FERMI
FAKE NEWS
RISK MITIGATOR

Project acronym:	FERMI
Project full title:	Fake nEws Risk Mitigator
Call identifier	HORIZON-CL3-2021-FCT-01
Start date:	01/10/2022
End date:	30/09/2025
Grant agreement no:	101073980

D2.1 FERMI starting point package

Work package: 2

Version: 1.5

Deliverable type: R - Document, report

Dissemination level: PU

Official submission date: M6

Actual submission date: M8/M18 (slightly adjusted)

Funded by
the European Union

Leading author(s):

Surname	First name	Beneficiary
Papadakis	Thanasis	INTRA
Gousetis	Nikos	INTRA
Vourtzoumis	Michalis	INTRA
Evangelatos	Spyros	INTRA

Contributing partner(s):

Surname	First name	Beneficiary
Bravos	George	ITML
Papargyri	Eleni	IANUS
Glöckner	Paul	BIGS
Stuchtey	Tim H.	BIGS
Valente	Catarina	INOV
da Costa	João Varela	INOV
Aziani	Alberto	UCSC
Giglio	Flavia	KUL
Garcia	Joaquin	ATOS
Rojas	Jairo	ATOS

Peer reviewer(s):

Surname	First name	Beneficiary
Dimakopoulos	Nikos	ITML
Kuch-Wesolowski	Robert	SPA

Ethics reviewer:

Surname	First name	Beneficiary
Fikenscher	Sven-Eric	BPA

Security reviewer:

Surname	First name	Beneficiary
Tobias	Mattes	BPA

Document Revision History

Version	Date	Modifications Introduced	
		Modification Reason	**Modified by**
0.1	02/02/2023	ToC	INTRA
0.2	10/03/2023	Class and Sequence Diagrams for technical offerings	INOV, UCSC, ATOS
0.3	20/03/2023	Class and Sequence Diagrams for technical offerings	ITML, BIGS
0.4	30/03/2023	Integration matrix of technical components	INOV, UCSC, ITML, BIGS, ATOS, INTRA
0.5	20/04/2023	Final use case requirements and scenarios	IANUS
0.9	25/04/2023	Integration of all chapters	INTRA
0.9.1	26/04/2023	Addition of final modifications for Chapter 2	KU Leuven
0.9.2	27-3 /04/2023	Security and Ethics Reviews	BPA
0.9.3	04/05/2023	Review	ITML
0.9.4	05/05/2023	Review	SPA
1.0	21/05/2023	Finalisation	BPA
1.1	28/08/2023	Revision	INTRA, BPA, UCSC, ITML, BIGS, ATOS, INOV
1.2	07/09/2023	Revision of chapter 2 paragraph 1.4 and chapter 3	INTRA, UCSC, BIGS
1.3	28/09/2023	Revision of chapter 2 paragraph 1.4 and chapter 3	INTRA, ITML, UCSC, INOV, ATOS
1.4	03/11/2023	Finalization of revisions on chapter 2 paragraph 1.4 and chapter 3	INTRA
1.5	27/11/2023	Revisions on chapter 3	INTRA, ITML, INOV

Executive summary

The aim of this deliverable is to provide a detailed description of the work performed within WP2 of the FERMI project. It summarises the derived user requirements following the MoSCoW (Must have, Should have, Could have and Won't have) method, based on the needs of the Law Enforcement Agencies [T2.1] serving as the basis of the technical developments that will follow in WP3 and WP4. A dedicated chapter addressing the societal landscape of the project has been provided demonstrating a fair balance of interests between law enforcement objectives and the protection of fundamental rights and democratic values, including freedom of speech, freedom of thought, conscience and religion, respect of private and family life, and freedom of assembly and association [T2.2]. In addition, the above-mentioned user requirements are translated into functional requirements and technical specifications connecting all technologies into a single platform [T2.3]. The platform architecture is also described in terms of interfaces, information flows, components interactions, and deployment views. Lastly, the use case scenarios [T2.4] based on the elicited user needs to fight disinformation and fake news are listed in order to facilitate the demonstration of the project's findings in WP5. Moreover, a set of key performance indicators are presented.

Table of Contents

List of figures and tables

Abbreviations

AI:	Artificial Intelligence
ATOS:	ATOS IT Solutions and Services Iberia SL
BFP:	Police Federale Belge
BIGS:	Brandenburgisches Institut für Gesellschaft und Sicherheit GGMBH
BPA:	Bavarian Police Academy
CJEU:	Court of Justice of the European Union
DMIA:	Ministere de l'Interieur
D&FN:	Disinformation and fake news
ECHR:	European Convention of Human Rights
ECtHR:	European Court of Human Rights
EU:	European Union
FMI:	Ministry of the Interior
GA:	Grant Agreement
GDPR:	General Data Protection Regulation
HLEG:	High Level Expert Group
IANUS:	IANUS Consulting Ltd
INTRA:	Netcompany-Intrasoft SA
INOV:	INOV Instituto de Engenharia de Sistemas e Computadores Inovacao
IT:	Information Technology
ITML:	Information Technology for Market Leadership
KPI:	Key Performance Indicators
KU Leuven:	Katholieke Universiteit Leuven
LEA:	Law enforcement agency
RAN:	Radicalisation Awareness Network
UC:	Use Case
UCSC:	Università Cattolica del Sacro Cuore
UI:	User Interfaces
UR:	(End-)user requirement
VUB:	Vrije Universiteit Brussel

1 User Requirements Elicitation: Updating and finetuning end users' needs and laying the ground for the experimentation protocol

1.1 Introduction and Foundational Terminology

The effort to lay the ground for the proper validation of the envisaged FERMI platform includes the conception of a first experimentation protocol, which will be subject to a further review and, if necessary, adjustments, at a later stage of the project (in the framework of WP5, in D5.1 to be exact). The key steps of the first outline that provides guidance on how to test the platform resulting in the experimentation protocol (see section 4) are summarised as follows:

1. End-User Requirements definition – This step captures the elicitation of the essential requirements as reported by end-users. To this end, an informal workshop and a survey were conducted.

2. Functional and Non-Functional Requirements definition – This step refers to the definition of functional and non-functional requirements.

3. Use cases and user scenarios definition – This task involves the definition of test cases to trace the fulfilment of requirements and key performance indicators (KPIs). The use cases will be fine-tuned in view of data availability by stakeholders and use case leaders. Some use cases will also be further defined with the contribution of technical partners, especially the ones to evaluate the requirements and KPIs that are more technical.

4. KPI definition – This step aims at identifying key performance indicators that can be used to further measure the use cases and user scenarios' successful implementation in the sense of grasping whether the key expectations of end-users as summarised in the end-user requirements have been met.

This deliverable has been largely structured accordingly. The notable exception is a societal landscape analysis, which is included after the end-user requirements' elicitation. Whilst not embedded in the experimentation protocol in a narrow sense, the societal landscape analysis is a key prerequisite for the application of the end-user requirements and the functional and non-functional requirements to testing proceedings. Thanks to the detailed analysis of how to balance the need to stem the tide of disinformation and fake news (D&FN) with basic human and civil rights, it is clarified what further conditions (non-user oriented and non-technical) must be taken into consideration when the platform is being evaluated such as the limitations of the legal mandate for government interference. D&FN must not be monitored or analysed by law enforcement agencies (LEAs) as such but only when such steps are justified, for example in the event the D&FN campaign includes illegal messages requiring an investigation.

Moreover, **the societal landscape analysis has identified three common elements that should be integrated in any definition of disinformation: 1) factual or misleading nature of the information; 2) intention of the actors to spread such information they know to be false to obtain economic gain or deceive the public; 3) public harm.** Accordingly, this definition will guide the further work of the FERMI project.

Before proceeding with the end-user requirements' elicitation, some basic definitions of the most crucial terms are shared below to provide a sufficient context and understanding of the subject matter,

Analysis takes the information we elicited and looks at where the gaps and impacts are; breaking down the information and looking at it through different angles.

Elicitation is the discovery, progressive elaboration and understanding of the needs of your stakeholders and customers.

Requirement: A requirement is a singularly documented physical and functional need that a particular design, product or process must be able to perform. It is a statement that identifies a necessary attribute, capability, characteristic, or quality of a system for it to have value and utility to a customer, organisation, internal user, or other stakeholders along the lines of their expectations.

End User: An end user is a person who ultimately uses or is intended to ultimately use a product and has a direct interaction with it. The end user stands in contrast to users who support or maintain the product, such as system administrators, database administrators, Information technology (IT) experts, software professionals and computer technicians.

Use Case (and user scenarios): Broadly, a use case is a description of the conditions under which a specific system/process/product is utilised by an actor in order to solve a problem or achieve a goal. More specifically, a use case is a methodology used in system analysis to clarify, organise and test system requirements. User scenarios are made up of a set of possible sequences of interactions between systems and users in a particular environment and related to a particular goal.

Need-to-know principle: A user must only be granted access to information that is relevant to their job duties, regardless of their level of security clearance or any other approvals they may have. This means that in order for a user to access information, he/she must have both the necessary permissions and a legitimate need-to-know that is directly related to their current role.

Functional requirements define what a product must do, what its features and functions are.

Non-functional requirements describe the general components of a system. They are also known as quality attributes.

1.2 Methodology

This section refers to the methodology followed for the implementation of two different tasks of the FERMI project, namely Task 2.1 “User Requirements Elicitation: Updating and finetuning end users' needs” and Task 2.4 “Experimentation protocol: Use cases' refinement and pathway towards FERMI validation”. The activities conducted followed a user-centered approach, with the primary objective of gaining a comprehensive understanding of the circumstances surrounding the end users, which enabled the elicitation of valuable insights into user requirements, which were subsequently utilised in the formulation of use-cases and drafting of scenarios. The two tasks, which focused on separate topics, were designed to be complementary in nature, with the exchange of feedback and good practices being an integral part of the research process. The approach was user-centered in total, yet the techniques deployed vary, in order to prevent gaps and delve deeply into the topic under research. Both quantitative and qualitative research was employed.

The section presents a detailed methodology for user requirements elicitation which leads to the definition of use cases (UCs) and scenarios. The methodology is structured into separate sections, with each sub-section providing a clear outline of the procedure and supporting the methodology as a whole. This approach ensures that the research is conducted in a systematic and rigorous manner, and that the resulting insights are reliable and valid.

Right from the start of the project, IANUS alongside the Coordinator BPA had developed a plan for how to reach the end-users. It was decided that all the partners of the consortium, especially the relevant stakeholders, would send an informative email to all EU relevant stakeholders of the FERMI project and invite them to participate in the survey which would be circulated by IANUS.

1.3 End-User Requirements analysis

As indicated above, requirements are an expression of what the end-users expect and want to see from a product, tool, process etc. In other words, it can be defined as the condition or capability that an end-user needs to achieve a certain objective. In that regard, FERMI adopted a user-centred approach for the requirements elicitation phase. The chosen approach gives great attention to the context, constraints of end-user uptake and opportunities the envisaged FERMI platform might provide for them, attempting to guide them through their own experience, in order to unveil not only the current parameters of their work, but also what they could potentially expect from the project and its results. In this vein, the user-centred approach followed a specific structure of activities, including setting and understanding the context, identifying stakeholders and end-users, requirements elicitation and documentation and requirements analysis and incorporation into use cases and scenarios.

After having discussed that the methodology would be user-centred and mixed in terms of data acquisition (quantitative and qualitative) a few further remarks on the context are warranted. More specifically, it was essential to consider a variety of investigation techniques from which to choose, according to what the circumstances would be in each case. At the beginning of a requirements elicitation process, one must plan with some degree of flexibility, so that risks and constraints that arise can be handled, minimising the impact. This approach may include several investigation techniques for a comprehensive requirements-gathering process, such as meeting with stakeholders to understand the constraints and expectations, conducting workshops with end-users, using questionnaires, creating user profiles, creating scenarios, comparing and generalising user requirements and creating use cases.

This is under no circumstances an exhaustive set of steps in a process, nor does it define the order in which these techniques are to be utilised. **The chosen approach was focused on a workshop and questionnaires** (**use cases and scenarios will be derived from the results thereof later on**). These were determined, after a number of internal discussions with the WP2 partners, as the ones best fit for the nature of the group of end-users we had, taking into consideration certain constraints that we faced. For instance, conducting a field observation to take a close look at the work environment and workflow of the LEA officers involved in the consortium, was unfortunately not a viable option, since there were authorisations issues to be taken into consideration which alongside with the limited timeframe of the task rendered this option not feasible.

Firstly, an online meeting with the end-users and the technical partners of the consortium was implemented to establish an initial contact in order to acquire some necessary guidance for the next steps of the procedure, by understanding the necessary context. After an intensive and detailed session, there were yet matters to be discussed and discovered, which would be the case for the next steps of the elicitation process. The meeting yielded significant understanding of the LEAs background and operational environment in the context of combating the spread of disinformation and fake news.

Capitalising on the elementary understanding built through this first consultation, the elicitation process continued with the organisation of an informal virtual workshop and the dissemination of a questionnaire. These constitute the two main pillars of this process.

