AI to analyze user biometrics (e.g., profile photos or voice data) to infer their political leanings.
- The AI Act prohibits systems that categorize natural persons based on biometric data to deduce or infer their political opinions, race, religious beliefs, or sexual orientation.
- This prevents from using facial analysis algorithms to classify social media users as "politically aligned" with specific foreign regimes based on their profile pictures.
- Prohibition on Predictive Risk Assessment (Article 5(1)(d)). It should be avoided using AI to assess or predict the risk of individuals committing criminal offences (e.g., election fraud) if the assessment is based solely on profiling or personality traits. This is prohibited unless used to support a human assessment based on objective facts.
- GDPR Interaction. The Guidelines emphasize that these prohibitions apply alongside the GDPR and the Law Enforcement Directive (LED). Specifically regarding scraping, the document notes that building facial databases from internet scraping would likely lack a legal basis under GDPR and be considered unlawful processing of personal data.

***Quotes*:**

- "Manipulative techniques are typically designed to exploit cognitive biases, psychological vulnerabilities, or situational factors that make individuals more susceptible to influence. Because of their adaptability, AI systems are also able to respond well to a person's individual circumstances or vulnerabilities and increase the effectiveness and impact of manipulation at scale"22.
- "'Deceptive techniques' deployed by AI systems should be understood to involve presenting false or misleading information with the objective or the effect of deceiving individuals and influencing their behaviour in a manner that undermines their autonomy, decision-making and free choices"23.
- "AI systems enabling 'social scoring' practices may lead to discriminatory and unfair outcomes for certain individuals and groups, including their exclusion from society, as well as social control and surveillance practices that are incompatible with Union values. [...] It also aims to safeguard and promote the Union values of democracy, equality (including equal access to public and private services), and justice"24.
- "The AI Act expressly excludes from its scope AI systems that are 'placed on the market, put into service, or used with or without modification exclusively for military, defence or national security purposes, regardless of the type of entity carrying out those activities'"25.

2. Multilayer Framework for Good Cybersecurity Practices for AI. ENISA Recommendation. June 2023.[26]

In this report, ENISA presents a scalable framework to guide NCAs and AI stakeholders on the steps they need to follow to secure their AI systems, operations and processes by using existing knowledge and best practices and identifying missing elements. The framework consists of three layers (cybersecurity foundations, AI-specific cybersecurity and sector-specific cybersecurity for AI) and aims to provide a step-by-step approach on following good cybersecurity practices in order to build trustworthiness in their AI activities.

Relevance:

- The document primarily approaches relevant areas through a cybersecurity and technical risk management lens, treating AI systems as "socio-technical" entities where technical failures can have broad social and political consequences.
- State-Sponsored Actors: The report identifies "state-sponsored attackers" and "cyberwarriors" as specific threat actors with the means, motives, and opportunities to target AI systems.
- It warns that by 2030, the misuse of AI will be a significant threat, explicitly citing "state-sponsored operatives" who might issue "deep fakes" to deceive.
- The document links AI security to the EU's Common Security and Defence Policy (CSDP), noting that AI is considered a technology that will play a "crucial role for defending the EU" against unconventional security threats.
- The report highlights "poisoning" as a major threat where attackers alter training data or models to modify an algorithm's behavior (e.g., to sabotage results or insert backdoors). Specific examples include data poisoning attacks on stop signs in autonomous driving scenarios.
- The text notes the risk of AI-generated content, such as deep fakes, which necessitates the ability to detect and contain these threats to assist security analysts. It identifies the risk of "model or data disclosure," where sensitive information about the model's configuration or training data is leaked, etc.
- The document addresses psychological aspects primarily in terms of understanding adversaries and the impact of AI on human trust and behavior. E.g., it emphasizes the need for a taxonomy of AI attackers that includes understanding their "psychological profiles," motives, and objectives.
- The document explicitly frames AI threats as "socio-technical," acknowledging that risks extend beyond technical failures to legal, ethical, and democratic concerns. E.g., "Bias" is identified as a specific threat that can lead to "algorithmic discrimination," where automated systems wrongly classify individuals or exclude them from services and rights
- The report asserts that adequate cybersecurity is necessary to "serve European values and the democratic rights of Europeans". It warns that social threats can impact "democracies and society as a whole". It calls for risk assessments that include social threats such as "lack of fairness, lack of interpretability/explainability/equality".

Quote:

"As explained in the NIST AI Risk Management Framework, AI systems are socio-technical in nature, meaning that the threats are not only technical, legal or environmental (as in typical ICT systems), but social as well. For example, social threats – such as bias, lack of fairness, lack of interpretability/explainability/equality – are directly connected to societal dynamics and human behaviour in all technical components of an AI system, and they can change during its life cycle."[27]

3. Fundamentals of Secure AI Systems with Personal Data. EDPB Guidance. April 2025.[28]

This training/guidance material is critical for bridging the technical requirements of secure AI with the legal obligations of GDPR (Data Protection and Security). The Publication outlines a training curriculum for cybersecurity professionals working with personal data and artificial intelligence (AI) in the European Union (EU).

***Relevance*:**

- The document addresses information and societal threats primarily through its taxonomy of AI risks and ethical considerations. It identifies specific domains such as "Misinformation" and "Pollution of the information ecosystem," noting risks related to the spread of false information, the

reinforcement of belief bubbles, and the "disinformation, surveillance, and influence at scale" conducted by malicious actors.

- It highlights "Socioeconomic & environmental harms" as a key risk area, warning against power centralization, increased inequality, the devaluation of human effort, and "Discrimination & toxicity," which includes unfair treatment and exposure to toxic content.
- Regarding psychological threats, the text explicitly references the EU AI Act's prohibition of AI systems that manipulate human behavior through "subliminal techniques" or exploit vulnerabilities based on age or disability to impair decision-making. It also warns of the "Loss of human agency and autonomy" and "Overreliance" on automated systems. While the specific term "hybrid threats" is not defined, the document covers overlapping concerns under "Malicious actors & misuse," describing the use of AI for cyberattacks, weapon development, and mass harm, which are components oftenassociated with hybrid warfare strategies.

Quote:

„AI systems that manipulate human behavior through subliminal techniques, impairing decision-making ability. <...> AI systems that exploit vulnerabilities of individuals based on age, disability, or economic situation."[29]

4. Guidance for Risk Management of Artificial Intelligence systems. European Data Protection Supervisor. November 2025.[30]

This Guidance aims to guide EUIs acting as data controllers in identifying and mitigating some of these risks. More specifically, they focus on the risk of non-compliance with certain data protection principles elicited in the EUDPR for which the mitigation strategies that controllers must implement can be technical in nature – namely fairness, accuracy, data minimisation, security and data subjects' rights. As such, the technical controls listed in this Guidance are by no means exhaustive, and do not exempt EUIs from conducting their own assessment of the risks raised by their specific processing activities. In doing so, it refrains from ranking their likelihood and severity.

Relevance:

- Hybrid threats are characterized by the convergence of cyber operations with physical acts of sabotage, such as the recruitment of individuals via social media for vandalism and arson, often orchestrated by state-aligned groups to destabilize EU society. Information threats, specifically Foreign Information Manipulation and Interference (FIMI), are a primary concern.
- Regarding psychological and broader societal threats, the documents note the use of "Fear, Uncertainty, and Doubt" tactics by hacktivists to amplify the psychological impact of their operations, such as publicizing tampering with operational technology systems.
- Furthermore, AI systems are described as "socio-technical" entities that introduce specific societal risks, including algorithmic bias and discrimination. The EU legislative framework explicitly addresses these by prohibiting AI practices that deploy subliminal or manipulative techniques to distort behavior, thereby aiming to protect fundamental rights and democratic values from these evolving threats.

Quote:

- "This Guidance focus on two specific aspects of the risk management process, namely risk identification and risk treatment; the risk analysis and the risk evaluation aspects are too dependent on the specific processing context and their assessment is better left to each organisation in line with their own risk criteria. This means that EUIs should perform a thorough analysis for each AI system they plan to use in order to also evaluate the likelihood and impact of the risks, and decide on the mitigating measures to address them, as well as on the residual risks."31

5. Guidance on Generative AI, strengthening data protection in a rapidly changing digital era. European Data Protection Supervisor. October 2025.[32]

These EDPS Orientations on generative Artificial Intelligence (generative AI) and personal data protection intend to provide practical advice and instructions to EU institutions, bodies, offices and agencies (EUIs) on the processing of personal data when using generative AI systems, to facilitate their compliance with their data protection obligations as set out, in particular, in Regulation (EU) 2018/1725. These orientations have been drafted to cover as many scenarios and applications as possible and do not prescribe specific technical measures. Instead, they put an emphasis on the general principles of data protection that should help EUIs comply with the data protection requirements according to Regulation (EU) 2018/1725.

Relevance:

- The document addresses societal and security concerns through the lens of fundamental rights and data integrity. It warns that generative AI can exacerbatesocietal threats by magnifying existing biases, leading to "unfair processing and discrimination" in critical areas such as public health, human resources, and public administration.
- Regarding information and security risks, the guidance highlights the danger of "hallucinations" (the generation of false information) and technical vulnerabilities such as "prompt injection" and "jailbreaks" that could be exploited by malicious actors. It mandates human oversight to mitigate risks to vulnerable populations and to prevent "automation bias".

Quote:

- "Biases in generative AI systems can lead to significant adverse consequences for individuals' fundamental rights and freedoms, including unfair processing and discrimination, particularly in areas such as human resource management, public health medical care and provision of social services, scientific and engineering practices, political and cultural processes, the financial sector, environment and ecosystems as well as public administration."33(Guidance on Generative AI, strengthening data protection in a rapidly changing digital era, p. 31)

6. First EDPS Orientations for EUIs using Generative AI. European Data Protection Supervisor. June 2024.[34]

These EDPS Orientations on generative Artificial Intelligence (generative AI) and personal data protection intend to provide practical advice and instructions to EU institutions, bodies, offices and agencies (EUIs) on the processing of personal data when using generative AI systems, to facilitate their compliance with their data protection obligations as set out, in particular, in Regulation (EU) 2018/1725. These orientations have been drafted to cover as many scenarios and applications as possible and do not prescribe specific technical measures. Instead, they put an emphasis on the general principles of data protection that should help EUIs comply with the data protection requirements according to Regulation (EU) 2018/1725.

Relevance:

- The document reflects concerns related to societal and information threats within the framework of fundamental rights and technical security. Regardingsocietal threats, the text warns that generative AI can magnify existing biases or create new ones, potentially leading to unfair discrimination in sensitive areas such as "political and cultural processes," "public administration," and the "provision of social services". It also highlights specific risks to "vulnerable populations and children," particularly in the context of automated decision-making.
- In terms of information threats, the guidance cautions against "hallucinations" (inaccurate or false information) and "harmful or toxic output", while also identifying specific security vulnerabilities

such as "prompt injection," "jailbreaks," and "model inversion attacks" that can compromise system integrity.

Quote:

- "Controllers should, in addition to the traditional security controls for IT systems, integrate specific controls tailored to the already known vulnerabilities of these systems - model inversion attacks, prompt injection, jailbreaks - in a way that facilitates continuous monitoring and assessment of their effectiveness."35

7. Guidelines on the scope of obligations for providers of General-Purpose AI (GPAI) models. July 2025.[36]

The Commission published guidelines on the scope of obligations for providers of general-purpose AI models under the AI Act. The aim is to provide legal certainty to actors across the AI value chain by clarifying when and how they are required to comply with these obligations. These guidelines are part of a broader package tied to the entry into application on 2 August 2025 of the EU-wide rules for providers of general-purpose AI models. They complement the General-Purpose AI Code of Practice that the Commission received from independent experts.

***Relevance*:**

- Document's primary focus is on determining the scope of obligations for general-purpose AI models-specifically regarding technical thresholds (such as training compute measured in FLOPs) and market classification-rather than analyzing specific geopolitical or psychological warfare tactics.
- The document addresses societal and information threats broadly through the regulatory framework of "systemic risk" and cybersecurity. Systemic risk is defined as a risk specific to high-impact capabilities that can cause negative effects on public health, safety, public security, fundamental rights, or "the society as a whole".
- Regarding information security, the guidelines specify that providers must report "serious incidents," which the AI Office interprets to include cybersecurity breaches such as "cyberattacks" and the "(self-)exfiltration of model parameters". Additionally, the text notes that even free and open-source licenses may include specific terms to restrict usage in applications that would pose a significant risk to public safety, security, or fundamental rights.

***Quotes*:**

- "The AI Act defines a 'systemic risk' as 'a risk that is specific to the high-impact capabilities of general-purpose AI models, having a significant impact on the Union market due to their reach, or due to actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or the society as a whole, that can be propagated at scale across the value chain' (Article 3(65) AI Act)"37.
- "The AI Office considers that this obligation covers serious cybersecurity breaches related to the model or its physical infrastructure, including the (self-)exfiltration of model parameters and cyberattacks, due to their possible implications for the obligations provided for in Article 55(1), points (b) and (d), AI Act"38.
- "While the licence must generally permit use for any purpose, licensors may include specific, safety-oriented terms that reasonably restrict usage in applications or domains where such use would pose a significant risk to public safety, security, or fundamental rights"39.

8. Guidelines on the AI system definition. February 2025.[40]

The guidelines on the AI system definition explain the practical application of the legal concept, as anchored in the AI Act. The Commission publishes guidelines on AI system definition to facilitate the first AI

Act's rules application. By issuing guidelines on the AI system definition, the Commission aims to assist providers and other relevant persons in determining whether a software system constitutes an AI system to facilitate the effective application of the rules. The guidelines on the AI system definition are not binding. They are designed to evolve over time and will be updated as necessary, in particular in light of practical experiences, new questions and use cases that arise.

Relevance:

- The document acknowledges societal concerns by outlining the overarching aim of the AI Act, which is to ensure a high level of protection for "health, safety, and fundamental rights in the Union, including democracy and the rule of law". It also notes that the regulation follows a risk-based approach, where only systems posing "significant risks to fundamental rights and freedoms" are subject to the strictest regulatory obligations and prohibitions, though the guidelines do not elaborate on the specific nature of these societal risks.

Quotes:

- "Aim is to promote innovation in and the uptake of AI, while ensuring a high level of protection of health, safety, and fundamental rights in the Union, including democracy and the rule of law"41. (Guidelines on the AI system definition, paragraph 1)
- "The AI Act's risk-based approach means that only those systems giving rise to the most significant risks to fundamental rights and freedoms will be subject to its prohibitions laid down in Article 5 AI Act, its regulatory regime for high-risk AI systems covered by Article 6 AI Act [...]"[42]. (Guidelines on the AI system definition, paragraph 63)

9. General-Purpose AI Code of Practice. July 2025.[43]

The Code of Practice helps industry comply with the AI Act legal obligations on safety, transparency and copyright of general-purpose AI models. The General-Purpose AI (GPAI) Code of Practice is a voluntary tool, prepared by independent experts in a multi-stakeholder process, designed to help industry comply with the AI Act's obligations for providers of general-purpose AI models. The code was published on July 10, 2025. It is complemented by Commission guidelines on key concepts related to general-purpose AI models. The Commission and the AI Board have confirmed that the code is an adequate voluntary tool for providers of GPAI models to demonstrate compliance with the AI Act. Following the endorsement, AI model providers who voluntarily sign it can show they comply with the AI Act by adhering to the code. This will reduce their administrative burden and give them more legal certainty than if they proved compliance through other methods.

Relevance:

- The document explicitly addresses the areas of concern through the framework of "systemic risks." Regarding hybrid and information threats, the Code identifies "Chemical, biological, radiological and nuclear (CBRN)" capabilities and "Cyber offence" (enabling attacks on critical infrastructure) as specified systemic risks. It further warns against information threats such as the generation of "illegal, violent, hateful, radicalising, or false content," specifically noting the model propensity to "hallucinate" or produce "misinformation" that could obscure truth or undermine democratic processes.
- In terms of psychological and societal threats, the document lists "Harmful manipulation" as a specified risk, defining it as the "strategic distortion of human behaviour or beliefs" through persuasion, deception, or personalized targeting that exploits individuals who cannot detect such influence. Societal threats are broadly covered under risks to "society as a whole," "fundamental rights" (including non-discrimination), and "public mental health".

- To manage these threats, the Code requires providers to implement "safety mitigations" and conduct "adversarial testing" (red-teaming) to evaluate the model's effectiveness against these risks throughout its lifecycle.

Quotes:

- "Harmful manipulation: Risks from enabling the strategic distortion of human behaviour or beliefs by targeting large populations or high-stakes decision-makers through persuasion, deception, or personalised targeting. This includes significantly enhancing capabilities for persuasion, deception, and personalised targeting, particularly through multi-turn interactions and where individuals are unaware of or cannot reasonably detect such influence. Such capabilities could undermine democratic processes and fundamental rights"44.

10. First Draft Code of Practice on Transparency of AI-Generated Content (December 2025)[45]

This document operationalizes the fight against information manipulation by establishing technical standards to make synthetic content detectable.

Relevance:

- Taxonomy of Threat Vectors: The Code establishes a "common taxonomy" to classify the types of content often used in FIMI campaigns. It distinguishes between "Fully AI-generated content" (autonomous creation) and "AI-assisted content" (where AI alters meaning, emotional tone, or factual accuracy). This classification includes specific psychological threat vectors such as "AI-generated text that imitates the style of a specific person" (impersonation) and alterations that change "contextual meaning" or "emotional tone" to manipulate the audience.
- Targeting Political Influence: The Code specifically addresses text published to "inform the public on matters of public interest", defined to include "political or social issue texts intended to persuade". This confirms that the regulatory framework explicitly recognizes and targets the core material generated by troll farms to influence elections.
- Risks to Integrity: The document acknowledges that AI systems raise new risks of "misinformation and manipulation at scale, fraud, impersonation and consumer deception," making it increasingly difficult for humans to distinguish authentic content. It frames transparency as a necessary safeguard to "mitigate the risk of deception" and "uphold trust" in the information ecosystem.

1.3. Studies and in-depth policy reviews

1. Interplay between the AI Act and the EU digital legislative framework. Parliamentary Study. Policy Department for Transformation, Innovation and Health Directorate-General for Economy, Transformation and Industry. European Parliament, October 2025.[46]

This study explores how the AI Act relates to various crucial pieces of EU digital legislation, such as the GDPR, the Data Act and the Cyber Resilience Act. It assesses overlaps and gaps between these acts, and shows that, while each of them is individually well targeted, their interplay creates significant regulatory complexity. Finally, it also provides reflections and suggestions for possible evolutions of the AI Act, and of EU digital legislation as a whole, keeping in mind the objective of ensuring that Europe can establish a competitive AI industry. This study was prepared at the request of the ITRE Committee.

Relevance:

- The document addresses "hybrid" issues in terms of technological integration and cybersecurity resilience, which are critical components of countering hybrid threats. The study analyzes "hybrid systems" that combine AI components with non-AI digital elements, particularly in the context of the Cyber Resilience Act (CRA). It highlights the complexity of assessing risks in these systems, noting

that risks may arise in one component (e.g., AI training) but cause harm in another (e.g., downstream application), making "continuous risk management" essential rather than just one-off conformity assessments.

- The study notes that the AI Act requires specific cybersecurity measures to protect against AI-specific attacks, such as "data or model poisoning," which are methods often associated with hybrid interference campaigns.
- Generative AI Risks: The study highlights that while the AI Act regulates the generation of content (e.g., transparency and labeling), the Digital Services Act (DSA) focuses on the moderation of that content once hosted. It recommends harmonizing these regimes to prevent "dispersed enforcement" where content generators and hosts make different risk assessments.
- Manipulative and Subliminal Techniques: The AI Act classifies as "unacceptable risk" and bans AI practices that deploy "subliminal, manipulative, or deceptive techniques to distort behaviour and impair informed decision-making, causing significant harm".
- Exploitation of Vulnerabilities: It is prohibited to exploit vulnerabilities related to age, disability, or socio-economic circumstances to distort behavior in a way that causes significant harm.
- Emotion Recognition: The use of AI systems to infer emotions in workplaces or educational institutions is banned, except for specific medical or safety reasons.
- Social Scoring and Biometrics: The AI Act prohibits "social scoring" (evaluating individuals based on social behavior or traits) and "biometric categorisation" systems that infer sensitive attributes like race, political opinions, or sexual orientation.
- Deployers of high-risk AI systems (e.g., public bodies) must conduct Fundamental Rights Impact Assessments (FRIA) to assess risks to the fundamental rights of individuals before deployment. This creates a higher compliance threshold for protecting society than standard product safety laws.
- The study discusses the tension between the GDPR and the AI Act regarding bias monitoring. The AI Act allows the processing of special categories of personal data (e.g., ethnicity) specifically for "bias monitoring, detection and correction" to prevent societal discrimination, creating a complex interplay with GDPR restrictions.

Quote:

- "Obligations tied to the detection and disclosure of AI-generated or manipulated content are another interaction expressly noted in the AI Act. The DSA requires VLOPs (Very Large Online Platforms) and VLOSEs (Very Large Online Search Engines) to mitigate systemic risks related to such content where it “resembles existing persons, objects, places or other entities or events and falsely appears to a person to be authentic or truthful”, e.g., by placing prominent markings on it. The AI Act’s Article 50, on the other hand, obliges providers of generative AI systems to ensure that AI-generated or manipulated content is marked as such."47 (Interplay between the AI Act and the EU digital legislative framework, p. 50)

1.4. European Court of Justice (CJEU/ECJ) Cases

1. La Quadrature du Net (Joined Cases C-511/18, C-512/18, etc.) (2020)[48]

Relevance:

- The Court ruled that Member States cannot cite "national security" to issue blanket mandates for mass data retention. This is crucial for the AI Act's dual-use scope: it prevents governments from easily labeling a civilian AI surveillance tool as "national security" just to bypass EU regulations.
- The Court of Justice of the European Union (CJEU) ruling in Joined Cases C-511/18, C-512/18, and C-520/18 (La Quadrature du Net and Others) reaffirms that EU law-specifically the ePrivacy Directive (2002/58) read in light of the Charter of Fundamental Rights-precludes national legislation that mandates the "general and indiscriminate" retention of traffic and location data for the purpose of

combating crime. However, the Court establishes a critical exception for national security: when a Member State faces a "genuine and present or foreseeable" serious threat to its national security, it may temporarily order the general retention of such data. This preventive measure must be strictly limited in time and subject to effective review by a court or an independent administrative body to prevent abuse.