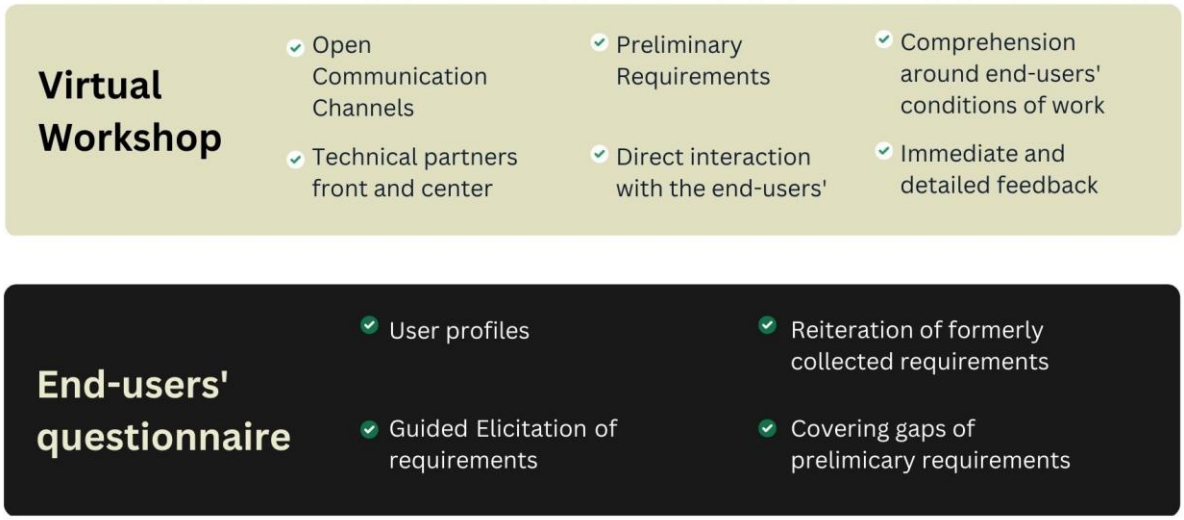


Figure 1 Schematic representation of the two pillars of the methodology

Throughout the elicitation process, requirements were analysed and documented properly. When the entire elicitation process was concluded, a more elaborated analysis process begun, which consisted of several sub-processes, such as: categorising requirements, filtering them to avoid overlapping and repetition, ensuring clarified and precise phraseology that derives from the end-users' perspective, as well as eliminating or re-defining any requirements that were irrelevant, not feasible or not compliant with the project objectives.

The tables, diagrams, use cases and other tools that are used within this document offer a rich representation of FERMI requirements. Towards this aim, we followed the MoSCoW Methodology to elicit and prioritise user requirements, while focusing on identifying the features or functions that are most important to the end user.[1] Accordingly, we categorised the requirements as follows:

1. Must have: These are the features or functions that are absolutely critical to the end users and must be included in the product for it to meet its needs. These requirements are non-negotiable and must be implemented to ensure user satisfaction.
2. Should have: These are the features or functions that are important to the end user but not essential. They can be deferred to a later phase or release, if necessary, but should still be considered for inclusion in the product.
3. Could have: These are the features or functions that are desirable but not critical to the end user. They can be considered if there is time and budget available, but should not be a priority over must-have or should-have requirements.

[1] Madsen, 'How To Prioritise Requirements With The MoSCoW Technique,' *Knowledgehut* (12 April, 2023). Available at: https://www.knowledgehut.com/blog/agile/how-to-prioritise-requirements-with-the-moscow-technique.

4. Won't have: These are the features or functions that are not relevant or necessary for the end user, and can be safely excluded from the product.

For the classification of the user requirements accordingly we followed the following approach. Whenever there was a consensus amongst end-users in the informal workshop the requirement that was being discussed was ranked a Must-have. In the event, there was a near-consensus with just one partner casting doubt on a certain measure, the requirement that was being discussed was ranked a Should-have. More controversial discussions resulted in the relevant requirement being incorporated as a Could-have and in the absence of any support the requirement was dismissed as a Won't have.

While drafting the survey for the end-users to gather information on their requirements a score was assigned on a scale of 1-5 based on its importance to the end-user, with 1 being the least important and 5 being the most important. The following step included the analysis of the survey results and sort the requirements based on their score. The requirements were sorted based on the most popular answer. Participants were mostly asked to rate statements on the necessity of platform components on the basis of five different options (ranging from "not important" to "very important") that could easily be transformed into a 5-point scale grasping end-user requirements with the most popular option participants could pick and choose from determining where to categorise that option on the 5-point scale. The few further questions either directly inquired into numbers (such as the scope of accuracy of tracing back the origin of D&FN percentage-wise) or asked participants to give straight "Yes" or "No" answers that could be easily categorised into a 5-point scale too by distinguishing between 20% intervals of Yes votes (0-20% approval, more than 20%-40% approval, more than 40%-60% approval, more than 60%-80% approval and more than 80% approval).

Requirements with a score of 5 are categorised as Must-have, those with a score of 4 are categorised as Should-have, those with a score of 2-3 are categorised as Could-have, and those with a score of 1 are categorised as Won't-have.[2] In the event of contradictory feedback from workshop and survey participants, a requirement was not ranked a Must-have. Other than that, the survey replies took precedence due to their comprehensive nature (we had a huge number of end-user participants that greatly exceeded the boundaries of the consortium, see below).

The following sub-section elaborates on Stakeholders Identification, the informal virtual Workshop and the Questionnaire.

1.3.1 Stakeholders and End-Users' identification

To collect the required data from the relevant stakeholders, the latter needed to be delineated in the first place. The term "stakeholders" may refer to entities, people or organisations (legal entities such as companies,

[2] Kravchenko, Bogdanova, and Shevgunov, 'Ranking Requirements Using MoSCoW Methodology in Practice,' *Lecture Notes in Networks and Systems* (2022). Available at: doi: 10.1007/978-3-031-09073-8_18.

standards bodies), that have a valid interest in the system under development. This rather abstract definition notwithstanding, the project's Grant Agreement (GA) implies that the end-user requirements' elicitation developed under task 2.1 shall be particularly addressed to the law enforcement officers of EU Member States who work on combating the criminal ramifications of D&FN, in particular by implementing technical solutions like the ones developed by the FERMI project. The GA explicitly says that "FERMI will facilitate EU Police Authorities to […] monitor the way that D&FN spread, both in terms of locations and within different segments of the society, and to put in place relevant security countermeasures."[3] The further stipulation to provide "[m]odern information analysis for Police Authorities, allowing them to efficiently fight criminals and terrorists who use novel technologies"[4] is even more illustrative. In other words, the consortium is not only required to support the fight of LEAs against the ramifications of D&FN, in this regard a special emphasis needs to be placed on the LEAs' fight against crime and terrorism.

This approach is fully supported by the topic of the call for proposals, urging that "[t]his topic requires the active involvement, as beneficiaries, of at least 3 Police Authorities [...] from at least 3 different EU Member States or Associated countries"[5] (which the FERMI consortium easily meets thanks to the involvement of SPA, FMI, BFP, DMIA (until 30 June, 2023) and Guardia Civil (since 01 December, 2023)). Moreover, the WP5 deliverables' content, which includes the revised experimentation protocol and the pilot evaluations, has been classified sensitive, which largely leaves us with the consortium LEAs as possible evaluators anyway. Quite tellingly, the platform-related training material is meant to be tailored to "officers in LEAs […], in order to enable their re-/up-skilling focusing on their capability to understand the spread of D&FN, to exploit the FERMI tools" etc.[6] Other than the EC and the consortium the matching deliverable (D5.4) can only be made available to "EU LEAs (Police and Border authorities from the EU)."[7]

Considering the huge role LEAs are expected to play as target group of the FERMI platform and further considering the specific legal authorisations and constraints that apply only to LEAs[8] (unlike other target groups that may share the interest in stemming the tide of D&FN but neither have any authority to launch investigations into criminal and terrorist activities nor do they have to comply with *LEA-specific* legal constraints), LEAs have been selected as the end-users of the platform tools. Accordingly, end-user requirements are elicited from LEA and LEA-affiliated experts and the experimentation protocol is tailored to the consortium LEAs (albeit other groups that hold a stake in the overall work that is being carried out in

[3] Gant Agreement, PART B, p.4.
[4] Grant Agreement, PART B, p.23.
[5] EU Commission, *Funding and Tender Opportunities, Disinformation and fake news are combated and trust in the digital world is raised* (TOPIC ID: HORIZON-CL3-2021-FCT-01-03) (2022). Available at: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-cl3-2021-fct-01-03.
[6] Grant Agreement, PART B, p.12.
[7] Grant Agreement, PART B, p.38.
[8] Those are summarised in section 2.

FERMI are engaged through communication, dissemination and exploitation activities too (see WP6, as far as RP1 is concerned, D6.1 and D6.2)).

1.3.2 Informal Online Workshop

The first step that was implemented to identify the end-user requirements was the organisation of an online workshop where the technical partners presented the proposed technologies and the LEA end-users expressed their views, their knowledge regarding the operations of their agencies and the currently available tools. Workshops are a great technique to get all relevant stakeholders in the same room and work together on bridging potential gaps in mutual understanding, as well as conversing on the issue at hand each time. Since organising an in-person workshop would require more time and planning in advance, a virtual workshop was chosen as the first pillar of the elicitation process. Holding a virtual workshop in the form of informal consultations was essential to collect more elaborate information from all end-users of the consortium. The nature of the workshop being a virtual one was necessary at the time, to sustain the momentum created while considering practical and time-related constraints.

For this virtual workshop, the purpose was to discuss the end-users' workflow, including the analysis of procedures, obstacles and problems faced, success stories and to procced on that basis by asking the above-mentioned LEA end-users of the consortium to provide information on how they attempt to grasp and mitigate D&FN-induced ramifications that are of relevance to them. The workshop also intended to open a channel of direct and constructive communication between all end-users within the consortium and the technical partners, particularly centered around end-users and any constraints, needs, challenges or even successes they face (see Figure 2). During that first informal consultation session, a list of procedures that potentially fit the workflow of our end-users was presented. This list was also communicated to the workshop participants requiring to receive a step-by-step description of these procedures, in order to understand the trail followed by end-users in completing the process and achieve the respective goal (if any).

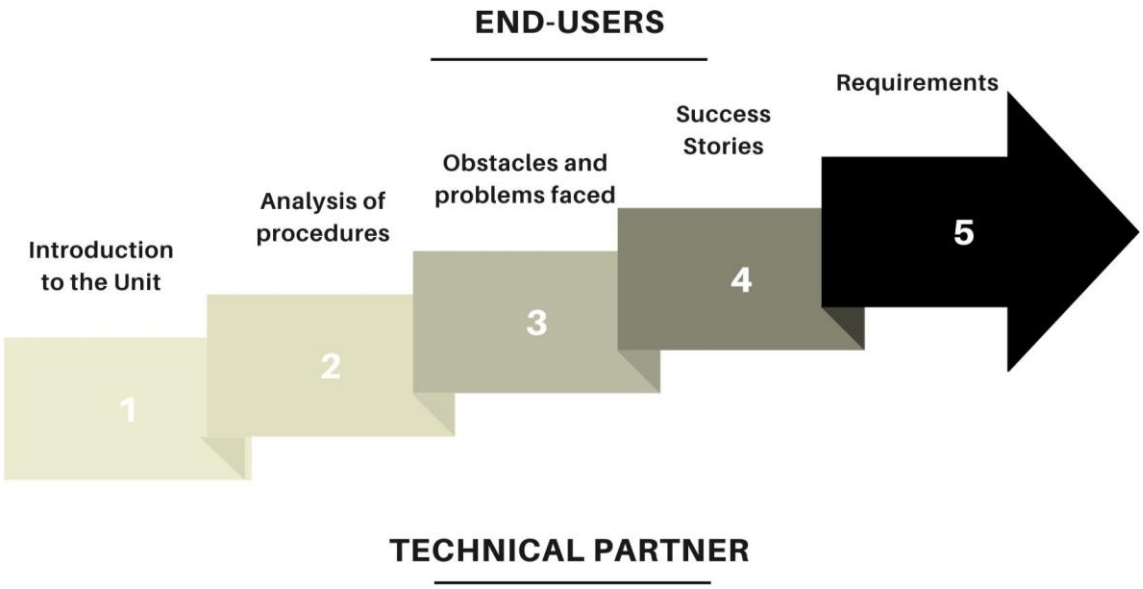


Figure 2 Virtual workshop flow of information

Any suggested end-user requirement was further elaborated in the discussion to avoid miscommunications and redundant information.

1.3.3 Questionnaire

Following the organisation of the online workshop the elaboration of the questionnaire took place. Questionnaires have been widely used to collect data from stakeholders and thus they constitute the backbone of many surveys.[9] Some of the benefits questionnaires offer as a research method are that they can keep the participants' identity from being revealed, the target audience, even if geographically spread, can be identified, and reached, and the majority of respondents are aware of their scope.

Questionnaires are widely used in research as a tool for collecting primary quantitative data. However, in the present report, the questionnaire developed under Task 2.1 aimed to extract both quantitative and qualitative data. To achieve this objective, a combination of open-ended and closed-ended questions was utilised, taking advantage of the benefits offered by both methods. Closed-ended questions restrict the respondent to a predetermined set of answers, such as "yes" or "no." This format allows for quantifiable and comparable results, facilitating statistical analysis. On the other hand, open-ended questions provide the respondents with the opportunity to express their opinions freely, without any constraints or preconceived ideas. This approach avoids bias and allows for spontaneous and personalised responses. The questionnaire utilised in this report incorporates both types of questions, resulting in a comprehensive data collection tool.

[9] International Institute of Business Analysis, *BABOK: A guide to the business analysis body of knowledge*® (2015). Available at: https://www.iiba.org/career-resources/a-business-analysis-professionals-foundation-for-success/babok/.

To assure that the questionnaire would be as user-friendly as possible particular emphasis was placed on the smooth and clear question sequence as well as on the wording used, which are key elements for a successful questionnaire.

Additionally, the privacy of the participants and the protection of their data was a top priority while developing the questionnaire. For this reason, to ensure that the questionnaire was compliant with the data privacy legislations, the draft was reviewed by the KU Leuven and VUB as legal and ethics experts, while the leader of T2.1 IANUS had appointed a data protection officer to monitor the implementation of the task. Additionally, even though the questionnaire was circulated via the EU Survey platform, the anonymous mode was activated (Figure 3).

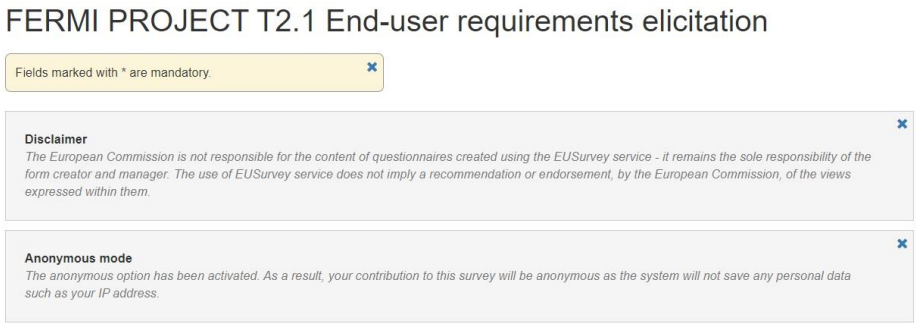


Figure 3 Anonymous mode of questionnaire

Once finalised, the questionnaire was circulated to Work Package (WP) 2 partners for feedback. The final version of the questionnaire (ANNEX A: Questionnaire to the LEAs) was circulated to various end-users inside and outside of the consortium on 20 January 2023. Although, the initial plan was to circulate the questionnaire at December 2022, the questionnaire was published later on due to the submission of D7.1 regarding the recruitment of research participants. In total, one hundred thirty four (134) answers were received, all via EU Survey (Figure 4).

Figure 4 Number of responses in the questionnaire

After the first pillar of the elicitation process was completed, the initial round of information needed some more concrete elaboration on specific issues pertaining to FERMI. As explained above, the questionnaire aimed at posing closed-ended questions to the end-users to get a specific "yes" or "no" answer, or a rank of the significance the end-users attributed to certain components, features and conveniences provided by the FERMI proposed technology. Few questions required elaborate replies. Another objective was to gather input by end-users outside of the consortium to get a richer range of replies. The questionnaire was open to all LEAs and LEA-affiliates since their operations are considered relevant to the spread of D&FN which can lead to online and offline crimes.

As mentioned before, a top priority while implementing the survey was the compliance with security and privacy measures, which is why the Questionnaire under Task 2.1 followed data protection and privacy norms and rules, shared an "Information Sheet" (ANNEX B: Information Sheet), requiring its acceptance by providing informed consent (by checking a box) and utilised the "Anonymity mode" of the EU Survey Platform, which was used to collect feedback. Wishing to further ensure that replies to the questionnaire could not be in any way "pinned" on a specific person, we omitted asking about the country of origin of the questionnaire participants. The structure of the questionnaire was such to create a coherent flow of the questions and prepare the end-users gradually.

Section A of the Questionnaire was a "Welcome" chapter, presenting information on FERMI.

Section B consisted of the Information Sheet and Consent Form.

Section C was intended to gather information on the end-users' profile and determine their experience in combating the ramifications of disinformation and fake news and the connection with their work in LEAs,

since the questionnaire was open not only to active law enforcement officers, but also to non-active duty LEA personnel and LEA-affiliates.

Section D was the last one, containing all questions that would result in clarifying or collecting end-user requirements. (These were informed by the ambition to go beyond the state-of-the-art as explained in the FERMI GA. End-users could then assess whether certain technical solutions matched their needs. More specifically, such questions addressed the role of AI-based tools in predicting D&FN-induced crimes and the deployment of LEAs, the accuracy level in identifying the origin of D&FN, distinguishing between physical persons and bots running accounts, predicting the exact kind of crimes induced by D&FN, the use of machine-learning, forecasting the victims, doing threat and risk assessments, predicting the environment and context, the role of economic and social factors, quantifying the costs, cultural aspects, big data analysis, community resilience, behavioural profiling, taking proper counter-measures, user-friendliness and grasping different forms of violent extremism).

The questionnaire was only finalised after receiving feedback from other consortium partners, including end-users, in order to make sure that it was coherent, comprehensive by the standards of end-users and precise in its wording. In order to share the questionnaire with other LEAs, the task leader contacted all consortium partners, especially the end-users, requesting them to not only participate in the survey, but also utilise their network and close cooperation with other LEAs, and Police Academies, to disseminate the questionnaire and increase participation. The questionnaire solidified certain aspects of the requirements and enriched the material we had. More on the questionnaire, including the exact results, can be found in Annex C: Results from End-Users Questionnaire.

1.4 End-User Requirements

Apart from a comprehensive portrayal of inputs that were gathered throughout the conducted activities, it is important to mention that the requirements in T2.1 are not referring to system specifications, which is part of different task which will be presented below.

The end-user requirements (UR) that are presented in the table that follows, are the product of the analysis laid out in the previous sections of the present document.

Table 1 FERMI User Requirements

FERMI Requirements List UR001-UR038				
UR ID		**Title**	**Priority**	**Origin**
UR001		The user is able to identify whether the X account spreading fake news online is a physical actor or a bot.	Must	Survey

UR002		The user is able to assess the origin of the disinformation with accuracy more than 80%.	Should	Survey
UR003		The user is able to identify key actors involved in spreading disinformation campaigns.	Should	All Sources
UR004		The user is able to contribute to the better allocation of law enforcement resources to prevent and respond to disinformation-induced crimes.	Should	All Sources
UR005		The user is able to grasp the social media interactions of those who are actively promoting D&FN.	Should	All Sources
UR006		The user has the ability to control or regulate the dissemination of information on social media platforms.	Won't	Workshop
UR007		The user is able to use graph data for analysis, based on fetching and transformation of all the responses, likes, and retweets of a disinformation post.	Must	All Sources
UR008		The user is able to estimate the most influential actor in the graph (social media account post) spreading D&FN.	Should	Survey
UR009		The user is able to automatically detect disinformation content on social media platforms.	Won’t	Workshop
UR010		The user through the platform is able to classify disinformation posts by category (e.g., political, health-related).	Could	Survey
UR011		The user is able to analyse the sentiment polarity of social media posts related to disinformation.	Must	Workshop
UR012		The user is able to have access to detailed reports, generated based on the data analysed. The reports should be customisable based on the user's needs and should be easy to understand and interpret.	Must	Workshop
UR013		The user is able to have access to interactive visualisations and dashboards generated by the platform to help law enforcement officers understand complex data patterns and trends.	Should	Workshop
UR014		The user is able to predict who are the potential victims of crimes related to D&FN.	Must	All Sources
UR015		The citizen is able to increase his/her knowledge about the socioeconomic and cultural aspects and the perception of disinformation among citizens.	Should	All Sources

UR016		The user is able to quantify the economic impact by making an approximation on the costs of violent extremism caused by disinformation and fake news.	Could	Survey
UR017	A	The user can identify the geographical unit in which the criminal event may more likely occur due to the D&FN	Should	Survey
	B	The user is able to manage risk based on community behavioural profiles and socioeconomic analysis.	Should	All Sources
UR018		The user is able to determine the economic factors that play a role in the ramifications of disinformation.	Should	Survey
UR019		The user is able to collaborate with other law enforcement agencies to combat the illegal ramifications of disinformation campaigns without the need of sharing the data outside of its facilities.	Should	Workshop
UR020		The user is able to track down the origin and distribution of disinformation campaigns related to violent extremism (right-wing extremism, left-wing extremism, health-related extremism).	Must	All Sources
UR021		The user is able to identify potential threats to public safety.	Should	Workshop
UR022		The user is able to automatically remove disinformation content on social media platforms.	Won't	Workshop
UR023		The user is able to measure the effectiveness of anti-disinformation campaigns.	Could	Workshop
UR024		The user is able to detect deepfake videos related to disinformation.	Could	Workshop
UR025		The user is able to verify the authenticity of images and videos related to disinformation.	Could	Workshop
UR026		The user is able to easily handle an AI-based tool to reliably predict the scope of disinformation-induced crimes.	Should	Workshop
UR027		The user is able to predict which kind of crimes the D&FN will eventually lead to.	Should	All Sources
UR028		The user is able to assess community resilience based on community behavioural profiles and socioeconomic analysis.	Should	All Sources

UR029		The user is able to evaluate the impact of disinformation campaigns on public opinion.	Should	Workshop
UR031		The user should be able to access accurate information regarding offline crimes stemming from D&FN campaigns, improved through incoming data collected from different LEAs/sources.	Should	Workshop
UR032		The user is able to use a software tool to predict the likelihood of an individual sharing disinformation on social media.	Could	Workshop
UR033		The user is able to measure the reach and impact of disinformation campaigns on social media (i.e., X).	Must	All Sources
UR034		The user has the ability to access personal information of social media users.	Should	Workshop
UR035		The user is able to use the platform in a user-friendly way.	Must	All Sources
UR036		The user complies with relevant data protection and privacy regulations while using the platform.	Must	Workshop
UR037		The user is able to process and analyse large volumes of data from various sources, including social media platforms through the utilisation of the FERMI platform	Must	All Sources
UR038		The user is able to provide near real-time alerts and notifications to law enforcement officers when new threats are detected. The alerts should be customised based on the user's preferences and job responsibilities.	Should	All Sources

2 FERMI societal landscape: Setting the baseline of Societal readiness and digital trust

This section of the deliverable addresses T2.2, which requires "finding a fair balance of interests between law enforcement objectives and the protection of fundamental rights and democratic values, including freedom of speech, freedom of thought, conscience and religion, respect of private and family life, and freedom of assembly and association."[10] In full accordance with the outline of FERMI, this balance concerns "regulatory gaps related to the spread of fake news online as well as the impact" thereof.[11]

In other words, LEAs and other players in and out of government may be concerned about (certain forms of) D&FN and aspire to embark on mitigation measures that, however, may easily run counter to the above-mentioned norms and principles. Considering that D&FN is a form of expression, any attempt at mitigating such expressions interferes with provisions protecting fundamental rights such as freedom of expression and, more broadly, the freedom to express conscience and religious beliefs, both in private and public settings, as well as freedom of assembly and association.

As further required by the GA, the aforementioned focal points are to be discussed through the lens of "legal instruments, ethical guidelines, case law and doctrinal research."[12] Accordingly, the ensuing remarks are guided by these demands. However, to put such an analysis in context, an overview of the EU's efforts to come to grips with delineating the terms 'disinformation' and 'fake news' is given beforehand. This amendment, which is not part of the GA, ensures that the subsequent analysis and the insights thereof can address the crucial issues in the current effort to fight D&FN in the EU, which FERMI – being an EU-funded research project – is supposed to advance. Moreover, a coherent project definition of the to-be-examined focal point can then be derived from such a delineation, which can guide the further work of FERMI.

In line with the GA's remark to place a special emphasis on "law enforcement objectives" the following observations particularly address the role of LEAs, the FERMI project's key target group to whom the user requirements and the platform are tailored, as clarified elsewhere in this deliverable. The role of further stakeholders is addressed as well, wherever necessary (social media stakeholders that are particularly affected by recent attempts to rein in the online spread of D&FN are a case in point.[13] Quite tellingly, social media stakeholders are explicitly mentioned in the GA.)[14]

[10] Grant Agreement, PART A, p.8
[11] Grant Agreement, PART A, p.8
[12] Grant Agreement, PART A, p.8
[13] The GA (PART A, p.8) also somewhat vaguely alludes to "societal implications created by FERMI project", which might cover disciplines as different as "law, sociology, economy, political sciences […]". These areas of study are mostly covered by the following remarks, which applies to the legal dimension ("law") but also to basic sociological and political norms, which are analysed to ensure that FERMI's experimentation protocol and overall conceptual approach are in full compliance with those. That being said, the legal, sociological and political "societal implications" of FERMI are discussed in greater detail in WP7 and its deliverables. Thanks to the project's compliance with data protection and ethics constraints, which are meant to guarantee that questionable or even unacceptable and possibly illegal impacts on society and politics are avoided, FERMI's negative "societal implications" are very marginal, if that. The economic framework of FERMI's implications is analysed in detail in the exploitation-related deliverables.
[14] Grant Agreement, PART A, p.8

Subsection 2.1 provides an introductory description of the phenomenon of disinformation, drawn from the analysis of relevant academic literature in different disciplines, as well as from relevant policy documents. Based on such description, it describes the main challenges of translating the features of the phenomenon in a legal definition of disinformation.

Subsection 2.2 describes the EU approach to disinformation, by analysing relevant EU policy documents and providing an overview of the national trends in regulating disinformation across Europe. It highlights the main regulatory challenges, as well as the role played by private entities in limiting the negative effects of disinformation.

Subsection 2.3 identifies the main concerns for fundamental rights and freedoms protected at the European level with regard to measures to tackle disinformation. Relevant decisions of the European Court of Human Rights (ECtHR) and Court of Justice of the European Union (CJEU) are analysed in order to extract general principles that must be respected in order to strike a balance between law enforcement purposes and fundamental rights and freedoms in enforcing measures to limit disinformation.

2.1 The phenomenon of disinformation. Interdisciplinary findings and the definitory challenge

2.1.1 Introduction

The phenomenon of disinformation and fake news has been an area of attention for policymakers across the globe since the revelations about the Russian interference in the United States 2016 elections.[15] A report of 2019 of the University of Oxford providing a global inventory of social media manipulation by governments and other politically involved actors showed that evidence of social media manipulation campaigns could be found in 70 countries in the world. The report highlights that the many issues connected to disinformation as a threat to democratic processes existed long before the use of social media technologies. Nevertheless, the use of social media and the Internet increased the scale and precision of disinformation operations, with a massive impact on societies and democracies.[16]

Disinformation and fake news is commonly understood as news that is fabricated and intended to mislead or deceive the public.[17] It may be aimed to have a political influence, but also a primarily economic motivation. While the distinction between these two categories of reasons behind disinformation campaigns may be clear in theory, in practice such distinction is blurred. Regardless of the hidden agenda behind the spread of D&FN,

[15] Hughes, Waismel-Manor, 'The Macedonian Fake News Industry and the 2016 US Election,' *PS: Political Science & Politics*, 54 (2021), 19-23.

[16] Bradshaw, Howard, *The global disinformation order: 2019 global inventory of organised social media manipulation* (n.2 Working Paper 2019: Project on Computational Propaganda, 2019).