- Regarding implementation and dual-use scenarios, the ruling clarifies that even in the context of national security, data retention cannot be systematic or continuous. The Court allows for the automated analysis and real-time collection of data only when a serious threat is shown to exist, provided that the criteria for analysis are specific, reliable, and non-discriminatory-explicitly prohibiting criteria based on sensitive attributes like racial origin, political opinions, or religious beliefs. Furthermore, the judgment underscores that while combating serious crime and safeguarding public security are vital objectives, they do not justify the same level of interference as national security; for these purposes, only targeted retention (based on specific geographical areas or groups of persons) or the "expedited retention" (quick-freeze) of data is permissible under EU privacy and GDPR standards.

2. Ligue des droits humains (Case C-817/19) (2022)[49]

Relevance:

- Concerned the processing of PNR (passenger name record) data. The Court held that automated processing systems must have reliable, non-discriminatory criteria and human review. This sets the judicial standard for the "Human Oversight" requirements in the AI Act.
- The CJEU ruling in Case C-817/19 (Ligue des droits humains), delivered on June 21, 2022, addresses the validity and interpretation of the Passenger Name Record (PNR) Directive (2016/681) in light of the Charter of Fundamental Rights. The Court upheld the Directive but imposed strict limitations on its implementation to ensure that the processing of passenger data is limited to what is "strictly necessary" for combating terrorism and serious crime. A central provision is the prohibition of the general and indiscriminate collection of PNR data for intra-EU flights; such measures are only permissible if a Member State establishes a "genuine and present or foreseeable" terrorist threat. In the absence of such a threat, PNR processing for domestic flights must be restricted to specific routes or airports based on objective risk criteria.
- Regarding national security and data protection, the ruling clarifies that the PNR regime cannot be extended to ordinary crime and must adhere to high standards of personal data protection and the GDPR. The Court explicitly forbids the use of artificial intelligence (AI) in the form of self-learning systems ("machine learning") for the automated advance assessment of passengers, as the opacity of such systems could deprive individuals of their right to an effective judicial remedy. Furthermore, the judgment mandates that any automated match must be subject to an individual human review before any action is taken. Passenger data must generally be deleted after six months unless a concrete connection to a security risk is established, ensuring that the surveillance regime does not evolve into a "digital panopticon" without proper oversight.

3. VD and SR. (Joined Cases C-339/20, C-397/20) (2020)[50]

Relevance:

- The Court of Justice of the European Union (CJEU) ruling in Joined Cases C-339/20 and C-397/20 (VD and SR), delivered on September 20, 2022, confirms that EU law-specifically the ePrivacy Directive (2002/58) read in light of the Charter of Fundamental Rights-precludes national legislation that mandates the general and indiscriminate retention of traffic and location data for the purpose of combating market abuse, such as insider dealing. The case arose after the French Financial Markets Authority (AMF) prosecuted two individuals using telephone call records obtained under French law, which required operators to retain such data for one year. The Court reaffirmed its established

jurisprudence that such broad retention constitutes a serious interference with the rights to privacy and personal data protection, even when the objective is to investigate serious economic crimes.

- Regarding the legal implications and national security, the judgment clarifies that a national court cannot restrict the temporal effects of a declaration of invalidity regarding such legislation. While the Court acknowledges that the Market Abuse Regulation and Directive grant authorities the power to access existing records, these powers must be interpreted in accordance with the GDPR and the ePrivacy Directive, which limit data retention to what is strictly necessary. The ruling highlights that general and indiscriminate retention is permissible only in the event of a serious and current threat to national security. In contrast, for the investigation of serious crime or market abuse, only targeted retention (based on specific categories of persons or geographical areas) or the "quick freeze" (expedited retention) of specific data is considered proportionate and lawful under Union law.

4. SCHUFA Holding AG (Case C-634/21) (Dec 2023)[51]

Relevance:

- Ruled that a credit score generated by an algorithm constitutes an "automated decision" under GDPR if it decisively influences the outcome. This prevents AI providers from hiding behind the claim that their tool is "just a recommendation," forcing dual-use tools (e.g., risk assessment software) to comply withstrict transparency rules.
- The Court of Justice of the European Union (CJEU) ruling in SCHUFA Holding AG (Case C-634/21), delivered on December 7, 2023, establishes that the automated generation of a credit score by a credit reference agency (CRA) constitutes "automated individual decision-making" under Article 22(1) of the GDPR if a third party (such as a bank) relies heavily on that score to make a final decision. The Court rejected SCHUFA's argument that it only provides "preparatory acts" and that the actual decision is made by the lender; instead, it found that when a poor score leads "in almost all cases" to the refusal of a loan, the establishment of the score itself is the de facto decision. This interpretation aims to prevent a "legal gap" where neither the CRA nor the lender would be fully responsible for explaining the logic behind the automated process to the data subject.
- The ruling has significant implications for AI security and ethics, as it clarifies that any organization providing automated risk-based scores-including those for identity verification or fraud detection-may be subject to the strict requirements of Article 22. Under this provision, such automated decisions are generally prohibited unless they are necessary for a contract, authorized by law, or based on explicit consent, and they must always be accompanied by safeguards such as the right to human intervention and the right to an explanation. Additionally, in the joined cases C-26/22 and C-64/22, the Court ruled that private agencies like SCHUFA cannot lawfully retain data from public insolvency registers (such as a "discharge from remaining debt") for longer than the six-month period provided for in the public register, as doing so constitutes an unjustifiable interference with the data subject's right to be forgotten and their ability to participate in economic life.

5. Dun & Bradstreet Austria GmbH (Case C-203/22) (Feb 2025)[52]

Relevance:

- Established that individuals have a right to know the "logic involved" in an automated decision, and that trade secrets cannot be used as a blanket refusal to explain how an AI system works. This directly impacts providers of high-risk dual-use AI who might try to hide their algorithms.
- The CJEU ruling in Dun & Bradstreet Austria GmbH (Case C-203/22), delivered on February 27, 2025, significantly bolsters the "right to an explanation" under Article 15(1)(h) of the GDPR in the context of automated individual decision-making. The Court clarified that when an individual is subjected to an automated decision with legal or significant effects (such as the refusal of a mobile phone contract based on a credit score), they are entitled to receive "meaningful information about the logic

involved". This means the data controller must provide a sufficiently detailed and intelligible explanation of the procedures and principles applied, enabling the person to understand how their personal data influenced the result and to challenge its accuracy.

- A key provision of the ruling addresses the tension between transparency and proprietary interests: the Court held that trade secrets and intellectual property rights cannot be used as a blanket defense to withhold information from the data subject. If a company claims that full disclosure would compromise its "secret sauce" (e.g., a proprietary algorithm), it must still submit the allegedly protected information to a court or supervisory authority. This body must then conduct a case-by-case balancing exercise to determine the extent of the access right, ensuring that the protection of business secrets does not effectively nullify the individual's fundamental right to verify the lawfulness of their data processing. While the Court noted that disclosing the complex algorithm itself may not be necessary-or even helpful-to the consumer, it mandated that the explanation must at least set out the "key parameters" and the extent to which variations in the data would have led to a different outcome.

6. Like Company v. Google Ireland Ltd. (Case C-250/25) (Pending, 2025)[53]

Relevance:

- The first referral specifically addressing Generative AI and Copyright (training data scraping). While primarily civil, the ruling will determine the liability of General Purpose AI models, which are inherently dual-use.
- The pending case Like Company v. Google Ireland Ltd. (Case C-250/25), referred by the Budapest Metropolitan Court in April 2025, represents the first preliminary ruling request to the CJEU specifically addressing the intersection of generative AI and EU copyright law. The dispute centers on whether Google's chatbot, Gemini, infringed the rights of the Hungarian news publisher Like Company by generating detailed summaries of its copyright-protected articles without authorization. A central provision at issue is Article 15 of the CDSM Directive (2019/790), which grants press publishers the right to control online uses of their content beyond "very short extracts". The Court is asked to determine if displaying AI-generated summaries that mirror original content constitutes a "communication to the public" and whether the act of training a Large Language Model (LLM)-which involves tokenization and pattern recognition-amounts to an act of reproduction under Article 2 of the InfoSoc Directive.
- The ruling's implications for AI security and ethics are profound, as it will clarify whether commercial LLM training on publicly available content falls under the Text and Data Mining (TDM) exception (Article 4 of the CDSM Directive) or whether such use is invalidated by its commercial nature and the economic harm caused to original rightsholders. If the CJEU rules in favor of the publisher, AI developers may be forced to obtain licences for both training and deployment in Europe, significantly increasing operational costs and potentially reshaping the development of dual-use AI technologies. Conversely, a ruling for Google could expand the scope of the TDM exception, allowing for broader "transformative" uses of data without compensation. While this case is primarily a copyright matter, it reflects the broader EU priority of ensuring that technological progress remains anchored in fundamental rights and a fair balance between innovation and the protection of original creative works.

7. C-413/23 P (EDPS v SRB). Pseudonymization and AI Model Training (GDPR). September 2025.[54]

The CJEU clarified that pseudonymized data is not automatically considered personal data for a recipient if they cannot reasonably re-identify the individuals, taking into account all available means. This offers a more nuanced, context-specific view on data use that is important for AI model development and data sharing within the EU.

Relevance:

- ***Legal Context.*** The dispute stems from a decision by the EDPS finding that the SRB failed to fulfill its obligations relating to the processing of personal data. This failure occurred during a procedure for granting compensation to shareholders and creditors of a banking institution following that institution's resolution.
- ***Key Legal Concepts***. The judgment focuses on the interpretation of Regulation (EU) 2018/1725, specifically addressing the Concept of 'personal data' (Article 3(1)), the Concept of 'pseudonymisation' (Article 3(6)), and the obligation to inform the data subject (Article 15(1)(d)) concerning the transmission of pseudonymised data to a third party.

8. C-807/21 (Deutsche Wohnen). GDPR Liability and Cyber-Fines. December 2023.[55]

The CJEU ruled that administrative fines under the GDPR can only be imposed if the infringement is due to negligent or willful conduct. It rejects the concept of strict liability for GDPR violations, which is highly relevant for assessing liability and financial risk in the event of a cybersecurity breach.

Relevance:

- The core issue addressed by the Court concerned the interpretation of provisions of Regulation (EU) 2016/679 (General Data Protection Regulation or GDPR). Specifically, the ruling focused on the conditions for imposing administrative fines on a legal person, analyzing the concepts of 'controller' (Article 4(7)), the powers of supervisory authorities (Article 58(2)), and the imposition of administrative fines (Article 83).
- Key Holdings (Operative Part). The Court interpreted Article 58(2)(i) and Article 83 of Regulation 2016/679, concluding two main points regardingadministrative fines levied against controllers that are legal persons and undertakings:
  - Requirement of Identifying a Natural Person: Article 58(2)(i) and Article 83(1) to (6) of the GDPR must be interpreted as precluding national legislation which permits an administrative fine to be imposed on a legal person (in its capacity as controller) for an infringement (referred to in Article 83(4) to (6)) only if that infringement has previously been attributed to an identified natural person.
  - Requirement of Intent or Negligence: Article 83 of Regulation 2016/679 must be interpreted as meaning that an administrative fine may be imposed pursuant to that provision only where it is established that the controller (which is both a legal person and an undertaking) intentionally or negligently committed the infringement referred to in Article 83(4) to (6).

9. C-446/21 (Meta Platforms Ireland). Data Minimization and Profiling (GDPR). October 2024.[56]

The Court reinforced the principle of data minimization, ruling that the GDPR limits a social network's use of a user's personal data collected outside that platform for targeted advertising. This is crucial for defining the ethical and legal boundaries of data-driven profiling and AI-assisted marketing practices.

Relevance:

- The principle of data minimization (Article 5(1)(c) GDPR) requires data to be adequate, relevant, and limited to what is necessary for the processingpurposes. This principle reflects the principle of proportionality.
- The principle of storage limitation (Article 5(1)(e) GDPR) requires data to be kept only for as long as is necessary for the purposes for which they were collected or further processed.
- The controller must refrain from collecting data in a generalised and indiscriminate manner and must avoid collecting data that are not strictly necessary for the purpose of the processing.
- The extensive processing of user data, monitoring a large part of online activities, is characterized by a serious interference with fundamental rights (Articles 7 and 8 of the Charter).

- The indiscriminate use of all of the personal data held by a social network platform for advertising purposes, irrespective of the data's sensitivity, does not appear to be a proportionate interference with users' rights.
- The storage of the personal data of users for an unlimited period for the purpose of targeted advertising must be considered to be a disproportionate interference.
- The Court concluded that the data minimization principle precludes any personal data collected by the operator (from the data subject or third parties, on or outside the platform) from being aggregated, analyzed, and processed for targeted advertising without restriction as to time and without distinction as to type of data.

10. C-200/23 (Agentsia po vpisvaniyata v OL). Non-Material Damage and Loss of Control (GDPR Art. 82). 2024.[57]

The Court ruled that a loss of control over personal data for a limited period (in this case, data made publicly available online) may be sufficient to constitute non-material damage. This lowers the threshold for claims under Article 82, which is highly relevant for the liability assessment of cybersecurity incidents and data breaches.

Relevance:

- Key Interpretations and Holdings. The Court provided interpretation on several key points concerning data protection rights when documents containingpersonal data are publicly disclosed in a commercial register:
- Disclosure Obligation (Directive 2017/1132). Article 21(2) of Directive 2017/1132 (concerning company law) does not impose an obligation on a Member State to permit the disclosure in the commercial register of a company's constitutive instrument containing personal data other than the minimum personal data required by law. This provision primarily concerns the voluntary disclosure of translations, not the content of the data itself.
- Status of the Register Authority (GDPR Controller/Recipient). The authority responsible for maintaining the commercial register (the Agency) is considered both a 'recipient' and a 'controller' of the personal data contained in the disclosed constitutive instrument. This classification applies particularly in so far as the authority makes the data available to the public, even where that instrument contains personal data not required by Directive 2017/1132 or by national law.
- Right to Erasure (Article 17 GDPR):
    - The processing of personal data not required by Directive 2017/1132 or national law is unlikely to satisfy the conditions for lawfulness under Article 6(1)(c) or (e) of the GDPR (legal obligation or public interest task), as such processing goes beyond what is necessary to achieve the objectives of the directive.
    - National legislation or practice that leads the commercial register authority to refuse any request for erasure of non-required personal data-merely because a redacted copy of the constitutive instrument was not previously provided to the authority-is precluded by Directive 2017/1132 (Article 16) and Article 17 of the GDPR.
- Handwritten Signature as Personal Data (Article 4(1) GDPR). The Court ruled that the handwritten signature of a natural person is covered by the concept of 'personal data' within the meaning of Article 4(1) of the GDPR, as it identifies the person and is linked to the documents it authenticates.
- Non-Material Damage (Article 82(1) GDPR):
    - A loss of control, for a limited period, by the data subject over their personal data, due to those data being made available online in the commercial register, may suffice to cause 'non-material damage' under Article 82(1) of the GDPR.
    - However, the data subject must demonstrate that they have actually suffered such damage, however minimal. This concept does not require the existence of additional tangible adverse consequences.

- Exemption from Liability (Article 82(3) GDPR). An opinion issued by a supervisory authority under Article 58(3)(b) of the GDPR (which is advisory and not legally binding) is not sufficient to exempt the controller (the register authority) from liability under Article 82(2) and (3) of the GDPR. The controller's liability is fault-based, and to be exempt, the controller must prove that it is in no way responsible for the event giving rise to the damage.

1.5. National sources of the Republic of Lithuania

Overview of the national policy practices of the Republic of Lithuania related to cybersecurity, data protection (GDPR) and artificial intelligence (AI Act) allows us to assess national progress in the context of this project and to identify potential shortcomings in political practice.

-

1. National Cybersecurity Status Report 2024[58] (Nacionalinė kibernetinio saugumo būklės ataskaita 2024).

The 2024 National Cybersecurity Status Report, submitted by the National Cybersecurity Center (NKSC) together with other institutions, shows a significant increase in the number of cyber incidents – a 63% increase per year. The report reveals that the largest share of incidents (59%) is made up of attacks based on social engineering. In addition, it is emphasized that the increasing threat arises from espionage carried out by groups supported by foreign countries. The report also emphasizes the importance of preventive measures, mentioning tools such as the malicious domain blocking tool “Vasaris” and the activities of distance learning platforms.

Relevance:

- The report identifies hybrid and information warfare as central elements of the current security landscape, driven primarily by hostile regimes in Russia and Belarus aimed at destabilizing the state. These actors employ hybrid attacks, espionage, and propaganda not only to damage critical infrastructure but also to undermine Lithuania's strategic interests, specifically its support for Ukraine and NATO integration.
- The document highlights the use of advanced technologies, including artificial intelligence and deepfakes, to generate disinformation that portrays Lithuania as a "failed state" or "Russophobic," thereby attempting to discredit national defense efforts and question the reliability of allied protection.
- Psychological and societal threats are inextricably linked to these operations, which are designed to polarize society, erode trust in democratic institutions, and instill a pervasive sense of insecurity or fear regarding potential global conflict.
- The report notes that hostile propaganda aims to break the public's will to resist by convincing the domestic audience that the country is not worth defending. Furthermore, a significant majority of cyber incidents (59%) rely on social engineering, exploiting human psychology and trust to manipulate individuals, while information attacks increasingly feature intimidation tactics, such as narratives about nuclear war.

Quote:

- "The main goal of the messages sent by Russia and Belarus about Lithuania and NATO is to encourage distrust in one's state, its institutions and one another, [and] to polarize society. This attempts to influence the stability of Western democratic states and the ability of citizens to make decisions"59. (National Cybersecurity Status Report 2024)

2. Report on the national cybersecurity exercise “Cyber Shield OpEx 2024” (Nacionalinių kibernetinio saugumo pratybų „Kibernetinis skydas OpEx2024“ ataskaita)[60].

The National Cyber Security Centre under the Ministry of National Defence (hereinafter referred to as the NKSC) organised the annual national cyber security exercise "Cyber Shield OpEx 2024" on 7-18 October 2024. The exercise aimed to develop practical cybersecurity skills of the exercise participants, test cyber incident management procedures, improve cooperation between institutions managing and/or investigating cyber incidents and cyber security entities, and develop public communication skills.

80 organisations providing critical services to the population of Lithuania successfully participated in the exercise. Most (65) of these organisationsmanaged incidents in a virtual cyber training ground environment, while 39 organisations actively improved their public communication skills. The lessons identified in the exercise have been recorded and will be taken into account when organising future cyber security exercises.

***Relevance*:**

- The exercise scenario modeled hybrid and information threats by depicting "hostile states" supporting "hacktivist groups" that launched attacks in retaliation for Lithuanian legislative decisions. These simulated attacks included "content distortion" on organizational websites aimed at causing confusion among employees and clients, as well as Distributed Denial of Service (DDoS) attacks to disrupt critical services provided to the population.
- Psychological and societal pressure tactics were integrated into the simulation through "public pressure" campaigns on social media designed to force ransom payments and the distribution of phishing emails to target citizens. The exercise explicitly focused on testing "public communication" and "media" responses, evaluating how organizations manage reputational damage and maintain public trust while under pressure from these simulated hostile information operations.

3. Resolution of the Seimas of the Republic of Lithuania "On the Principles of the Use of Artificial Intelligence Technologies in the Public Sector" (Lietuvos Respublikos Seimo rezoliucija „Dėl dirbtinio intelekto technologijų naudojimo viešajame sektoriuje principų"). TAR, 2024-05-16, Nr. 8947[61].

The resolution establishes principles and guidelines for the application of AI in the public sector, emphasizing transparency, openness, equality and ethics. This is an important political document defining AI ethical guidelines at the national level.

Relevance:

- Regarding societal concerns, the resolution addresses the potential risks of AI integration into public administration rather than external security threats. It establishes safeguards to protect society from technological misuse, mandating principles such as "human oversight" to control administrative decisions, "equality" to prevent discrimination based on attributes like race or political views, and "transparency" to ensure the public is informed about AI influence. The document emphasizes the "primacy of human interests," ensuring that AI serves as a tool for social good rather than a source of societal destabilization.

Quote:

- "Only the proper use of artificial intelligence tools can ensure the protection of all human rights and freedoms, the country's economic and national security interests"[62]. (Resolution of the Seimas of the Republic of Lithuania "On the Principles of the Use of Artificial Intelligence Technologies in the Public Sector")

4. Methodological information (guidelines, recommendations, etc.) of the State Data Protection Inspectorate (Valstybinės duomenų apsaugos inspekcijos metodinė informacija – gairės, rekomendacijos ir kt.)[63].

The State Data Protection Inspectorate (SDI) is preparing a wide range of methodological information, guidelines and recommendations to help data controllers (both public and private sectors) and data processors properly implement the requirements of the General Data Protection Regulation (GDPR) and the Law on the Legal Protection of Personal Data of the Republic of Lithuania (LPPD). The SDI methodological material acts as a consulting aid, helping organizations not only to avoid violations, but also to actively implement data protection already at the design stage (privacy by design).

Relevance:

- "Recommendation on Prevention of Cyber Incidents" addresses psychological threats through the lens of social engineering (phishing), describing it as a method that exploits "human qualities" and trust to deceive victims into revealing sensitive information or transferring money. Additionally, it covers ransomware, which functions as a form of psychological pressure and blackmail (šantažas), forcing victims to decide whether to pay criminals to recover encrypted data. The provided provisions focus on mitigation strategies for these specific tactical threats, such as recognizing fake websites, refusing to pay ransoms, and reporting incidents to the police or the National Cyber Security Centre.

Quote:

- "Cyber attacks based on social engineering principles are a method whereby the aim is to exploit human qualities and extract certain sensitive information, e.g., login data for social network accounts, information systems, bank accounts [...].[64]" (Recommendation on Prevention of Cyber Incidents).

1.6. National sources of selected EU Member States

Identifying "best policy practices" in the overlapping fields of cybersecurity, GDPR, and the AI Act involves looking at national governments' strategic planning, the issuance of proactive regulatory guidance, and the establishment of effective institutional frameworks.

Based on recent national documents and policy initiatives focusing on AI Act and NIS2 implementation, France, Germany, and the Netherlands exhibitleading practices, particularly in providing detailed guidance on the intersection of data protection and AI.