[17] Johnson, Marcellino. *Bad Actors in News Reporting: Tracking News Manipulation by State Actors*. RAND Corporation (2021), p.2; United Nations General Assembly, *Countering disinformation for the promotion and protection of human rights and fundamental freedoms* (Report of the Secretary-General, 2022), p.2; European Union External Action Service, *1st EEAS Report on Foreign Information Manipulation and Interference Threats. Towards a framework for networked defence* (European Union, 2023), p.4.

the phenomenon needs to be interpreted in light of a broader tendency in the media system and in the political landscape. In fact, it characterises the so-called "post-truth" age. The expression refers to an era in which objective facts and the truth are less influential in shaping public opinion that in the past. On the contrary, emotion and personal belief are gaining influence. The trend is associated with a general decline in public trust in institutional figures and their claims.[18]

Behavioural sciences provide some insight about why the exposure to disinformation affects the decision-making processes of individuals. When reading inaccurate information, prior accurate knowledge may facilitate a critical evaluation shielding from the misleading effect of disinformation. The so-called "epistemic cognition" is formed by the ability to acquire prior knowledge and the motivation and skills to support it, in order to activate effective processes of reasoning, problem-solving and behavioural decisions. However, well-functioning epistemic cognition may not constitute a shield strong enough when individuals are exposed to disinformation. First of all, exposure to inaccurate information may cause confusion, even in subjects that possess enough prior knowledge to affirm that such information is inaccurate. This effect has been demonstrated by analysing reading times of individuals of both accurate information and misleading information about well-known facts. After exposure to inaccurate information, for which the reading time is usually longer due to the confusion caused by the misleading content, the individuals tend to require longer reading times even for information that is accurate, and on which they have previous knowledge. This occurrence shows the potential of false information to generate confusion even in subjects with a solid epistemic cognition. As a consequence, people with low confidence in their prior knowledge are in doubt due to the uncertainty they already had about the information they possessed, while people with high confidence are equally influenced due to their belief of not being impacted by false information, which leads them to not accurately evaluate it. This process ultimately brings people to rely on false information and successfully integrate them in their baggage of knowledge, polluting their understanding of reality and decision-making.[19]

As a response to the spreading of fear in many countries that disinformation may constitute a threat to democracy, many governments have adopted countermeasures, including regulatory frameworks for social media platforms. The challenge of adopting policies and laws to address threats deriving from disinformation comes with the need to find a common definition of a very complex phenomenon, and subsequently translate it into an operational definition from a legal standpoint. The search for a common definitory framework is accompanied by many difficulties, which have significant repercussions on the measures adopted to tackle disinformation and their impact on fundamental rights and freedoms at stake.

[18] Buckingham, 'Teaching media in a 'post-truth' age: fake news, media bias and the challenge for media/digital literacy education,' *Culture and Education*, 31 (2019), 213-231.

[19] Rapp, Salovich, 'Can't We Just Disregard Fake News? The Consequences of Exposure to Inaccurate Information,' *Policy Insights from the Behavioral and Brain Sciences*, *5* (2019), 232–239.

2.1.2 The EU approach to the definitory challenge

In 2018, the European Commission tasked a High-Level Expert Group (HLEG) to provide policy suggestions on how to tackle disinformation. The result of the work of the Group was the report "A multi-dimensional approach to disinformation: report of the independent high-level group on fake news and online disinformation", published in the same year. The report aimed to provide a common understanding of disinformation, and subsequently provide insights on how to address the phenomenon at the European level.[20]
The report addressed the problem of defining fake news and disinformation in the policy context. It defines disinformation as **"false, inaccurate or misleading information designed, presented or promoted to intentionally cause public harm or for profit"**. When specifying the concept of public harm, the HLEG referred to threats to democratic values and processes such as elections, which may manifest in a number of relevant sectors, such as health, finance, education. Notably, the report clarified that the term "disinformation" should be preferred to "fake news". First of all, the phenomenon of disinformation is not limited to news that are completely false, but may also refer to forms of content that are presented in a way that promotes a misleading understanding of reality. Also, disinformation may encompass a broad range of online content which is not limited to traditionally intended news, such as audiovisual content, targeted advertising and organised trolling. Moreover, the expression "fake news" has been acquiring a political connotation, due to the use made by politicians and supporters of certain political parties that are aimed at dismissing news that they find disagreeable. Therefore, the term is associated with certain political strategies with the potential to undermine independence of media.[21]
The report also specified that disinformation does not include other illegal forms of speech, such as hate speech, incitement to violence and defamation, which are regulated by already existing legal instruments at the EU level. Also, it does not include expressions such as satire and parody, which purposely distort reality, but do not intend to cause any public harm.[22]
After the publication of the report, the European Commission adopted the communication "Tackling Online Disinformation: a European approach", aimed to build a European approach to tackle online disinformation, the first policy document addressing the phenomenon at the EU level. The Commission adopted a definition of disinformation very similar to that of the report of the HLEG. According to the communication, disinformation is any "**verifiable false or misleading information that is created, presented and disseminated for economic gain or to intentionally deceive the public, and may cause public harm**".[23]
While the definition slightly differed from that adopted by the HLEG, it includes all the key elements of the Report.

[20] European Commission, Directorate-General for Communication Networks, Content and Technology, *A multi-dimensional approach to disinformation: Report of the independent high level group on fake news and online disinformation* (Publications Office of the European Union, 2018).
[21] *Ibidem.*
[22] *Ibidem.*
[23] European Commission, *Tackling Online Disinformation: A European Approach* (*Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, COM/2018/236*). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0236.

Another contribution to the challenge of grasping a definition of disinformation came from the report "Information disorder: Toward an interdisciplinary framework for research and policy making", commissioned in 2017 by the Council of Europe. In the report, the authors Wardle and Derakhsan defined disinformation as any "**information that is false and deliberately created to harm a person, social group, organization or country**". The report distinguished disinformation from misinformation, which was false information, but did not aim to cause any harm. It also excluded from the category of disinformation what was called 'malinformation', which was any true information used to inflict harm to a person, organisation or country.[24] While this distinction is valuable, the definitions of the Commission and of the HLEG report will be used as a reference in order to outline the issues related to define disinformation, as they are the cornerstone of the chosen EU approach to tackle disinformation.

While there are some differences among the definitions of the European Commission and the HLEG, they present three common elements: 1) factual or misleading nature of the information; 2) intention of the actors to obtain economic gain or deceive the public; 3) public harm. With regard to the latter element, the HLEG requires the intent to cause public harm or gain a profit. Instead, the European Commission requires the intent to have an economic gain and deceive the public, and the public harm is only a potential but not a necessary consequence of the dissemination. Besides the positive definition, both HLEG and the Commission have adopted a negative definitory approach, by precising the content already regulated under EU law which is outside the scope of disinformation. Besides the exclusion of expressions like satire and parody, both the communication of the Commission and the HLEG report specify that content already made illegal pursuant to EU law is outside the scope of the definition.[25] Under EU law, four types of content were defined as illegal by means of sector-specific legislation harmonising the law of EU Member States with regard to them: 1) child sexual abuse material; 2) racist and xenophobic hate speech; 3) terrorist content; 4) content infringing Intellectual Property rights. Such expressions are therefore otherwise regulated and do not fall under the EU definition of online disinformation.[26]

In order to be translated in legal terms, a definition of disinformation should not be broad or vague enough to allow arbitrary interpretations from the authorities enforcing the provisions against it. Taking into consideration the definitions adopted in the EU policy debate on disinformation, a number of difficulties may obstacle the elaboration of a definition which can lawfully be used when imposing measures against the spread of disinformation.

[24] Wardle, *Information disorder: Toward an interdisciplinary framework for research and policy making*, Council of Europe (2017).
[25] Ó Fathaigh, Helberger, Appelman, 'The perils of legally defining disinformation,' *Internet policy review*, 10 (2021), 2022-2040.
[26] De Streel et al., *Online Platforms' Moderation of Illegal Content Online: Law, Practices and Options for Reform* (Policy Department for Economic, Scientific and Quality of Life Policies Directorate-General for Internal Policies, European Parliament, 2020).

2.1.3 The legal issues of defining disinformation

The first difficulty in defining disinformation comes from the possible spill-over effects deriving from identifying what can be considered as false or misleading information. An ideal legal definition of disinformation which can be deemed as compliant with human rights laws and standards should be clear and narrow, with precise elements that are not susceptible to be confused with other types of online content. However, it is difficult to build such a clear and narrow definition, as the understanding of what is true in a given context may be challenging. Compared to other forms of illegal content, such as terrorist content and hate speech, disinformation is much more likely to be confused with other types of legal and not harmful content. This difficulty can cause negative consequences for freedom of expression, as it may result in pre-emptive self-censorship of individuals who refrain from sharing opinions and information whose trustworthiness they cannot verify.[27]

Secondly, the intent of the disinformation actors to obtain economic gain or to deceive the public may be very difficult to prove, especially in light of the factor of "public harm" included in the definition of disinformation. Both the HLEG and the European Commission identify the public harm of disinformation in the alteration or undermining of democratic processes, such as elections. Given the large scale of disinformation campaigns, the public harm referred to in the EU approach is often caused by a broad number of users sharing the same content over and over again. Proving the intent of the single subjects participating in the dissemination of disinformation, or even only identifying them in the online ecosystem, is challenging.[28] This is true even when the intent is economic gain or deceiving the public (like in the Commission's definition) and not directly causing the public harm in question (like in the HLEG's definition).

Finally, the definition of disinformation poses issues with regard to the concept of "public harm" caused by it. Studies of democratic theory and the analysis of EU policy documents allow to identify three normative goods that are threatened by disinformation, all connected with the risks for the democratic processes that the policies to tackle disinformation aim to prevent.

The first good threatened by disinformation is self-determination. Self-determination refers to the ability of a people to rule themselves freely from external domination, or from domination of internal elites. Thus, disinformation may affect self-determination when it impairs such ability, or the ability of a people to give themselves rules. An example may be found in how certain disinformation campaigns aim to alter democratic processes, such as elections and public deliberations. With globalisation, foreign actors are inevitably involved in democratic processes of states, due to the interdependencies between countries. However, the concept of self-determination limits the extent of such influence.[29]

Secondly, disinformation may represent a threat to democratic representation and accountability, as it has the potential to influence elections, which enable citizens to select their representatives and hold them accountable.

[27] Pielemeier, 'Disentangling Disinformation: What Makes Regulating Disinformation So Difficult?,' *Utah Law Review*, 917 (2020).
[28] *Ibidem*.
[29] Tenove, 'Protecting democracy from disinformation: Normative threats and policy responses,' *The International Journal of Press/Politics*, 25 (2020), 517-537.

Disinformation may affect elections by spreading false claims and damaging fair competition among candidates and parties. The use of social media platforms and in particular the use of false accounts to spread certain information is also a way to influence democratic processes.[30]

Finally, disinformation may constitute a threat to democratic deliberation, as it may impair exchanges between people that are necessary in order to ensure a well-informed public decision making. Authors argue that, in order to promote deliberation in democracies, epistemic quality, moral respect and democratic inclusion need to be preserved. Disinformation may harm epistemic quality by promoting false claims at a large scale and thus discouraging people from relying on adequate sources of information. Also, it may undermine moral respect toward certain social groups that might struggle to participate in the political discourse. Moreover, the use of social media by actors promoting disinformation may reduce opportunities for citizens to confront themselves with diverse views, affecting their democratic inclusion in the political debate.[31]

In the context of large-scale disinformation campaigns, which rely on a combination of very diverse online content, the isolation and measurement of harm to these normative goods may be challenging, due to the long-lasting effects of disinformation. The uncertainty in defining *ex ante* what type of disinformation may cause the public harm in question may, as a result, lead to overly broad definitions and measures that risk to censor legal content, thus generating chilling effects on freedom of expressions as described above.[32]

2.1.4 Conclusions

D&FN refers to the spreading phenomenon of fabricating news with the aim to deceive the public, with the aim to have a political influence or for an economic gain. Such phenomenon needs to be understood in the broader context of the "post-truth era", an expression that refers to the tendency of attaching more relevance to emotion and personal beliefs than objective facts in shaping personal opinions. This trend is a direct result of a decline of public trust in institutional figures. Behavioural sciences have proved that the exposure to disinformation may affect the decision-making processes of individuals, thus having an impact on public opinions that play a role in democratic processes. The effort of many countries in enforcing measures to tackle disinformation poses the problem of finding a common definition to be used when adopting laws and policies in this domain.

The EU was active in the debate on how to define disinformation. Both the HLEG and the European Commission provided a definition of the phenomenon in related policy documents. **The findings of these documents identified three common elements that should be integrated in the definition of disinformation: 1) factual or misleading nature of the information; 2) intention of the actors to spread such information they know to be false to obtain economic gain or deceive the public; 3) public harm.** However, the translation of such definition in legal terms requires for it not to be excessively vague, so to

[30] *Ibidem*.
[31] *Ibidem*
[32] Pielemeier, 'Disentangling Disinformation: What Makes Regulating Disinformation So Difficult?,' *Utah Law Review*, *917* (2020).

avoid arbitrary interpretations by authorities responsible of tackling disinformation. This objective poses a number of difficulties.

The first difficulty in defining disinformation comes from the possible spill-over effects deriving from identifying what can be considered as false or misleading information, as disinformation is more susceptible to be confused with other legal and not harmful content. The second challenge comes from the necessity to prove the intent of disinformation actors to deceive the public or have an economic gain. To prove this intent is very difficult, especially when considering the "public harm" as an element of the definition, and identifying it as the effect of undermining democratic processes. This effect is often caused by a broad number of actors sharing online content over and over again, thus making it difficult to trace it back to single individuals and their intent. Finally, the identification of what constitute "public harm" also causes some issues. The normative goods threatened by disinformation – self-determination of individuals, democratic representation and accountability, and democratic deliberation – are actually undermined after large-scale disinformation campaigns. Therefore, it is challenging to define *ex ante* what type of disinformation has the potential to cause public harm intended as detrimental to the normative good and questions.