France

France's Data Protection Authority, the Commission Nationale de l'Informatique et des Libertés (CNIL), is a key benchmark for providing clear, actionable guidance on reconciling AI development with GDPR principles. This proactive approach offers legal clarity before the AI Act's high-risk systems obligations become fully applicable. CNIL published several recommendations to promote the responsible use of AI while ensuring compliance with personal data protection.

These recommendations confirm that GDPR requirements are sufficiently balanced to address the specific challenges of AI. They provide concrete and proportionate solutions to inform individuals and facilitate the exercise of their rights.

1. CNIL Recommendations on AI and GDPR Compliance[65].

Relevance:

- The recommendations take into account the EU Artificial Intelligence Act recently adopted. Indeed, where personal data is used for the development of an AI system, both the GDPR and the AI Act apply. CNIL's recommendations have therefore been drawn up to supplement them in a consistent manner regarding data protection.

Germany

Germany's strength lies in its long-term strategic integration of cybersecurity into its national AI and industrial policy, particularly emphasizing security-by-design for critical infrastructure:

2. Cyber security strategy for Germany. 2021[66].

The Cyber Security Strategy for Germany 2021 creates a strategic framework for Federal Government policy on cyber security for the next five years, subject to the availability of budgetary funds.

Relevance:

- Cyber Security Strategy sets out four crosscutting guiding principles:
- Establishing cyber security as a joint task for government, private industry, the research community and society.
- Reinforcing the digital sovereignty of government, private industry, the research community and society.
- Making digital transformation secure.
- Setting measurable, transparent objectives.
- The strategic objectives are implemented in three action areas:
    - Action Area 1 – "Remaining safe and autonomous in a digital environment";
    - Action Area 2 is "Government and private industry working together";
    - Action Area 3 – "Strong and sustainable cyber security architecture for every level of government".

3. The Federal Government's Artificial Intelligence Strategy and AI Action Plan (and related information, programmes, initiatives, etc.)[67].

The Federal Cabinet adopted the Federal Government's Artificial Intelligence Strategy (AI Strategy) in 2018. The AI Strategy is designed to strengthen Germany's position in the international competition for research, development and applications in AI. The AI Strategy is updated to take account of new developments and needs.

Relevance:

- Various different initiatives now focus on the issues of pandemic response, sustainability, environmental protection, climate action and national and international networking. In 2023 the AI Action Plan was launched as the update to the AI Strategy.

The Netherlands

The Netherlands is recognized for its proactive approach to regulating algorithmic risks through its Data Protection Authority (AutoriteitPersoonsgegevens or AP) and its clear vision for institutional cooperation on AI Act enforcement.

Every six months, the Dutch Data Protection Authority (AP) publishes its Report AI & Algorithms Netherlands (RAN). It also draws observations and analyses, based on the knowledge gathered through the network, conversations and consultations with experts and society.

4. AI & algorithmic risks: developments in the Netherlands (and related information, links, initiatives, programmes, etc.)[68].

5. AP and RDI: Supervision of AI systems requires cooperation and must be arranged quickly (and links to related information)[69].

6. Guidelines on AI Literacy[70].

Relevance:

- The use of generative AI raises legal concerns. One aspect of this is the training of models on large quantities of 'scraped' data, which can include personal data. Additionally, new risks for users and the increased concentration of power in big tech firms demand scrutiny.
- Many countries identify large scale spreading of disinformation and manipulation through the use of Generative AI as risks.
- The output of Generative AI is based on probability estimates. The user does often not have insight into uncertainty, plausible alternative answers, or references to origins of the content. This makes it hard to interpret the output and increases the risk of incorrect conclusions and actions. This can result in discrimination and arbitrariness in a person's or organisation's approach.
- Training a generative AI system with data containing unbalanced information of groups, possibly without realising it, may reinforce people's biases. After all, they will then be shown 'overrepresented' groups more often. For example, AI-generated images may be disproportionately more likely to show men as doctors and women as nurses.
- The development of generative foundation models takes place at a small number of organisations. Mainly big tech firms have sufficient financial means to invest in this. This makes it hard for new providers to enter the market, and reinforces the existing power of big tech companies.

2. Keywords

In the area of the analysis of the recent policy practices within the scope of the AICP-FIMI project, several categories of keywords can be distinguished.

2.1. Operational Threats & Targets

- "Foreign Information Manipulation and Interference (FIMI)": identified as a sophisticated hybrid attack used to erode trust in democratic institutions.
- "Social Media Bots & Troll Farms": the primary targets for detection by AICP-FIMI platform to enhance resilience against cyber threats.
- "Hybrid Threats": a convergence of cyber operations, sabotage, and information warfare intended to destabilize society.
- "Deepfakes & AI-Generated Content": specific technical vectors used to manipulate information flows and distort public debate.
- "Social Engineering": a primary method for cyber incidents, exploiting human psychology and trust (e.g., phishing).
- "Hallucinations": a risk associated with Generative AI where models produce false information, complicating the information ecosystem.
- "Psychological Manipulation": tactics involving "Fear, Uncertainty, and Doubt" used to amplify disruptive impacts.

2.2. Legal & Regulatory Framework

- "AI Act": the central horizontal regulatory framework establishing a risk-based approach to AI safety and fundamental rights.
- "GDPR (General Data Protection Regulation)": the foundational data protection law that interacts complexly with AI development, particularly regardingdata scraping and profiling.

- "Cyber Resilience Act (CRA)": mandates cybersecurity requirements for products with digital elements (hardware/software), applicable to your cloud platform.
- "Cyber Solidarity Act": establishes a European cybersecurity alert system that such platform as AICP-FIMI could potentially integrate with.
- "NIS2 Directive": focuses on the security of network and information systems for critical sectors, including public administration.
- "Digital Services Act (DSA)": regulates online platforms and provides mechanisms for researchers to access data necessary for analyzing bot networks.

2.3. Compliance & Governance Concepts

- "High-Risk AI Systems": classification for AI used in sensitive areas like elections, requiring strict compliance obligations.
- "Prohibited AI Practices": banned activities such as biometric categorization of political opinions, untargeted facial recognition scraping, and subliminal manipulation.
- "Trustworthy AI": the EU's strategic goal to ensure AI is lawful, ethical, and robust.
- "Transparency Obligations": requirements to mark and label AI-generated content to enable detection.
- "Fundamental Rights Impact Assessment (FRIA)": a mandatory assessment for deployers of high-risk AI systems to evaluate risks to democracy and the rule of law.
- "Human Oversight": a required safeguard to ensure AI systems remain under human control and do not undermine agency.

2.4. Technical & Operational Mechanisms

- "Data Minimization": a GDPR principle requiring that data collection be limited to what is strictly necessary, preventing indiscriminate retention.
- "Security-by-Design": the requirement to integrate security measures (like vulnerability handling) from the earliest design phase of the platform.
- "Biometric Categorization": the use of AI to infer sensitive attributes (e.g., political opinions) from biometric data, which is explicitly prohibited.
- "Web Scraping": the collection of data from the internet; "untargeted scraping" for facial recognition is prohibited.
- "Rapid Response System": a mechanism for reporting time-sensitive threats to election integrity.
- "Automated Decision-Making": Processes that legally require explainability and transparency regarding the "logic involved".

2.5. Strategic Initiatives & Institutions

- "European Democracy Shield": a package of measures to protect elections, media, and civil society from foreign interference.
- "Whole-of-Society Approach": a strategy linking civilian, military, and private sectors to build collective resilience.
- "Strategic Compass": the EU's plan for strengthening security and defense, including countering hybrid threats.
- "AI Office & AI Act Service Desk": new support structures established to assist with regulatory compliance.
- "ENISA (European Union Agency for Cybersecurity)": the agency providing technical guidelines and threat landscape reports.

Conclusions

General Conclusions

1. Strategic Alignment and Urgency

- The relevant movements in the recent policy practices show that Foreign Information Manipulation and Interference (FIMI) is a critical security threat, characterized as a sophisticated "hybrid attack" used by authoritarian regimes (specifically Russia and Belarus) to erode trust in democratic institutions.
- The EU has explicitly called for tools to trace "coordinated inauthentic behaviour" and "bot-driven amplification," validating the direct operational need for your platform. The launch of initiatives like the European Democracy Shield and the Security Action for Europe (SAFE) indicates a favorable environment for securing funding and institutional support.

2. High-Risk Regulatory Classification

- AICP-FIMI platform will operate in a strictly regulated environment. Because it interacts with electoral processes and democratic integrity, it will likely beclassified as a High-Risk AI System under the AI Act. This classification triggers mandatory obligations regarding fundamental rights impact assessments, human oversight, and strict data governance. Furthermore, the platform falls under the scope of the Cyber Resilience Act (CRA), meaning it must meet specific cybersecurity requirements for hardware and software products.

3. The "Prohibited Practices" Minefield

- The project faces significant legal challenges regarding automated collection of information. The sources explicitly warn against "untargeted scraping of facial images" to build databases and the "biometric categorization" of individuals to infer political opinions. Using AI to classify users as "politically aligned" with foreign regimes based on biometric data is prohibited. Additionally, the CJEU case law (La Quadrature du Net) prohibits general and indiscriminate data retention without a specific, present threat to national security.

4. The Transparency Imperative

- Recent CJEU rulings (SCHUFA and Dun & Bradstreet) have established that if an automated system makes a decision that significantly affects an individual (e.g., labeling them a "bot" or "troll," potentially leading to censorship), the provider must be able to explain the "logic involved".

Recommendations

- Integrate the project with EU Ecosystems, e.g., design the platform to feed into the Rapid Response System used during elections and potentially the Cyber Solidarity Act’s alert system. Position the platform as a tool for the FIMI Information Sharing and Analysis Centre (FIMI ISAC) to reinforce civil society capabilities.
- To avoid falling under the AI Act's ban on biometrically inferring political opinions, detection models should focus on technical behavioral patterns (e.g., posting frequency, cross-platform coordination, automated scripting) rather than analyzing the political content or biometric data of user profiles.
- Instead of web scraping (which carries legal risks under GDPR and AI Act), project should utilize the research access mechanisms provided by the Digital Services Act (DSA). This is the primary legal

pathway mentioned for researchers to access Very Large Online Platform (VLOP) data for analyzing systemic risks.

- To comply with CJEU rulings (Ligue des droits humains) and High-Risk AI requirements, it should be ensured that any automated flagging of accounts as "bots" is subject to human review before significant action is taken.
- The developed AI models should be interpretable. If the platform flags a user as a bot, it must be able to generate a "meaningful information" report explaining why (e.g., "User posted 500 times in one hour"), complying with the Dun & Bradstreet ruling.
- As required by the Cyber Resilience Act and NIS2, integrated security-by-design measures (vulnerability handling, encryption) should be taken into account from the very first design phase. The platform itself will be a target for state-sponsored cyberattacks.
- The platform should be equipped with possibilities to detect the specific transparency markers mandated by the Code of Practice on Transparency of AI-Generated Content (e.g., C2PA metadata, imperceptible watermarks). The platform should verify if social media content complies with the mandatory "AI generated" labeling.
- The collection of facial images from the internet for the purpose of identification should be excluded as this is a Prohibited AI Practice (Article 5(1)(e)).

References

1. European Commission Coordinated Plan on Artificial Intelligence 2021 Review. https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review
2. Report of the High Representative of the Union for Foreign Affairs and Security Policy to the Council - "Annual Progress Report on the Implementationof the Strategic Compass for Security and Defence (2025). https://data.consilium.europa.eu/doc/document/ST-8023-2025-INIT/en/pdf
3. European Commission. European approach to artificial intelligence, policy review (2025). https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence
4. The AI Continent Action Plan. https://digital-strategy.ec.europa.eu/en/library/ai-continent-action-plan
5. Apply AI Strategy (Communication from the Commission Artificial Intelligence for Europe) COM(2025) 723 final. https://eur-lex.europa.eu/resource.html?uri=cellar:194ae542-a421-11f0-97c8-01aa75ed71a1.0001.02/DOC_1&format=PDF
6. Progress Report on the Digital Rulebook Implementation. Commission Policy. European Commission. September 2025. https://commission.europa.eu/document/download/68237940-a9e3-460c-bf49-932d432a6bba_en?filename=VIRKKUNEN_APR_SPI_2025_70_EN.pdf
7. European Democracy Shield: Empowering Strong and Resilient Democracies. JOIN(2025) 791 final. https://commission.europa.eu/document/download/2539eb53-9485-4199-bfdc-97166893ff45_en?filename=JUST_template_comingsoon_standard_1.pdf
8. European Parliament Motion for a Resolution on Hybrid Provocations. European Parliament. October 2025. https://www.europarl.europa.eu/doceo/document/B-10-2025-0437_EN.pdf
9. Eighth progress report on the implementation of the 2016 Joint framework on countering hybrid threats and the 2018 Joint communication on increasing resilience and bolstering capabilities to address hybrid threats. Joint staff working document. European Commission. October 2024. https://defence-industry-space.ec.europa.eu/system/files/2025-01/SWD_Annual-Progress-Report-2024.PDF
10. Union Rolling Work Programme for European cybersecurity certification. Policy and Legislation. February 2024. https://ec.europa.eu/newsroom/dae/redirection/document/102292
11. Council Conclusions on the Future of Cybersecurity. Council of the EU. May 2024. https://data.consilium.europa.eu/doc/document/ST-10133-2024-INIT/en/pdf
12. ENISA NIS360 2024 report: A comprehensive look at cybersecurity maturity and criticality of NIS2 sectors. March 2025. https://www.enisa.europa.eu/news/enisa-nis360-2024-report

13. ENISA Threat Landscape (ETL). October 2025. https://www.enisa.europa.eu/sites/default/files/2025-10/ENISA%20Threat%20Landscape%202025_0.pdf
14. 2024 Report on the State of Cybersecurity in the Union. December 2024. https://www.enisa.europa.eu/sites/default/files/2024-11/2024%20Report%20on%20the%20State%20of%20the%20Cybersecurity%20in%20the%20Union.pdf
15. Guidelines on prohibited artificial intelligence (AI) practices. February 2025. https://ec.europa.eu/newsroom/dae/redirection/document/112367
16. Multilayer Framework for Good Cybersecurity Practices for AI. ENISA Recommendation. June 2023. https://www.enisa.europa.eu/sites/default/files/publications/Multilayer%20Framework%20for%20Good%20Cybersecurity%20Practices%20for%20AI.pdf
17. Interplay between the AI Act and the EU digital legislative framework. Parliamentary Study. Policy Department for Transformation, Innovation and Health Directorate-General for Economy, Transformation and Industry. European Parliament, October 2025. https://www.europarl.europa.eu/RegData/etudes/STUD/2025/778575/ECTI_STU(2025)778575_EN.pdf
18. C-511/18 - La Quadrature du Net and Others. https://curia.europa.eu/juris/liste.jsf?language=en&num=C-511/18
19. C-817/19 - Ligue des droits humains. https://curia.europa.eu/juris/liste.jsf?num=C-817/19
20. Joined Cases C-339/20, C-397/20 - VD and SR. https://curia.europa.eu/juris/liste.jsf?num=C-339/20&language=en
21. Case C-634/21 - SCHUFA Holding AG. https://curia.europa.eu/juris/liste.jsf?num=C-634/21
22. Case C-203/22 - Dun & Bradstreet Austria GmbH. https://curia.europa.eu/juris/liste.jsf?num=C-203/22
23. Case C-250/25 - Like Company v. Google Ireland Ltd. https://curia.europa.eu/juris/liste.jsf?num=C-250/25
24. C-413/23 P (EDPS v SRB). Pseudonymization and AI Model Training (GDPR). September 2025. https://curia.europa.eu/juris/document/document.jsf?text=&docid=305584&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=8121207
25. C-807/21 (Deutsche Wohnen). GDPR Liability and Cyber-Fines. December 2023. https://curia.europa.eu/juris/document/document.jsf?text=&docid=282192&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=8123480
26. C-200/23 (Agentsia po vpisvaniyata v OL). Non-Material Damage and Loss of Control (GDPR Art. 82). 2024. https://curia.europa.eu/juris/document/document.jsf?text=&docid=290701&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=1164237
27. National Cybersecurity Status Report 2024 (Nacionalinė kibernetinio saugumo būklės ataskaita 2024). https://www.nksc.lt/doc/Nacionaline-kibernetinio-saugumo-ataskaita-2024.pdf
28. Report on the national cybersecurity exercise "Cyber Shield OpEx 2024" (Nacionalinių kibernetinio saugumo pratybų „Kibernetinis skydas OpEx 2024" ataskaita). https://www.nksc.lt/doc/KS2024_OPEX_Pratybu_ataskaita.pdf
29. Resolution of the Seimas of the Republic of Lithuania "On the Principles of the Use of Artificial Intelligence Technologies in the Public Sector" (Lietuvos Respublikos Seimo rezoliucija „Dėl dirbtinio intelekto technologijų naudojimo viešajame sektoriuje principų"). TAR, 2024-05-16, Nr. 8947. https://www.e-tar.lt/portal/lt/legalAct/611a93c0135e11efbcbfb318996800a8
30. Methodological information (guidelines, recommendations, etc.) of the State Data Protection Inspectorate (Valstybinės duomenų apsaugos inspekcijos metodinė informacija – gairės, rekomendacijos ir kt.). https://vdai.lrv.lt/lt/naudinga-informacija/rekomendacijos-gaires-ir-kt/work/
31. CNIL Recommendations on AI and GDPR Compliance. https://www.cnil.fr/en/ai-and-gdpr-cnil-publishes-new-recommendations-support-responsible-innovation and https://www.cnil.fr/en/ai-how-comply-regulations
32. Cyber security strategy for Germany. 2021. https://www.bmi.bund.de/SharedDocs/downloads/EN/themen/it-digital-policy/cyber-security-strategy-for-germany2021.pdf?__blob=publicationFile&v=4

33. The Federal Government's Artificial Intelligence Strategy and AI Action Plan (and related information, programmes, initiatives, etc.).https://www.bmftr.bund.de/EN/Research/EmergingTechnologies/ArtificialIntelligence/artificialintelligence_node.html andhttps://www.bmftr.bund.de/DE/Forschung/Schluesseltechnologien/KuenstlicheIntelligenz/KiAktionsplan/kiaktionsplan_node.html
34. AI & algorithmic risks: developments in the Netherlands (and related information, links, initiatives, programmes, etc.). https://www.autoriteitpersoonsgegevens.nl/en/themes/algorithms-ai/ai-algorithmic-risks-developments-in-the-netherlands
35. AP and RDI: Supervision of AI systems requires cooperation and must be arranged quickly (and links to related information). https://www.autoriteitpersoonsgegevens.nl/en/current/ap-and-rdi-supervision-of-ai-systems-requires-cooperation-and-must-be-arranged-quickly
36. Guidelines on AI Literacy. https://www.autoriteitpersoonsgegevens.nl/documenten/aan-de-slag-met-ai-geletterdheid
37. Fundamentals of Secure AI Systems with Personal Data. EDPB Guidance. April 2025. https://www.edpb.europa.eu/system/files/2025-06/spe-training-on-ai-and-data-protection-technical_en.pdf
38. Guidance for Risk Management of Artificial Intelligence systems. European Data Protection Supervisor. November 2025. https://www.edps.europa.eu/data-protection/our-work/publications/guidelines/2025-11-11-guidance-risk-management-artificial-intelligence-systems_en
39. Guidance on Generative AI, strengthening data protection in a rapidly changing digital era. European Data Protection Supervisor. October 2025. https://www.edps.europa.eu/data-protection/our-work/publications/guidelines/2025-10-28-guidance-generative-ai-strengthening-data-protection-rapidly-changing-digital-era_en
40. First EDPS Orientations for EUIs using Generative AI. European Data Protection Supervisor. June 2024. https://www.edps.europa.eu/data-protection/our-work/publications/guidelines/2024-06-03-first-edps-orientations-euis-using-generative-ai_en
41. Guidelines on the scope of obligations for providers of General-Purpose AI (GPAI) models. July 2025. https://digital-strategy.ec.europa.eu/en/library/guidelines-scope-obligations-providers-general-purpose-ai-models-under-ai-act
42. Guidelines on the AI system definition. February 2025. https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application
43. General-Purpose AI Code of Practice. July 2025. https://digital-strategy.ec.europa.eu/en/policies/contents-code-gpai
44. First Draft Code of Practice on Transparency of AI-Generated Content (December 2025). https://ec.europa.eu/newsroom/dae/redirection/document/123074

Annex 2
Case Studies

1. Hypothetical Case Study: Implementation of the „AICP-FIMI“ Project

1.1. Regulatory Environment Analysis

The AICP-FIMI project operates at the intersection of cybersecurity, data protection, and the regulation of artificial intelligence. The regulatory environment is characterized by a "risk-based approach" where the platform's potential impact on democratic processes necessitates strict compliance.

A. The AI Act[1] (High-Risk & Prohibited Practices) The AICP-FIMI platform will likely be classified as a High-Risk AI System because it operates in the sensitive area of elections and democratic processes. This classification imposes obligations to ensure systems are safe, transparent, and human-centric.

- **Prohibited Practices:** The project must strictly avoid practices defined as "unacceptable risks" under Article 5. Specifically, the untargeted scraping of facial images from the internet to create recognition databases is prohibited (Article 5(1)(e)). Furthermore, biometric categorization to infer political opinions is banned (Article 5(1)(g)), meaning the platform cannot analyze profile photos to deduce political alignment.
- **Transparency:** There is a duty to mark and enable the detection of AI-generated content (Article 50), which is relevant if the platform is used to analyze or flag such content.

B. General Data Protection Regulation (GDPR)[2] Compliance with GDPR is a prerequisite for AI development.

- **Lawfulness and Minimization:** Data processing must be limited to what is "strictly necessary". The "data minimization" principle precludes aggregating user data indiscriminately for profiling without restriction (CJEU *Meta Platforms*)[3].
- **Automated Decision-Making (Article 22):** If the platform's identification of a "bot" leads to significant effects (e.g., account suspension), it constitutes automated decision-making. Recent CJEU case law (*Schufa*[4], *Dun & Bradstreet*)[5] establishes that providers must explain the "logic involved" and cannot hide behind trade secrets.
- **Data Retention:** General and indiscriminate retention of data is prohibited unless there is a present, serious threat to national security (CJEU *La Quadrature du Net*)[6].