Such uncertainties in legally defining disinformation may lead to overly broad definitions that can risk to censor legal content, creating a chilling effect for freedom of expression, and a negative impact for other fundamental rights. For these reasons, it is important to identify the strengths and pitfalls of the current EU approach to disinformation, and to recognise which fundamental rights can be impacted by measures to limit its spread, in order to provide guidelines to competent authorities on how to tackle D&FN while guaranteeing their respect.

2.2 The EU approach to disinformation

2.2.1 Introduction

The European Union is also active in the policy debate about disinformation. Given the strict link between disinformation campaigns and the use of online platforms to carry them out, the EU approach mainly focuses on online disinformation, which may be considered to fall under the broader category of content moderation policies. However, contrary to other types of content that are considered illegal under EU law, online disinformation is not *per se* illegal. In other words, there are no EU legal instruments prohibiting the spread of disinformation and imposing obligations to take down this type of content. Nevertheless, the EU recognised that disinformation may be harmful with regard to the formation of informed and pluralistic opinions, and damaging to democratic processes. While a legislative initiative was taken at the EU level to make other types of content illegal, online disinformation was so far not subject to EU law. Instead, a number of soft-law documents and co-regulation initiatives with private stakeholders constitute the EU approach to disinformation. The European Commission's above-mentioned decision to commission a study on the phenomenon of disinformation, which resulted in the HLEG's report on the matter, is a case in point. In the

same year, the Commission adopted the communication "Tackling Online Disinformation. A European Approach". Finally, the "Action plan against Disinformation" was adopted by the Commission and the High Representative of the Union for Foreign Affairs and Security Policy, in order to contribute to the discussion in the European Council on how to effectively tackle the challenges of the field.[33] The abovementioned documents' attempt at mitigating disinformation will be further discussed below.

2.2.2 The HLEG report on disinformation. Toward a common understanding of disinformation at the EU level

In identifying the key principles that should guide the effort to tackle disinformation, the HLEG report puts a particular emphasis on the importance of protecting freedom of expression, safeguarded in both the EU and international human rights framework.[34] The document recalls that, pursuant to the EU Charter of fundamental rights, any limitations to freedom of expression must be provided by law, proportionate and justified in order to protect rights and freedoms of others, or to pursue general interest objectives.[35] The attention of the HLEG to the protection of this right is to be underlined, as freedom of expression represents one of the values which necessitate a careful balance with the goal to limit online disinformation.

When suggesting policy actions to tackle disinformation, the report first identifies the general objective of improving transparency in online ecosystems. Transparency is identified as a means to provide users with the necessary knowledge to better assess the veracity of online content. To this end, the report calls upon online platforms to take a number of actions to ensure the transparency of sources of information, the decision-making process behind sponsoring certain content, and privilege high quality content in order to achieve dilution of disinformation. The report also highlights the importance of fact-checking practices. Moreover, the HLEG encourages the enhancement of information sharing by providing privacy-compliant access to datasets regarding disinformation and disinformation actors, in particular with a view to support research on disinformation dynamics. With regard to users, the Group identifies two main sets of actions: on the one hand, it endorses the promotions of policies to increase media literacy and awareness of citizens when dealing with online content. On the other hand, it supports the creation of tools to empower users and enable them to have control over the content displayed as a result of searches and activities online. Furthermore, the Group recalls the importance of preserving press freedom and pluralism and information, calling upon public authorities to commit to supporting policies to safeguards these values, and reaffirming a negative obligation for them with regard to avoiding interferences with media independence. Such actions would guarantee the implementation of a safe online ecosystem which fosters the dissemination of quality information over disinformation. Finally, the report elaborates on a number of future steps in order to evaluate the state of disinformation in Europe,

[33] European Commission, *Action Plan against Disinformation* (Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, 2018).
[34] European Commission, Directorate-General for Communication Networks, Content and Technology, *A multi-dimensional approach to disinformation: Report of the independent high level group on fake news and online disinformation* (Publications Office of the European Union, 2018).
[35] European Union, *Charter of Fundamental Rights of the European Union*, Art. 11.

assess the effectiveness of the measures taken and implement the EU policy to tackle disinformation through a multi-stakeholder effort.[36]

The HLEG report represents the first notable effort within the EU to provide a common theoretical framework on the phenomenon of disinformation, and suggest policy approaches to it. The document identifies a number of key principles that have been recalled by other policy documents addressing the matter, and its definitory efforts have been shaping the reflection around the perils of providing a legal definition of disinformation.

2.2.3 The communication of the European Commission on a European approach to tackle disinformation and the Code of Practice on Disinformation

The communication of the European Commission of 2018 was aimed to build a European approach to tackle online disinformation.[37] The document is the result of a consultation with the High-Level Expert Group, and of the conclusion it drew it is 2018 report. The theoretical framework delineated in the report guided the Commission in drafting the communication, which may be considered as the first of a number of policy initiatives aimed at addressing large-scale disinformation.

The document recognizes disinformation as a threat to the existence of free and independent media, considered as an essential element to guarantee open and democratic societies an effective public participation in the political debate. It focuses, in particular, on the use of new technologies, such as social media, to disseminate disinformation and enhance its negative impact on democracies, due to their potential to become echo chambers for disinformation campaigns. The Commission acknowledged that such campaigns may be perpetrated by both domestic and foreign actors, public or private stakeholders, with severe consequences in terms of internal security and effect on the public debate preceding and influencing policy making.[38]

The communication identifies a number of steps contributing to the spread of disinformation. First, the Commission notes that the creation of disinformation may involve very different types of content. It may consist not only of written articles, but also of false pictures and audiovisual content. Second, the power of amplification through social media and other online media is recognised. For example, the algorithm-based criteria used by social media platforms to disseminate information contribute to sharing of certain content among users more likely to be influenced and attracted by it, thus enhancing polarisation in society. Advertising models based on algorithms facilitate the placement of types of content that appeals to certain categories of users. Also, the use of fake accounts with no authentic user behind them may increase the spread of disinformation online. Third, the communication focuses on the role of users in disseminating disinformation, and the attitude to share content without prior verification of its veracity.[39]

[36] European Commission, Directorate-General for Communication Networks, Content and Technology, *A multi-dimensional approach to disinformation: Report of the independent high level group on fake news and online disinformation* (Publications Office of the European Union, 2018).

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

[38] *Ibidem.*

[39] *Ibidem.*

The communication lists four principles to guide the policies against online disinformation. It encourages the improvement of transparency as regards the origin, production, dissemination of disinformation, in order to detect attempts of manipulation. It also supports the promotion of diversity of information, as to ensure the ability of citizens to make informed decisions. In the third place, it highlights the importance of providing indications about the trustworthiness of information in order to enhance its credibility. Finally, the Commission points out the importance of promoting inclusive and all-encompassing solutions, involving awareness-raising, the collaboration with relevant stakeholders (media groups, online platforms, journalists) and the cooperation with public authorities.[40]

The first action the communication proposes in order to comply with the aforementioned principles is to improve the transparency and accountability of the online ecosystem. To this end, the Commission identifies the online platforms as key actors which should commit to increase their effort to tackle online disinformation. Self-regulation is considered as a valid means through which to achieve this goal. A result of this first goal identified by the European Commission was the adoption of the 2018 Code of Practice on Disinformation, which enshrines self-regulatory standards to tackle online disinformation.[41] Leading players among tech companies and online platforms were therefore called to voluntarily adhere to such code, in order to adopt a coordinated approach to the challenges posed by disinformation. According to the 2018 Code of Practice, the signatories committed to implement policies to scrutinise advertisement placements, including measures to avoid the placement of advertisement content created by disinformation actors through verification tools. The companies adhering to the Code also committed to improve transparency of political advertising enabling its public disclosure. At the same time, the Code includes a specification about the importance of preserving fundamental rights such as the freedom of expression, in order not to impair the political debate. The signatories were also called to intensify their efforts as regards closing fake accounts, and the prevention of misuse of their platforms. The Code also imposed a commitment to empower users by prioritising verifiable and high-quality information, improving transparency about information and targeting technologies in online ecosystems and enhance diversity. Finally, the stakeholders committed to support research by providing access to relevant datasets about disinformation.

The initiative of the Code of Practice was criticised as amounting to a "privatization of censorship": the tool substantially left it to the Internet platforms to achieve a fair balance between the need to moderate disinformation online and the respect of fundamental rights and freedoms. This conclusion was drawn not only considering the powers of online platforms to implement policies to remove problematic content, but also taking into account the commitments of the Code aimed to privilege certain contents over others, on the basis of a claim of authenticity, accuracy and relevance.[42] While assessing the legality of content may be a challenging task for actors playing a role in content moderation, the assessment on its authenticity and accuracy

[40] *Ibidem.*

[41] European Union, *EU Code of Practice on Disinformation* (European Union, 2018). Available at: https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation.

[42] Monti, *The EU Code of Practice on Disinformation and the Risk of the Privatisation of Censorship, in Democracy and Fake News. Routledge,* (2020), 214-225.

may be even more difficult. Moreover, the Code does not refer to any systems to appeal to such decisions based on which certain content might be downgraded and gain much less visibility than other. All in all, the Code of Practice fell within the trend of delegating broad content moderation tasks to private actors with regard to disinformation. This tendency reveals another problematic trait, as private enforcement of content moderation principles is less visible than the governments' intervention, and therefore less accountable, especially when the specific rules guiding the decision-making on content moderation are not delineated clearly, like in the Code.[43]

Furthermore, the 2018 Code of Practice was criticised as it did not live up to its promise to provide access to datasets relevant to research activities in the context of disinformation. In 2020, an assessment on the implementation of the Code was commissioned by the Directorate-General for Communications Networks, Content and Technology of the European Commission. The study stated that the pillar of the Code aimed to empower the research community proved to be the least advanced. The surveyed stakeholders of the research community complained about the still quite ineffective cooperation between researchers and online platforms, and the access to platforms' datasets through privacy-compliant procedures still being arbitrary and episodic.[44]

The necessity to integrate feedbacks on the first code resulted in the initiative to draft the Strengthened Code of Practice on Disinformation, adopted in 2022 to repeal the first version of the code. The 2022 Code of Practice reiterates the objectives already pursued by the first Code, while considerably expanding on the commitments taken in order to achieve them. The section of the new code dedicated to empowering the research community establishes for the signatories to provide prompt access to data on disinformation that are necessary to undertake research, in cooperation with an independent, third-party body overseeing researches and research proposals.[45] It should be noted that the Code specifies the non-application of this commitment to the access by government bodies and law enforcement authorities, which fall outside the scope of data access-related provisions of the instrument.[46] The 2022 Code reaffirms the commitment to empower users by privileging authoritative sources of information. The Code explicitly mentions the use of recommender systems designed to promote these types of news, based on transparent criteria to select such information. While the use of recommender systems to prioritise or deprioritise certain information still causes concerns as regards the respect of freedom of expression and pluralism of information, the new Code envisages a commitment to inform users of enforcement measures performed in order to counter disinformation, and to provide them with an appeal mechanism against such enforcement actions.

Among the other actions envisaged by the communication of the European Commission, the document also aims to strengthen the role of fact checkers in online environment, to foster online accountability by ensuring

[43] Kuczerawy, 'Fighting online disinformation: did the EU Code of Practice forget about freedom of expression?,' *Disinformation and Digital Media as a Challenge for Democracy: European Integration and Democracy Series*, 6 (2019).
[44] Plasilova et al., *Study for the assessment of the implementation of the Code of Practice on Disinformation* (European Commission, 2020). Available at: https://digital-strategy.ec.europa.eu/en/library/study-assessment-implementation-code-practice-disinformation.
[45] European Union, *The Strengthened Code of Practice on Disinformation* (European Union, 2022). Available at: https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation.
[46] *Ibidem*.

traceability of actors creating and disseminating of information, and recur to new emerging technologies to better tackle online disinformation.

Furthermore, the communication poses the objective of supporting Member States in protecting democratic processes and elections from cyberattacks, including disinformation campaigns with the potential to undermine democratic values. Other actions in the document regard the improvement of media literacy, the support for quality journalism and the creation of awareness-raising strategies for the public against disinformation.

2.2.4 The Action Plan of the European Commission against disinformation

Following the adoption of the communication on a European approach to tackle disinformation, the Commission published in the same year an Action Plan against disinformation, jointly with the High Representative for Foreign Affairs and Security Policy.[47] The Action Plan underlines a more explicit link between disinformation and potential harms to democratic processes, by framing the phenomenon of disinformation as a form of hybrid warfare that might possibly undermine the Member States' democracies and the European project as a whole. The document is also clearly directed to public authorities, and calls upon governments and relevant regulatory bodies in the Member States to cooperate with them.

The first pillar of the Action Plan aims to improve the capabilities of EU institutions to tackle disinformation, including by reinforcing threat analyses and intelligence assessments measures. The second pillar aims to improve a coordinated response to disinformation by creating a Rapid Alert System to provide alerts on disinformation campaigns in a timely manner. The tool is conceived to be implemented by strategic communications departments of the Member States, and to enhance the information sharing between competent national authorities and EU bodies. The third pillar of the Action Plan reiterates the importance of mobilising the private sector in the fight against disinformation, and encourages the signatories of the Code of Practice on Disinformation to promptly implement the commitments taken by signing the Code. The fourth pillar aims to increase public awareness on the phenomenon of disinformation, and places a special emphasis on the importance of fact checkers and researchers in this context. Moreover, the Commission commits to keep supporting independent media and journalism in order to ensure pluralism of information and protect citizens from actions aimed to manipulate the public debate.

2.2.5 The national initiatives to tackle disinformation across Europe. The law enforcement involvement

The EU adopted an approach not entailing the adoption of legal instruments aimed to make disinformation illegal across its Member States. Instead, it preferred to regulate the phenomenon through a number of policy documents aimed at establishing general principles that should guide the actions against disinformation, without providing any bindings rules. Moreover, the EU action is characterised by a significant focus on the

[47] European Commission, *Action Plan against Disinformation* (European Union, Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, 2018).

cooperation with private actors, which remain at the center of the main initiatives taken to tackle online disinformation. This tendency is exemplified by the 2018 and 2022 Codes of Practice on Disinformation, establishing a framework of standards to which online platforms and other relevant players may adhere on a voluntary basis.