C. Cybersecurity and Digital Services Act (DSA)

- **Digital Services Act[7] (DSA):** This is the primary legal mechanism for accessing data from Very Large Online Platforms (VLOPs) for research into systemic risks, rather than unauthorized scraping.
- **Cyber Resilience Act[8] (CRA) & NIS2[9]:** The platform constitutes a product with digital elements and must meet security-by-design standards. As a tool aimed at public administration or critical infrastructure, it falls under the scope of NIS2, requiring high cybersecurity maturity.

1.2. Compliance Mapping

The following stages outline specific legal provisions (compliance mapping) relevant to the AICP-FIMI project:

Legal Act	Provision	Requirement for AICP-FIMI
AI Act	**Art. 6**	Classify as High-Risk AI; conduct Fundamental Rights Impact Assessment (FRIA).

AI Act	**Art. 5(1)(e)**	Prohibition: Do not use untargeted scraping of facial images for databases.
AI Act	**Art. 5(1)(g)**	Prohibition: Do not use biometric data to categorize political opinions.
AI Act	**Art. 14**	Implement Human Oversight measures to prevent automation bias.
GDPR	**Art. 35**	Conduct a Data Protection Impact Assessment (DPIA).
GDPR	**Art. 5(1)(c)**	Data Minimization: Collect only data strictly necessary for botdetection.
GDPR	**Art. 22 & 15**	Provide "meaningful information about the logic involved" in automated decisions (Right to Explanation).
DSA	**Art. 40**	Utilize the lawful data access mechanism for researchers/vetted organizations rather than web scraping.
CRA	**Security**	Implement security-by-design and vulnerability handling throughout the lifecycle.

1.3. Most Important Implementation Stages

Based on the project lifecycle and regulatory requirements, the four most critical stages are:

1. **Design and Impact Assessment Phase:** Establishing the legal basis and architectural safeguards.
2. **Data Acquisition and Processing Phase:** The collection of training and operational data.
3. **Model Training and Testing Phase:** ensuring robustness against bias and adversarial attacks.
4. **Operational Deployment and Oversight:** Human-in-the-loop operations and incident reporting.

1.4. Practical Steps at Each Stage

Stage 1: Design and Impact Assessment

- **Privacy by Design:** Integrate security measures from the earliest conceptual phase, not as an add-on.
- **Conduct Assessments:** Perform a Data Protection Impact Assessment (DPIA) and a Fundamental Rights Impact Assessment (FRIA)**.** These must assess risks to democracy, non-discrimination, and freedom of expression.
- **Define Logic:** Clearly define the behavioral markers for "bot" identification (e.g., posting frequency) rather than political content to avoid prohibited profiling.

Stage 2: Data Acquisition

- **Secure Access:** Do not scrape social media indiscriminately. Apply for access to data via the **Digital Services Act (DSA)** framework for researchers to analyze systemic risks.
- **Data Minimization:** Filter data at the point of collection. Exclude biometric data (profile photos) unless strictly necessary and legally justified; never use them for political categorization.
- **Pseudonymization:** Implement pseudonymization techniques. Be aware that pseudonymized data may still be "personal data" if re-identification is possible (CJEU *SRB* case[10]), so strict access controls are required.

Stage 3: Model Training and Testing

- **Adversarial Testing (Red-Teaming):** Conduct "red-teaming" to test the model against "data poisoning" and "prompt injection" attacks, which are common risks for AI systems.

- **Bias Mitigation:** Use special categories of data (e.g., political orientation data) only within the strict confines of bias monitoring and correction as allowed by the AI Act, to ensure the model does not unfairly target specific groups.
- **Transparency Markers:** Train the system to detect technical transparency markers (e.g., C2PA metadata) mandated by the Code of Practice on Transparency.

Stage 4: Operational Deployment and Oversight

- **Human Oversight (Human-in-the-Loop):** Implement a review process where a human verifies the AI's "bot" classification before any significant action (like reporting to a platform for takedown) is taken. This is required by CJEU case law (*Ligue des droits humains*)[11].
- **Explainability:** Ensure the system can generate a report explaining *why* an account was flagged (e.g., "posted 500 times in 1 hour"), satisfying the "right to explanation".
- **Incident Reporting:** Integrate with the Cyber Solidarity Act[12] ecosystem to report large-scale FIMI incidents or cyber threats detected by the platform.

1.5. Practical Challenges due to Legal Clarity/Specificity

A. The "Subliminal Techniques" Ambiguity The AI Act prohibits "subliminal techniques" that distort behavior. However, there is a lack of clarity on where legitimate political persuasion ends and "harmful manipulation" begins. The AICP-FIMI platform may struggle to programmatically distinguish between a sophisticated FIMI bot and a passionate human activist using persuasive language, creating a risk of false positives that could infringe on freedom of speech.

B. Data Access vs. Scraping While the DSA provides a pathway for data access, the practical implementation of this for third-party AI developers remains complex. Relying on scraping risks violating GDPR (lack of legal basis) and the AI Act (prohibited biometric scraping). The *Like Company v. Google Ireland*[13] case highlights the uncertainty regarding the liability of using scraped data for AI training under copyright and database laws.

C. The "National Security" Exception Member States often cite national security to justify broad data processing, but the CJEU (*La Quadrature du Net*) has ruled that "general and indiscriminate" data retention is only permitted during a "genuine and present" threat. AICP-FIMI developers face a challenge in defining data retention periods that are operationally useful for tracking long-term bot farms without violating this strict judicial standard.

D. Explainability of "Black Box" Models The CJEU (*Dun & Bradstreet*) requires revealing the "logic involved" in automated decisions, rejecting trade secret defenses. However, modern deep learning models (often used for bot detection) act as "black boxes" where the specific logic is hard to articulate. This creates a technical compliance gap: the law requires an explanation that the technology may not be easily capable of providing.

2. Real-World Case Analysis

2.1. Graphika

Type: Cybersocial Intelligence & Network Analysis. **Core Philosophy:** "Structural" detection -focusing on how actors behave and how networks are formed rather than just the content itself[14].

1. Regulatory & Compliance Context

- **High-Risk Classification:** As a provider of "takedown intelligence" to platforms and governments, Graphika operates in a high-stakes environment comparable to the "High-Risk" designation under the AI Act.

- **Data Privacy (GDPR):** Graphika's methodology involves mapping millions of social media accounts to identify "coordinated inauthentic behavior" (CIB). This raises GDPR challenges regarding the processing of personal data (even public profiles) for profiling purposes. The company mitigates this by focusing on *behavioral patterns* (Actor/Behavior/Content framework) rather than individual content moderation.
- **Government Ties:** Graphika has faced scrutiny for its contracts with the US Department of Defense and intelligence communities, which highlights the need for transparency to avoid accusations of acting as a state-proxy censor.

2. Technical Architecture & Stages

- **Stage 1: Design (ABC Framework):** Graphika utilizes the **ABC Framework** (Actors, Behavior, Content) to categorize threats. This moves analysis beyond simple "fake news" detection to identifying the *infrastructure* of an operation.
- **Stage 2: Data Acquisition (ATLAS):** The proprietary **ATLAS** platform ingests data from social media APIs. Unlike untargeted scraping, Graphika often works with data provided by platforms (e.g., Meta, Twitter/X) for forensic analysis of specific datasets.
- **Stage 3: Analysis (Cybersocial Mapping):** The core technology maps "nodes" (accounts) and "edges" (interactions) to visualize communities. For example, it identified the "Spamouflage" network by tracking how disparate accounts coordinated to amplify Chinese state messaging.

3. Practical Steps for FIMI Countering

- **Network Mapping:** Graphika maps the "cybersocial terrain" to detect anomalies, such as a sudden influx of accounts in a specific cluster.
- **Cross-Platform Tracking:** A key practical step is tracking actors as they migrate. For instance, Graphika tracked the "Harlan Report" persona across TikTok, X, and YouTube, identifying it as a "Spamouflage" asset masquerading as a U.S. conservative outlet.
- **Forensic Reporting:** The output is often deep-dive intelligence reports (e.g., "The Americans," "Bad Reputation") provided to platforms to justify the removal of networks.

4. Challenges

- **"Black Box" of Attribution:** Attributing a network to a state actor (e.g., establishing that "Spamouflage" is Chinese state-linked) is complex. Graphika relies on "high confidence" assessments based on behavioral fingerprints, but this lacks the absolute certainty of legal evidence.
- **AI-Destabilized Truth:** Graphika analysts have noted that the rise of generative AI "destabilizes the concept of truth itself," making it harder to establish a baseline of reality against which to measure manipulation.

2.2. Factmata (Acquired by Cision)

Type: Automated Narrative Monitoring & Content Scoring. **Core Philosophy:** "Narrative" detection-using NLP to cluster opinions and score content for toxicity and bias[15].

1. Regulatory & Compliance Context

- **Automated Decision Making (GDPR Art. 22):** Factmata's core product relies on AI scoring algorithms (e.g., toxicity, hate speech, hyper-partisanship). Under GDPR, if these scores automatically lead to ad-blocking or censorship, the logic must be explainable.
- **Bias Mitigation (AI Act):** To prevent algorithmic bias, Factmata adopted an "expert-in-the-loop" model, recruiting over 2,000 experts (journalists, scientists) to annotate training data. This directly addresses the AI Act's requirement for high-quality data governance.

2. Technical Architecture & Stages

- **Stage 1: Design (Taxonomy):** Development of 12 scoring dimensions, including non-objectivity, racism, sexism, and clickbait.
- **Stage 2: Model Training (Expert Annotation):** Unlike systems trained solely on general internet data, Factmata built a proprietary annotation platform where experts actively generated the training dataset to ensure the AI could handle nuance.
- **Stage 3: Integration (Cision One):** Following its acquisition by Cision in 2022, Factmata's tech was integrated into the Cision One platform to monitor "narratives" across a massive media database.

3. Practical Steps for FIMI Countering

- **Topic Clustering:** The AI groups similar opinions into "narratives" (e.g., "COVID-19 is a hoax") regardless of the specific phrasing used. This allows analysts to track the *idea* rather than just keywords.
- **Pre-Virality Detection:** The system is designed to flag harmful narratives in the "pre-virality phase," allowing brands or governments to intervene before the damage spreads to the mainstream.
- **Brand Safety:** For advertisers, Factmata provides a "trust rating," allowing brands to programmatically avoid placing ads on sites flagged for hate speech or propaganda.

4. Challenges

- **Nuance & Sarcasm:** While the expert training data helps, NLP models still struggle with high-context sarcasm or cultural nuances compared to human analysts.
- **Commercial vs. Civic Goals:** Post-acquisition, the focus shifted heavily toward brand reputation management (protecting companies from boycotts) rather than purely defending democratic processes, highlighting a tension between commercial viability and civic defense.

2.3. NewsGuard

Type: Human-Centric Journalistic Auditing. **Core Philosophy:** "Journalistic" compliance-rejecting algorithms in favor of trained analysts applying transparency criteria[16].

1. Regulatory & Compliance Context

- **Transparency & Explainability:** NewsGuard is the most compliant with the "Right to Explanation" (GDPR). Every rating (0-100) is accompanied by a detailed "Nutrition Label" explaining exactly why a site passed or failed specific criteria.
- **Apolitical Criteria:** To avoid regulatory pushback on bias, NewsGuard utilizes nine static, apolitical criteria (e.g., "Does not repeatedly publish false content," "Discloses ownership").
- **Defamation Liability:** As a provider of reputation scores, NewsGuard faces legal risks. In 2025, a federal judge dismissed a defamation lawsuit against them, ruling that their ratings are protected opinions based on disclosed methodology.