However, many jurisdictions across the EU and beyond adopted domestic legislation to counter disinformation in the last ten years, establishing provisions to make certain forms of disinformation illegal.

A notable example among EU Member States may be found in France. The "Freedom of Press Law", which dates back in its first version to 1881, prohibits the "publication, dissemination or reproduction, by any means, of fake news" when such fake news are disseminated in bad faith and may disturb public peace. Moreover, in 2018 a new law was enacted against the dissemination of false information. The "Law on the Fight Against Manipulation and Disinformation" imposes obligations upon online platforms to set up a system for users to flag false information. Subsequently, the online platforms are obliged to adopt measures to fight disinformation actors and limit the dissemination of online disinformation, including by enhancing transparency about the origin of such content and promoting content produced by press agencies, radio and television services. Furthermore, three months prior to elections, a judicial authority may order proportionate and necessary measures to fight the deliberate and massive dissemination of false or misleading information online. The motion may be presented before a court by a political party, a candidate or a public prosecutor, and the court must act within 48 hours. Finally, the French National Council on Audiovisual, a regulatory body for radio and television broadcasting, may withdraw a license of an operator under the influence of a foreign State, which broadcasts false information capable of influencing the vote during the elections period, or to cause harm to the fundamental interests of the Nation.[48]

Another example is Lithuanian legislation on false information. Pursuant to the "Law on the provision of Information to the Public", a prohibition may be found regarding the diffusion of disinformation and information that is offensive to a person and degrades human dignity. Under the Lithuanian law, "disinformation" is defined as intentionally disseminated false information".[49]

A number of legislative provisions are enacted across the EU referring to a very similar definition of disinformation, specifically in criminal laws of the Member States. An example is the Maltese Criminal Code, criminalising the malicious spread of false news which is likely to alarm the public opinion or disturb public order or the public peace. The notion of false news with an alarming effect on the public is recalled in the Czech Republic's Criminal Code, which criminalises the act of intentionally causing a threat to the population, or a portion of it, by spreading alarming and false information.[50]

[48] Levush, *Government Responses to Disinformation on Social Media Platforms: Argentina, Australia, Canada, China, Denmark, Egypt, European Union, France, Germany, India, Israel, Mexico, Russian Federation, Sweden, United Arab Emirates, United Kingdom.* The Law Library of Congress, Global Legal Research Directorate (2019).
[49] Ó Fathaigh, Helberger, Appelman, 'The perils of legally defining disinformation,' *Internet policy review*, 10 (2021), 2022-2040.
[50] *Ibidem*.

Similar provisions are enacted in the criminal legislation of Austria, Croatia, Cyprus, Greece, Hungary, Romania and Slovakia.[51]

Outside of the EU, a law was recently adopted by the Turkish legislator concerning the spread of disinformation. A recent amendment to the Turkish Criminal Code criminalises the public dissemination of false or misleading information.[52] In particular, the law targets individuals disseminating such false information about the internal and external security, public order, and public health, in order to create distress, fear and panic among the public and disturb public peace. The law does not provide a definition of false and misleading information, nor makes it explicit references to standards in order to assess whether the goods protected by it are affected by the dissemination of false information in question. This vagueness in the provision has raised concerns among scholars about the legal certainty of the law, and the actual possibility to foresee when certain online content may be censored pursuant to it.[53]

The regulatory approach to disinformation, which often entails the adoption of criminal laws across Europe, results in the involvement of law enforcement authorities to implement the relevant provisions. However, the role of such authorities often goes beyond enforcing formal legislation on disinformation, as they are often deputed to implement governmental measures to counter it. From this point of view, an example can be found in Italy, where prior to the elections of 2018 the Ministry of Interior implemented a system to enable users with reporting the existence of networks spreading false allegations online. The competent law enforcement authorities for online crime, after reviewing the report, were empowered to decide whether to pursue legal actions against the spread of false information. A similar situation occurred in Spain, where the Ministry of Interior announced that law enforcement authorities were deputed to monitor the online ecosystem in order to detect false information. Such initiatives received attention as they raised concerns with regard to fundamental rights, and in particular the freedom of expression. With regard to the Italian online reporting service of disinformation, the UN rapporteur on freedom of expression expressed concerns about the possibility that the system could have a chilling effect for freedom of expression, as it may function as a "pipeline" for criminal prosecutions.[54] The growing involvement of law enforcement authorities in the initiatives to tackle disinformation, along with the many laws adopted in various States and making disinformation online illegal, calls for a careful assessment about the compliance of such measures with fundamental rights and the rule of law.

[51] *Ibidem*.

[52] Navarro, 'Free Speech: A Right in Crisis as Turkish Parliament Passes New "Disinformation" Bill,' *CICLR Online, 64* (2023), Available at: https://larc.cardozo.yu.edu/ciclr-online/64/.

[53] Yildirim, 'Silenced, Chilled, and Jailed,' *Verfassungsblog on matters constitutional* (2022). Available at: https://verfassungsblog.de/silenced-chilled-and-jailed/.

[54] van Hoboken, O. Fathaigh, 'Regulating Disinformation in Europe: Implications for Speech and Privacy,' *UC Irvine Journal of International, Transnational, and Comparative Law*, 6 (2021).

2.2.6 Conclusions

The EU approach to disinformation represents a compromise between self-regulation of online environments and the adoption of national or supranational laws to address the phenomenon. On the one hand, self-regulation carries a number of issues deriving from the delicate position of content moderation in relation to fundamental rights. Leaving the regulatory power to tackle online disinformation to Internet platforms would result in a variety of different and uncoherent approaches to the problem, with huge discretion left to private stakeholders which are not as bound to the respect of constitutional rights and freedoms as public authorities. The EU took a step to surpass the self-regulation scheme with the endorsement of the Code of Practice on Disinformation, first in 2018 and later in 2022. This initiative falls under the umbrella of what can be defined "audited self-regulation".[55] The Code of Practice enshrines standards to tackle online disinformation to which the players in the industry may voluntarily adhere. The Code of Practice 2022 includes a commitment for signatories to comply with reporting and transparency obligations before the European Commission and other relevant EU regulatory bodies, in order to keep track of the implementation of the Code. Like for any audited self-regulation scheme in the Internet domain, the functioning of this mechanism largely depends on the independence and role of auditors and on the level of commitment of private actors involved in the regulating operation.[56]

On the other hand, statutory regulation requires the direct involvement of states through adoption of legal instruments or measures to fight online disinformation.[57] The adoption of this scheme may be witnessed in a number of EU national jurisdictions, where the legislators have adopted laws or other administrative measures putting in charge public authorities of tackling online disinformation. The statutory regulation of disinformation may have the merit to not leave the definition and regulation of disinformation to different private actors in the online ecosystem. However, the delegation by public authorities to online platforms in order to promptly detect and remove disinformation is not excluded by the statutory regulation scheme. Such delegation may force online platforms to make a major use of artificial intelligence techniques to comply with national legislation, to the detriment of an effective safeguard of fundamental rights at stake. Moreover, the difficulties related to legally defining disinformation may raise doubts about the compliance of national legislations with the rule of law, and in particular the requirements of clearness and preciseness of laws imposing sanctions and limiting rights and freedoms.[58] Finally, the involvement of law enforcement authorities in tackling disinformation may create very specific concerns with regard to fundamental rights and freedom. The next two sections are dedicated to exploring these potential issues, in particular concerning freedom of expression and the right to privacy and data protection.

[55] Marsden, Meyer, Brown, 'Platform values and democratic elections: How can the law regulate digital disinformation?,' *Computer law & security review*, 36 (2020).
[56] *Ibidem*.
[57] *Ibidem.*
[58] *Ibidem.*

2.3 Balancing law enforcement purposes with fundamental rights in measures to tackle disinformation

2.3.1 Introduction

Measures to tackle online disinformation have the potential to impact various fundamental rights protected pursuant to both the European Convention of Human Rights (ECHR) and the EU Charter of Fundamental Rights.

The measures aiming to detect and counter disinformation entail an interference with freedom of expression. The protection of freedom of expression as a fundamental right is guaranteed in Article 10 of the ECHR.[59] The right to freedom of expression is also protected pursuant to the EU Charter of Fundamental Rights, namely via Article 11.[60]

The rights to privacy and data protection are also recognised as fundamental rights under both the ECHR and the EU Charter of Fundamental Rights. The right to respect for private and family life, to home and correspondence is recognised under Article 8 of the ECHR.[61] The EU Charter of Fundamental Rights protects the right to respect for private and family life, home and communications pursuant to Article 7.[62] The EU Charter includes a specific provision on the protection of personal data, which is Article 8.[63] The rights to privacy and data protection are particularly important in the context of law enforcement activities to grasp online disinformation. In fact, counter-disinformation measures may lead to the necessity for law enforcement authorities to trace disinformation actors. The personal data processing operations deriving from these measures amount to an interference with the right to privacy and data protection, and should be legitimate and justified based on the protection afforded to this fundamental right in Europe.

Other fundamental rights may also be impacted by measures to limit online disinformation. Freedom of assembly and association is also protected as a fundamental freedom by the ECHR. Pursuant to Article 11 of the ECHR, everyone has a freedom to peacefully assemble and the freedom of association with others. Similarly to freedom of expression, the right may only be restricted when the limitation is prescribed by law, and it is necessary and proportionate in a democratic society, to protect legitimate interests, such as national security, public safety, prevention of disorder or crime, protection of health or morals, or the protection of rights and freedoms of others.[64] The right is also enshrined in the EU Charter, pursuant to Article 12. The article specifies that the right is protected in particular with regard to political, trade union and civic matters.[65] Nowadays, most of the movements organising assemblies and creating associations of individuals rely on online platforms to recruit members and organise their gatherings. Therefore, the safeguards afforded to online freedom of expression are functional to ensure the freedom of assembly and association as well. The activities

[59] European Union, *European Convention of Human Rights*, Art. 10.
[60] European Union, *Charter of Fundamental Rights of the European Union*, Art. 11.
[61] European Union, *European Convention of Human Rights*, Art. 8.
[62] European Union, *Charter of Fundamental Rights of the European Union*, Art. 7.
[63] European Union, *Charter of Fundamental Rights of the European Union*, Art. 8.
[64] European Union, *European Convention of Human Rights*, Art. 11.
[65] European Union, *Charter of Fundamental Rights of the European Union*, Art. 12.

in the context of the FERMI project should take into account this functional link and consider that the same strict requirements to limit freedom of expression also apply to restrictions to the rights under Article 11 ECHR.

Freedom of thought, conscience and religion, enshrined in Article 9 ECHR, is linked to both freedom of expression and freedom of assembly and association. The right includes the freedom to change belief or religion, and freedom to manifest such belief or religion, either alone or in a community, in private or in public. This right can only be limited when prescribed by law, and when the limitation is necessary in a democratic society, in light of interests such as public order, health or morals, public safety, or protection of rights and freedom of others.[66] The freedom of thought, conscience and religion is also included in the EU Charter of Fundamental Rights, which lists worship, teaching, practice and observance activities among those protected pursuant to the right. Moreover, Article 10 of the EU Charter protects the right to conscientious objection, in accordance with domestic law of the EU Member States.[67] Both the freedom of assembly and association, and the freedom of expression, are functional to exercising the freedom of thought, conscience and religion, as online manifestations of conscience and religion can be categorised as protected expressions, and the right to manifest a belief or religion in a community may imply the freedom of association and assembly. Moreover, the rights to privacy and data protection also play a crucial role in ensuring the exercise of the aforementioned rights. In fact, the processing of special categories of personal data, such as data revealing racial and ethnic origin, political opinions, religious or philosophical beliefs or trade union memberships may impact on the ability to freely express political opinions or express their religious or philosophical beliefs. The importance of protecting these sensitive categories of data due to their functionality to the exercise of other rights is confirmed by the heightened protection afforded to them in both the GDPR[68], the Law Enforcement Directive[69], and the Convention 108 for the protection of individuals with regard to the processing of personal data.[70]

The present section enshrines an analysis of the main ECtHR and CJEU judgements on freedom of expression and the rights to privacy and data protection which have a relevance for measures against disinformation The present legal analysis aims to provide the law enforcement community and the FERMI consortium with guidelines and key principles for the law enforcement activities in this realm. As outlined above, the right to freedom of expression and the rights to privacy and data protection are likely to be particularly impacted by

[66] European Union, *European Convention of Human Rights*, Art. 9.
[67] European Union, *Charter of Fundamental Rights of the European Union*, Art. 7.
[68] [68] European Union, *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC*, Art. 9. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016L0680. – Referred to as GDPR as follows.
[69] European Union, *Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA* (Law Enforcement Directive), Art. 10. – Referred to as Law Enforcement Directive as follows.
[70] Convention 108 for the Protection of Individuals with Regard to the Automatic Processing of Individual Data, Art. 6.

content moderation activities, and in particular by measures to limit disinformation. Furthermore, their protection is instrumental to protect other fundamental rights that can be impacted by such measures.

2.3.2 Freedom of expression

Freedom of expression is protected as a fundamental right in different legal texts across Europe. First of all, it is enshrined in Article 10 of the European Convention of Human Rights. The provision states that everyone has a right to freedom of expression, including "the right to hold opinions and to receive and impart information and ideas without interference by public authorities and regardless of frontiers".[71] The same Article establishes a possibility to limit the right. The limitations must be prescribed by law and must be proportionate and necessary to achieve a number of goals in the interest of a democratic society. In particular, the Article allows restrictions of the freedom if these restrictions pursue objectives related to national security, territorial integrity or public safety, the prevention of crimes, the protection of health and morals, the protection of the reputation or rights of others, the prevention of the disclosure of confidential information, or the maintenance of the impartiality and independence of judicial authorities.[72] Similarly, freedom of expression is protected pursuant to the EU Charter of Fundamental Rights. Article 11 of the EU Charter reiterates that this includes holding opinions and receiving and imparting information and ideas without interferences by public authorities and regardless of frontiers. It also specifies that freedom of media and pluralism must be respected.[73]

The ECtHR was very active in defining freedom of expression and putting a clear emphasis on its key role in a democratic society and the necessity to strictly construct any exceptions to the right. It is notable that the court affirmed that not only "information" or "ideas" that are considered inoffensive are accepted as a manifestation of freedom of expression, but also statements that may shock and disturb, in order to guarantee pluralism and tolerance in the society.[74] The importance of this freedom lies it its close connection with the right of people to take part in the cultural, political and social life in the country where they conduct their life. In this sense, freedom of expression guarantees both the possibility to express one's ideas or opinions, but also to receive those of others.[75]

Both in the European legal landscape and at an international level, special attention is dedicated to the expression of political opinions. The ECtHR stated that the very nature of political opinions is often polemical and virulent, and nonetheless their dissemination is in the public interest, as long as such opinions do not constitute incitement to violence, hatred and intolerance.[76] At an international level, political opinions are recognised as deserving of special protection, due to the risks that limitations to expressing such opinions may pose to democracies. In particular, the limitation of political expressions may result in a way for states to target

[71] European Union, *European Convention of Human Rights*, Art. 10.
[72] European Union, *European Convention of Human Rights*, Art. 10.2.
[73] European Union, *Charter of Fundamental Rights of the European Union*, Art. 11.
[74] The Observer and The Guardian v. United Kingdom App no. 13585/88, (ECtHR 26 November 1991).
[75] Pustorino, *Introduction to International Human Rights Law* (Springer Nature, 2023).
[76] Baldassi and Others v. France App no. 15271/16, 15280/16, 15282/16 et al., (ECtHR 11 June 2020).

political dissenters, or silence minorities, in the name of preserving law and order. For example, the ground of the protection of morals may be used to impose a majoritarian conception of morality, to the detriment of minoritarian views that are nonetheless deserving protection in order to ensure a pluralism of ideas that is necessary in democratic societies.[77]

As regards the limitations that can be imposed on freedom of expression, the first condition to restrict it in any way is that such restriction is provided by law. Similarly, to restrictions applied to other fundamental rights, the law providing it should be accessible and the restrictions enshrined in it should be foreseeable for individuals. Also, the restriction should be functional to the protection of significant collective interests, as listed in Article 10 of the ECHR. When assessing whether a restriction of freedom of expression is lawful, the criteria of necessity and proportionality of the measures taken should be considered. On one hand, the measure should be necessary with a view to the protection of the interest at stake. On the other hand, the measures should be proportionate to the objective that the restriction aims to achieve.[78]

A consolidated view in international human rights law also recognised an external limit to freedom of expression. In fact, it can be limited through the prohibition of incitement to violence, hatred and racial discrimination.[79] Under EU law, this principle is exemplified by two laws currently in force that establish limits to freedom of expression in online environments. The Council Decision 2008/913/JHA was adopted in order to fight against various forms of expressions underlying racism and xenophobia. Among others, the instrument imposes on Member States to prohibit forms of expressions consisting of publicly inciting to violence or hatred directed against a group of persons, or a member of the group, based on race, color, religion or origin.[80] Another example is Regulation (EU) 2021/784, which was enforced in order to fight the dissemination of terrorist content online by imposing rules on hosting service providers to take down this type of content.[81] When addressing the balance between the need to tackle terrorist propaganda and freedom of expression, the Regulation points out that "the expression of radical, polemic or controversial views in the public debate on sensitive political questions should not be considered to be terrorist content". It is noteworthy that, notwithstanding the many references in the law to the importance of safeguarding freedom of expression, the Regulation gave rise to many criticisms prior to its adoption, due to concerns with regard to freedom of expression. In particular, the definition of terrorist content enshrined in the Regulation was considered very broad, with the potential to include radical or problematic content, that is nevertheless legal and deserving the protection afforded to freedom of expression.[82] Other concerns derived from the provisions of the Regulation possibly leading to a massive use of artificial intelligence to detect and remove terrorist content, and the

[77] Gunatilleke, 'Justifying limitations on the freedom of expression,' *Human Rights Review*, 22 (2021), 91-108.
[78] Pustorino, *Introduction to International Human Rights Law* (Springer Nature, 2023).
[79] *Ibidem*.
[80] European Union, *Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law*.
[81] European Union, *Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online*.
[82] Kuczerawy, *The proposed Regulation on preventing the dissemination of terrorist content online: safeguards and risks for freedom of expression* (Center for Democracy and Technology, 2018).

consequent risks for erroneous removals and difficulties in overseeing the decision-making procedures behind the taking down of certain content.[83]

2.3.3 Freedom of expression concerns in the measures to tackle disinformation

Across Europe, many legislators have been active in adopting legislative and administrative measures to address disinformation. Such measures amount to an interference with freedom of expression. The COVID pandemic has increased the proportion of this phenomenon, as many governments have adopted stringent rules to counter disinformation in response to the severity of the health crisis. While online disinformation may pose a significant threat to democracies, these measures constitute a risk for fundamental rights, as they may cause a "chilling effect" for freedom of expression. A chilling effect is understood as a negative effect that a state action may have, resulting in dissuading persons from exercising their rights or fulfilling professional obligations, for fear of being subject to state actions resulting in sanctions or also informal consequences, such as threats and attacks.[84] The measures adopted to fight online disinformation may result in the practice of self-censorship for citizens and for special categories of professionals, such as journalists, that play a key role in ensuring the maintenance of democracies. As reported by human rights organisations such as Amnesty International, many governments took advantage of measures to tackle disinformation in order to sanction journalists and social media users expressing political dissent and criticising the governments in question. These occurrences are often possible because of the vagueness of the legal definitions of disinformation that make this type of content illegal. The vagueness and excessive broadness of such definitions enables state actors to interpret them arbitrarily, with serious risks for the fundamental rights at stake.[85]

A trend to delegate functions to online platforms in order to tackle disinformation may also be observed, by imposing obligations upon these private actors with regard to detecting and removing content labelled as disinformation. This delegation is not without controversies. The lawmakers delegate to private entities the difficult judgement necessary in order to find a fair balance between the need to take down disinformation-related content and safeguard freedom of expression. This implies that the online platforms should foresee effects and intents of the dissemination of certain content. Moreover, in practice, content moderation online is largely carried out through algorithmic processes. The use of algorithms in content moderation makes it difficult to assess on a case-by-case basis whether a removal of online disinformation is made in compliance with the rule of law, and duly appreciating the context and circumstances of the spread of potentially harmful content.[86]

[83] *Ibidem*.
[84] Pech, *Concept of Chilling Effect: Its Untapped Potential to Better Protect Democracy, the Rule of Law, and Fundamental Rights in the EU* (Open Society Foundations, 2021).
[85] Vese, 'Governing fake news: the regulation of social media and the right to freedom of expression in the era of emergency,' *European Journal of Risk regulation*, 13 (2022), 477-513.
[86] Castets-Renard, 'Algorithmic Content Moderation on Social Media in EU Law: Illusion of Perfect Enforcement,' *University of Illinois Journal of Law, Technology & Policy*, 283 (2020).

From a substantial point of view, the vagueness of rules on disinformation may lead to divergent applications from the various online platforms. Furthermore, time and resources constraints of human content moderators, when involved in the process, may also play a role in the difficulty of uniform application, along with a lack of knowledge that would enable them to contextualise online content in order to objectively judge whether or not it constitutes disinformation. This difficulty in handling content moderation has pushed online platforms to make increasing use of algorithmic tool to automatize content moderation. These solutions are also adopted due to the strict timeframe imposed by certain national legislations to comply with take-down orders. However, the ability of these tools to effectively moderate content without producing unwanted discriminatory effects has not yet been fully demonstrated. Besides, even when online platforms act in order to comply with state legislation, the opacity of their moderating operations remains a point of concern, in relation to the possibility to truly oversee them handling forms of online disinformation. The effective functioning of remedies in place to complaint against decision to remove certain content is also a problematic element in the role of online platforms in this context.[87]

2.3.4 The balance between freedom of expression and the need to tackle online disinformation. Lessons from the ECtHR and CJEU case law

As observed, measures involving law enforcement to fight disinformation must entail narrow and precise limitations to freedom of expression. The public need to tackle online disinformation needs to be balanced with the importance of free speech, and in particular the expression of political opinions, in democratic societies. Also, limitations to freedom of expression that are too vague or too broad might cause a chilling effect on the freedom of expression, impairing the expression of legal content in case of doubts about its trustworthiness.

The ECtHR provides some guidelines in its case law to assess when a limitation of freedom of expression is acceptable pursuant to Article 10 of the ECHR.

First of all, it is important to reiterate that Article 10 ECHR does not only protect ideas that are perceived as agreeable and inoffensive, but also those that may offend, shock or disturb the state or a sector of the population. This broad understanding of freedom of expression serves the purpose to preserve pluralism, tolerance and broadmindedness in democracies.[88] Moreover, the Court has specified that freedom of expression also covers the dissemination of information strongly suspected to be untruthful. In fact, stating otherwise would deprive people of the right to express an idea or opinion about what they read in the media, unreasonably limiting the freedom of expression.[89]

At the same time, Article 10 of the ECHR allows states to impose limitations on freedom of expression when they are prescribed by law and necessary and proportionate in a democratic society. Such limitation may be

[87] Sander, 'Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation,' 43 *Fordham International Law Journal*, 939 (2020).
[88] Handyside v United Kingdom App no. 5493/72 (ECtHR 7 December 1976).
[89] Salov v. Ukraine App no. 65518/01 (ECtHR 6 September 2005).

considered as lawful when responding to collectively relevant interests, including other fundamental rights at stake, but also national security, public safety and prevention of public disorder and crimes.[90]

With regard to the lawfulness of the interference with the freedom of expression, it is important to recall that the prescription of the limitation must be based on domestic law. Such domestic law should be accessible to the public, for example through the publication in the official gazette of the country imposing the restriction.[91] In the case of measures requiring law enforcement authorities to act against online disinformation, a legal basis must therefore be adopted by the legislative authorities in the country, in order to empower the former to act and restraint any forms of expression.

The ECtHR took a very stringent position about the falsification of history by way of Holocaust denial claims, in the case Garaudy v. France.[92] The Court held that Holocaust deniers' online campaigns were ultimately aimed to incite racial hatred against victims of Holocaust and defaming them. In this context, the it found that Article 17 of ECHR was infringed, prohibiting the destruction of rights and freedoms of others. Therefore, groups going against such provision by falsifying history were not afforded the protection of freedom of expression under Article 10.[93]

The aim to protect others' fundamental rights and freedoms may be compared with that to prevent public disorder. As mentioned above, various national legislations across Europe refer to objectives of avoiding disturbances to public order and public peace as a justification to limit online disinformation. In this regard, the case law of the ECtHR suggests that the ECtHR is reluctant to recognize a justification to limiting freedom of expression on the basis of a vague reference to the protection of public order. The provision should clearly be justified in light of a pressing social need as enshrined in the list of Article 10.2 of the ECHR mentioned above.[94] In this regard, the proportionality test to assess whether a limitation to freedom of expression is lawful is of particular importance. According to the traditional approach to the proportionality test, the state imposing a restriction should first of all pursue a compelling and legitimate interest. Secondly, a rational nexus should be evident between the measure and the protection of the identified interest, meaning that the measure should be suitable to the objective pursued. Third, the measure should be necessary to achieve the objective of public interest, meaning that the achievement of the objective cannot be possible by recurring to alternative actions. Finally, the measure should be proportionate to the objective. From this point of view, the restriction to the freedom should be balanced with the gain in terms of protecting the compelling interests at stake.[95] While in the case of hate speech it is easier to identify the groups negatively affected by false information and the rights infringed by its spread, in the case of public order, anti-disinformation measures are more likely to pursue abstract interests, and, as already explained above (para) the link with the public harm caused by disinformation

[90] European Union, *European Convention of Human Rights*, Art. 10.
[91] NIT S.R.L. v. Republic of Moldova App no. 28470/12 (ECtHR 5 April 2022).
[92] Garaudy v. France App no. 65831/01 (ECtHR 24 June 2003).
[93] *Ibidem*.
[94] Perinçek v. Switzerland App no. 27510/08 (ECtHR 15 October 2015).
[95] Gunatilleke, 'Justifying limitations on the freedom of expression,' *Human Rights Review*, 22 (2021), 91-108.

should be carefully evaluated, taking into account all the challenges implied in the effort of defining the phenomenon.