2. Technical Architecture & Stages

- **Stage 1: Criteria Definition:** Establishing the nine weighted criteria. For example, "Does not repeatedly publish false content" carries the highest weight (22 points).
- **Stage 2: Human Analysis:** Unlike Factmata or Graphika, NewsGuard's primary "engine" is a team of 50+ trained journalists who manually review sites. They do not rely on AI for the rating process.
- **Stage 3: AI Auditing (New):** Recently, NewsGuard has pivoted to *auditing* AI. They conduct "Red Teaming" on LLMs (like OpenAI's Sora or DeepSeek) to test how often these models regurgitate FIMI narratives.

3. Practical Steps for FIMI Countering

- **Right of Reply:** A critical procedural step is contacting the website publisher *before* issuing a negative rating to ask for comment. This procedural fairness is rare in the industry and mitigates liability.
- **Misinformation Fingerprints:** They maintain a machine-readable catalog of top false narratives (e.g., specific myths about the Ukraine war). Platforms can use this data to train their own AI to recognize these specific storylines.
- **Browser Integration:** The ratings are deployed directly to users via browser extensions, placing a "Green" (trusted) or "Red" (caution) shield next to links in search results.

4. Challenges

- **Scale vs. Nuance:** The reliance on human analysts limits scalability compared to AI-driven competitors like Logically or Blackbird.AI. They have rated ~35,000 top sites, but cannot instantly rate every new blogthat pops up.
- **Political Targeting:** Despite their "apolitical" stance, NewsGuard has been the subject of congressional investigations by US Republicans alleging anti-conservative bias, demonstrating the political risks of being an arbiter of truth.

2.4. Comparative Summary

Based on the case studies of Graphika, Factmata (now Cision), and NewsGuard, the primary challenges in countering Foreign Information Manipulation and Interference (FIMI) center on the trade-offs between automated scalability and human nuance, as well as the increasing difficulty of attribution in an AI-saturated landscape.

Graphika faces the distinct technical and reputational challenge of attributing "cybersocial" threats to state actors. While their mapping of networks like "Spamouflage" allows them to track cross-platform coordination, the rise of generative AI "destabilizes the concept of truth itself," making it increasingly difficult to establishthe ground truth necessary for attribution.

Furthermore, their close work with government entities has attracted political scrutiny and allegations regardingfunding sources, complicating their perceived neutrality.

Factmata, conversely, illustrated the commercial and technical struggle of pure AI detection. Before its acquisition by Cision to power brand safety tools, Factmata faced challenges in scaling its operations and ensuring its NLP models could handle the high-context nuance of human speech without "expert-in-the-loop" annotation. Their trajectory highlights the difficulty of maintaining a standalone business model focused purely on civic disinformation defense versus commercial reputation management.

NewsGuard operates at the opposite end of the spectrum, grappling with the "scale-nuance tradeoff." By relying on human journalists to rate credibility using nine criteria, they offer high explainability but lack the speed to match the automated scale of their competitors. This human-centric approach also exposes them to intense political polarization; they have faced congressional investigations and lawsuits (though often dismissed) alleging bias against conservative outlets.

Moreover, the rapid proliferation of AI-generated content has forced NewsGuard to pivot toward auditing AI models themselves-revealing, for instance, that models like DeepSeek produce disinformation at an 83% failure rate-a task that requires constantly evolving methodologies to keep pace with synthetic media generation.

Schedule 1: Detailed Compliance & Technical Governance Checklist for AICP-FIMI

This annex serves as a practical execution guide for the "Design," "Development," and "Deployment" phases, integrating legal prohibitions from the EU AI Act with technical safeguards for autonomous agents.

1. Regulatory "Red Lines" (Prohibited & High-Risk Practices)

Objective: Ensure the architecture strictly avoids practices banned under the EU AI Act and GDPR before any code is deployed.

Compliance Domain	Requirement Description	Implementation Verification Step
Biometric Scraping	**Strict Prohibition:** The platform must **not** scrape facial images from social media to build recognition databases (AI Act Art. 5(1)(e)).	Verify web scrapers explicitly exclude .jpg/.png collection from user profiles. Audit data ingestion pipelines to ensure no vectorization of facial features.
Political Profiling	**Strict Prohibition:** No biometric categorization (e.g., analyzing profile photos) to infer political opinions (AI Act Art. 5(1)(g)).	Ensure classification models rely solely on *behavioral metadata* (posting frequency, network patterns) and *text semantics*, never on physical attributes.
Subliminal Techniques	**Strict Prohibition:** Avoid techniques that distort behavior or impair informed decision-making (e.g., hyper-personalized, manipulative bot responses).	Any counter-narrative content generated by the system must be clearly labeled as AI-generated (transparency markers).
Data Minimization	**GDPR Art. 5(1)(c):** Collect only data strictly necessary. Indiscriminate retention is banned (CJEU *La Quadrature du Net*).	Implement "TTL" (Time-To-Live) on all raw scraped data. Use "Semantic Caching" to store abstract patterns rather than raw user data where possible.

2. Agentic Architecture & Protocol Compliance

Objective: Standardize agent communication and tool use to prevent "Black Box" liability and ensure interoperability (Model Context Protocol & Agent2Agent).

Technical Component	Requirement	Technical Control / Action Item
Tool Interfaces (MCP)	Agents must connect to external tools (social media APIs, databases) via **Model Context Protocol (MCP)** to standardize security borders.	Implement MCP Servers for all data connectors (e.g., Twitter/X API, Neo4j). Enforce "User Consent" prompts before an agent executes a "Write" action (e.g., reporting a bot).
Inter-Agent Comms (A2A)	Use **Agent2Agent (A2A)** protocol for coordination between "Research Agents" and "Analysis Agents" to ensure auditability.	Create "Agent Cards" (JSON metadata) for every agent, defining capabilities and auth requirements. Use A2A "Task Lifecycle" states (pending, working, completed) to track long-running FIMI investigations.
Identity & Access (RBAC)	Treat Agents as **Non-Human Identities (NHIs)** with granular permissions.	Assign "Read-Only" access to Research Agents. Require "Human-in-the-Loop" (HITL) tokens for any agent attempting a

		"High Risk" action (e.g., public alerting).
Agentic RAG	Move from static retrieval to **Agentic RAG** with self-correction capabilities.	Implement "Iterative Retrieval": If initial data is insufficient, the agent rewrites the query rather than hallucinating. Enable "Citation Matching": The model must link every claim to a specific document ID (provenance).

3. Algorithmic Transparency & Explainability (GDPR Art. 22)

Objective: Satisfy the CJEU *Schufa* and *Dun & Bradstreet* rulings requiring disclosure of the "logic involved" in automated assessments.

Metric/Feature	Requirement	Implementation Strategy
Chain of Thought (CoT)	Log the reasoning steps, not just the final output.	Enable "CoT Logging" in the orchestration layer. The log must show: *Query -> Plan -> Tool Call -> Observation -> Conclusion*. Store CoTlogs in an immutable audit trail.
Logic Disclosure	Provide meaningful information about how a "Bot" classification was reached.	The UI must display *why* an account was flagged (e.g., "Flagged because posting frequency > 500/hr AND network centrality > 0.9"). Avoid generic "AI Score" labels.
Hallucination Checks	Prevent fabrication of evidence.	Implement a "Verifier Agent" step: A secondary model checks if the generated report matches the retrieved source data (Grounding Check).

4. Operational Governance & Security (TRAPS Framework)

Objective: Implement the **TRAPS** (Trusted, Responsible, Auditable, Private, Secure) framework for agent lifecycle management.

TRAPS Component	Operational Control	Risk Mitigation
Trusted	**Grounding:** Agents must query specific knowledge bases (DSA archives), not general internet scraping.	Mitigates risk of processing copyright/illegal content.
Responsible	**Bias Testing:** Regular "Red Teaming" to ensure the model does not disproportionately flag specific linguistic or ethnic groups.	Mitigates violation of non-discrimination laws (AI Act).
Auditable	**"Flight Recorder":** Full traceability of agent actions.	Allows forensic reconstruction of why a FIMI alert was triggered.
Private	**PII Redaction:** Output guardrails must strip names/IDs before reports leave the secure enclave.	Prevents accidental data leaks (GDPR).
Secure	**Circuit Breakers:** Automated stops if an agent exceeds rate limits or accesses unauthorized APIs.	Prevents runaway agents or "Reward Hacking."

5. Deployment & "Automatability" Roadmap

Objective: Align the project timeline with the maturity of compliance-automating technologies to reduce overhead.

Phase 1: Design (Automatability Trigger)

- *Action:* Do not activate automated takedown features until Compliance-Automating AI (agents that autonomously check other agents against EU rules) is integrated.
- *Rationale:* Reduces the risk of premature regulation violations and high human oversight costs.

Phase 2: Hybrid Teaming (Deployment)

- *Action:* Deploy as a "Cyborg" workflow (Human-AI hybrid) rather than fully autonomous.
- *Rationale:* Studies show hybrid teams outperform fully autonomous agents by 68.7% in quality. Humans handle judgment; agents handle data scale.

Phase 3: Continuous Monitoring

- *Action:* Use Guardrails (e.g., NeMo, simple deterministic rules) to filter inputs/outputs for "Prompt Injection" attacks or "Jailbreaks" attempting to weaponize the FIMI platform.
- *Rationale:* Protects against adversarial attacks typical in high-stakes geopolitical environments.

References

1. Artificial Intelligence Act: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202401689
2. General Data Protection Regulation: https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng
3. CJEU Case C-446/21 - Meta Platforms Ireland, 2024. https://curia.europa.eu/juris/document/document.jsf?text=&docid=290674&pageIndex=0&doclang=EN&mode=lst&dir=&occ=first&part=1&cid=8907758
4. CJEU Case C-203/22 - Dun & Bradstreet Austria GmbH, 2025. https://curia.europa.eu/juris/liste.jsf?num=C-203/22
5. CJEU Case C-634/21 - SCHUFA Holding AG, 2023. https://curia.europa.eu/juris/liste.jsf?num=C-634/21
6. CJEU Case C-511/18 - La Quadrature du Net and Others, 2020. https://curia.europa.eu/juris/liste.jsf?language=en&num=C-511/18
7. Digital Services Act. https://eur-lex.europa.eu/eli/reg/2022/2065/oj/eng
8. The Cyber Resilience Act: https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=OJ:L_202402847
9. NIS2 Directive: https://eur-lex.europa.eu/eli/dir/2022/2555/2022-12-27/eng
10. CJEU Case C-413/23 - EDPS v SRB, 2025. https://curia.europa.eu/juris/document/document.jsf?text=&docid=305584&pageIndex=0&doclang=EN&mode=req&dir=&occ=first&part=1&cid=8121207
11. CJEU Case C-817/19 - Ligue des droits humains. https://curia.europa.eu/juris/liste.jsf?num=C-817/19
12. Cyber Solidarity Act: https://eur-lex.europa.eu/eli/reg/2025/38/oj/eng
13. Case C-250/25 - Like Company v. Google Ireland Ltd. https://curia.europa.eu/juris/liste.jsf?num=C-250/25
14. Graphika. https://www.graphika.com/
15. Cision. https://www.cision.com/
16. NewsGuard. https://www.newsguardtech.com/