Finally, it should be noted that political speech has a primary position in the protection of freedom of expression pursuant to the case law of the ECtHR. Considering the frequent connection between disinformation and the threat to political processes during elections in the EU policy on disinformation, it is important to outline the view of the Court with regard to the protection of the expression of political opinions. In this regard, it is acknowledged that, while free elections and the freedom to express political opinions constitute a key element in the functioning of democracies, the two may occasionally conflict and require restrictions on freedom of expression. For example, misleading information shared by candidates in order to mislead voters would not be protected under Article 10 of the ECHR, as long as it is demonstrated that the intent behind the dissemination was indeed to impair the ability of citizens to obtain accurate information with a view to elections. With regard to false allegations, the need to rectify them as soon as possible in order to avoid to mislead voters is recognised by the Court. However, the principle of fairness should still apply in the procedures to take down said content.[96] In this context, it is questionable whether national provisions ordering the take-down of content in an extremely short timeframe ensure the required procedural guarantees.[97] Conclusively, with particular regard to political speech, a lawful restriction of freedom of expression should ensure, through an appropriate and fair procedure, the possibility to verify the link between the false information and the caused harm to democratic processes, as well as the assessment about the non-veracity of the information itself.[98]

The CJEU did not directly address the question of the balance between measures to tackle online disinformation and freedom of expression. However, the CJEU analysed the legality of measures to limit online disinformation in the Baltic Media Alliance case, concerning the decision of the Lithuanian radio and Television Commission to require that a channel targeting the Russian-speaking minority in Lithuania could only be broadcasted in pay-TV-packages.[99] The licence to broadcast the programme was held by a company registered in the United Kingdom, and the restriction of the programme was enforced in order to counter the diffusion of false information aimed at destabilising the Lithuanian state. In particular, the channel was held responsible for disseminating information on the collaboration of Lithuanians with the perpetrators of the Holocaust and on the spread of neo-Nazi internal policies.[100] The decision revolved around the question whether such measure would be in compliance with Article 3 of the Audio-visual Media Services Directive,[101] which imposes an obligation to Member States to ensure freedom of reception and not to restrict

[96] Brzeziński v. Poland App no. 47542/07 (ECtHR 25 July 2019).
[97] Bayer et al., *The fight against disinformation and the right to freedom of expression* (Policy Department for Citizens' Rights and Constitutional Affairs Directorate-General for Internal Policies, European Parliament, 2021).
[98] van Hoboken, O. Fathaigh, 'Regulating Disinformation in Europe: Implications for Speech and Privacy,' *UC Irvine Journal of International, Transnational, and Comparative Law*, 6 (2021).
[99] C-622/17, Baltic Media Alliance v. Lietuvos radijo [2019] ECLI:EU:C:2019:566.
[100] C-622/17, Baltic Media Alliance v. Lietuvos radijo [2019] ECLI:EU:C:2019:566, par. 79.
[101] European Union, *Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (AVMSD), OJ L 95 of 15 April 2010*.

retransmissions on their territories of audio-visual media services from other Member States. In this regard, the Court decided that the provision in question was not applicable to the case, as the broadcasting of the programme was limited by certain conditions established in the domestic law of the Member State, but its retransmission as such was not prohibited in its territory.[102] While the case does not refer specifically to striking a balance between fighting disinformation and preserving freedom of expression, scholars have argued that the judgement suggests a less stringent approach to measures limiting disinformation, rather than to those removing this type of content altogether. In other words, the measures are more likely to be considered proportionate to the legitimate aim pursued when the dissemination of disinformation is only limited, but a direct removal of it is not foreseen by anti-disinformation measures.[103]

The General Court, one of the courts forming part of the CJEU and competent to decide on actions taken against institutions of the EU by individuals or Member States[104], had the chance to address a matter concerning the balance between freedom of expression and the necessity to curb harmful information in the case Kiselev v. Council of the European Union.[105] The case concerned a Decision[106] and a Regulation[107] adopted by the Council of the EU establishing restrictive measures against actions undermining or threatening the territorial integrity, sovereignty and Independence of Ukraine. The applicant was the Head of the Russian Federal State news agency "Rossiya Segodnya", which played a central role in disseminating the government propaganda in favour of deploying Russian forces in Ukraine. Based on his active support of such campaign, he was included in the list of individuals subject to the established restrictive measures in 2014.[108] The Court assessed whether the measure amounted to a legitimate limitation of the right to freedom of expression of the applicant. It recalled that, according to its case law, the right to freedom of expression can be limited when a) the limitation has a legal basis in EU or Member States' law; b) the limitation is aimed to achieve an objective of general interest, recognised as such by the EU; c) the limitation must not be excessive.[109] The Court argued that, while the notion of "active support" to policies aimed at destabilising Ukraine was not enshrined in the measure, it was clear from the legal basis of the measure that the actions of the applicant would fall under the its scope of application. In fact, due to the pivotal role played by media in modern societies, his actions clearly had the potential to have a remarkable impact.[110] With regard to the pursuit of an objective of general interest, the Court held that the measures were aimed at exerting pressure on the Russian government in order to stop policies and actions to destabilise Ukraine. In turn, such measures were therefore justified by the need to

[102] C-622/17, Baltic Media Alliance v. Lietuvos radijo [2019] ECLI:EU:C:2019:566, par. 81.
[103] Bayer et al., *The fight against disinformation and the right to freedom of expression* (Policy Department for Citizens' Rights and Constitutional Affairs Directorate-General for Internal Policies, European Parliament, 2021).
[104] European Union, *The Treaty on the Functioning of the European Union*, Art. 263.
[105] 8 T-262/15, Kiselev v. Council [2017] ECLI:EU:T:2017:392.
[106] European Union, *Council Regulation (EU) No 269/2014 concerning restrictive measures in respect of actions undermining or threatening the territorial integrity, sovereignty and independence of Ukraine* (2014).
[107] European Union, *Council Regulation (EU) No 269/2014 concerning restrictive measures in respect of actions undermining or threatening the territorial integrity, sovereignty and independence of Ukraine* (2014), p. 1.
[108] Kiselev v. Council (n. 91), par. 3.
[109] Kiselev v. Council (n. 91), par. 69.
[110] Kiselev v. Council (n. 91), par. 76-77.

preserve peace and strengthen international security.[111] With regard to the non-excessiveness of the measures, the Court held the measures in question could not be considered as disproportionate with respect to the pursued goal of public interest. In fact, the applicant actively engaged in propaganda that could result in destabilising Ukraine, in light of a previous decision of the Russian Public Collegium for Press Complaints, affirming that the applicant had disseminated propaganda against Ukraine contrary to social responsibility, harm minimisation, truth, impartiality and justice, in order to manipulate the Russian public opinion on the matter.[112] His programmes were also defined as war propaganda by the Latvian National Electronics Mass Media Council.[113] It can be noted that the decision of the General Court strongly focused on establishing a causal link between the information disseminated by the applicant and the possible negative consequences on the stability and integrity of Ukraine, and the previous decisions by competent bodies with regard to his propaganda-oriented activities were a crucial element in the decision of upholding the validity of the measures taken. While the decision does not expressly address the dissemination of disinformation, a parallelism may be drawn with regard to the necessity to directly link the dissemination of disinformation and the public harm that may be caused by it.

2.3.5 The rights to privacy and data protection

The right to privacy and data protection is protected both at an international and EU level. The right to respect for private and family life, to home and correspondence is recognised under Article 8 of the ECHR.[114] Similarly, the EU Charter of Fundamental Rights protects the right to respect for private and family life, home and communications pursuant to Article 7.[115] Contrary to the ECHR, the EU Charter also includes a specific provision on the protection of personal data. Article 8 of the Charter establishes a right for individuals to the protection of their personal data. This right is substantiated in the obligation to process data fairly and for specified purposes. Article 8 of the Charter also demands that personal data processing should always be carried out on the basis of a lawful basis, which may be either the consent of the data subject to or another legitimate basis provided by law. Moreover, the provision guarantees a right for individuals to access their personal data that have been collected, and to request their rectification. Finally, Article 8 of the Charter also imposes that an independent authority should be in charge of overseeing the compliance with the provision.[116] Notwithstanding the absence of an explicit reference to a right to data protection in the ECHR, the ECtHR has been active in recognising this right as part of the right to private and family life pursuant to Article 8 of the ECHR. The jurisprudence of the ECtHR reiterates the definition of personal data as any information related to an identified or identifiable individual, by making reference to the Convention no. 108 for the protection of individuals with regard to automated processing of personal data, which was adopted in 1981 by the Council

[111] Kiselev v. Council (n. 91), par. 80-81.
[112] Kiselev v. Council (n. 91), par. 98.
[113] Kiselev v. Council (n. 91), par. 105.
[114] European Union, *European Convention of Human Rights*, Art. 8.
[115] European Union, *Charter of Fundamental Rights of the European Union*, Art. 7.
[116] European Union, *Charter of Fundamental Rights of the European Union*, Art. 8.

of Europe and updated in 2018.[117] Similarly to what is established under the EU data protection framework, the Court recognises that personal data may take very different forms, as it is considered as any information that may directly or indirectly lead to the identification of a natural person. In the ECtHR case law, a broad number of measures and operations are considered as an interference with the right to data protection, which may vary in its severity depending on the context and the objectives justifying the processing. For example, the collection of the information of a subscriber associated with a specific individual's dynamic IP address from an Internet provider by the police is considered as an interference with the right to data protection.[118] While the reasonable expectation to private life is considered as one of the elements that may be considered while assessing whether an interference with data protection is lawful, the Court stated that a subscriber not hiding his or her dynamic IP address while using an Internet service does not lower the expectation of the natural person to private life.[119] This point is particularly interesting when it comes to the collection of public information related to a specific individual. For example, the gathering of public information about an individual on his or her political activity has also be considered as a processing of personal data that requires a legitimate purpose and justification. In this regard, the Court affirmed that public information may still be protected as personal data when it is collected in a systematic manner, even if the methods of collection cannot be considered as secret surveillance measures.[120]

The EU legislator took a significant step in defining the right to data protection and regulating it as an autonomous right in its secondary legislation. The General Data Protection Regulation (GDPR), adopted in 2016, corroborates the definition of "personal data" as any information relating to an identified or identifiable natural person. In particular, the person should be identified or identifiable, directly or indirectly, by reference to identifiers such as a name, an identification number, location data, an online identifier. Moreover, information referring to the physical, physiological, genetic, mental, economic, cultural or social identity of a natural person may also be considered as personal data.[121] Subsequently, the right to data protection is very broad in the EU legal framework, as the GDPR affords a number of guarantees to data subjects whose personal data are processed. The definition of personal data is complemented by the definition of "processing", which is as comprehensive as the former. According to the GDPR, a very varied set of operations, including collection, disclosure by transmission, storage, alteration, may amount to processing of personal data.[122] The definitions of "personal data" and "processing" of the GDPR form also part of the Directive (EU) 2016/680 (Law Enforcement Directive), concerning the protection of personal data in the context of prevention, investigation, detection, or prosecution of criminal offences or the execution of criminal penalties.[123] The Law Enforcement Directive complements the GDPR in the EU data protection framework. It represents a sectoral

[117] Convention 108 for the protection of individuals with regard to automated processing of personal data. The Convention entered into force in 1985.
[118] Benedik v. Slovenia, App no. 62357/14 (ECtHR 14 July 2018).
[119] *Ibidem*.
[120] Rotaru v. Romania App no. 28341/95 (ECtHR 4 May 2000).
[121] GDPR, (n. 54), Art. 4(1).
[122] GDPR, (n. 54), Art. 4(2).
[123] GDPR, (n.55).

legal instrument, as it applies in the presence of two cumulative criteria. First of all, it applies when the purpose of the processing of personal data falls within law enforcement objectives. In particular, the processing should be carried out for purposes of prevention, investigation, detection or prosecution of crimes, or the execution of criminal penalties. The second criterion requires that the processing should be performed by competent authorities in the abovementioned activities.[124] The directive specifies that the competent authorities may be either public authorities deputed to law enforcement activities in criminal matters, or any other body entrusted by EU law or Member States' law to exercise public powers and functions related to the fight against criminal offences.[125]

Under both the EU legal framework on data protection and the case law of the ECtHR, special categories of data are afforded a higher level of protection, due to their sensitiveness. Under both the GDPR and the Law Enforcement Directive, specific provisions are dedicated to such categories of data, as the processing is in this case subject to more stringent limitations and stronger safeguards. Among the special categories of data, the GDPR and the Law Enforcement Directive list the data revealing political opinions, religious or philosophical beliefs.[126] The same data are also afforded special protection pursuant to Convention 108[127], and are considered as "sensitive" in the view of the ECtHR. In this regard, the Court affirmed that such data should not be processed in accordance with ordinary domestic rules, and national authorities should take into account the need for heightened protection when performing processing operations concerning this information.[128] These data are considered to be particularly sensitive as their processing may result in high risks for fundamental rights and freedoms, such as discriminatory effects. Moreover, the processing of such data without adequate safeguards may cause injury to individuals' dignity, as they attain to the intimate sphere of data subjects, or be detrimental to the presumption of innocence.[129]

2.3.6 Privacy and data protection concerns in the measures to tackle disinformation

When analysing the trends governing the fight against disinformation in the EU, one of the identifiable policy objectives is to enhance access to platforms data and ensure transparency of platforms' ecosystems. The above-mentioned HLEG's report on disinformation addressed access to platforms data in 2018. The report lists a number of suggestions to increase transparency on online platforms, in order to counter disinformation. Transparency is considered as a cornerstone in tackling online disinformation, as it would allow an efficient fact-checking of information and put users in the condition of better evaluate reliability of online content. This goal would be achieved by providing adequate information about the claims made on the Internet, the way and the reasons why they are disseminated and the fundings allowing their spread. To this end, the report

[124] Law Enforcement Directive, (n. 55), Art. 2.
[125] GDPR, (n. 55), Art. 3(7).
[126] GDPR (n. 54), Art. 9, Law Enforcement Directive (n. 55), Art. 10.
[127] Convention 108 for the Protection of Individuals with Regard to the Automatic Processing of Individual Data, Art. 6.
[128] Catt v. United Kingdom App no. 43514/15 (ECtHR 14 January 2019).
[129] Quinn and Malgieri, 'The Difficulty of Defining Sensitive Data—The Concept of Sensitive Data in the EU Data Protection Framework,' *German Law Journal*, 22 (2021), 1583-1612.

encourages platforms to enable privacy-compliant access to data about disinformation actors, in order to analyse the dynamics behind disinformation and appropriate fact-checking strategies. One of the purposes of the access to platforms' data would be to encourage academics to carry out research on disinformation and provide effective responses to the issue.[130]

Transparency and access to platforms' data are generally considered as an asset in the debate on how to tackle disinformation. However, while disinformation does not constitute an illegal content *per se* at a EU level, a tendency can be observed in the EU Member States and beyond taking legislative initiatives to criminalise certain forms of online disinformation. Such initiatives at a national level have resulted in a growing involvement of law enforcement authorities in detecting and fighting disinformation, and increasing monitoring activities of online environments.[131]

The monitoring activities of online environments by law enforcement authorities in order to tackle disinformation are at the center of a discussion which goes beyond the European borders, and focuses on the implications of the involvement of government and police authorities in online content moderation. While criminal law policies and law enforcement practices may influence content moderation on Internet platforms through takedown order, the existence and functioning of such platforms also shapes the way law enforcement authorities carry out their functions as regards criminal matters. In the United States, similarly to the EU, law enforcement makes increasing use of social media to gather publicly available information for targeted investigation or for intelligence purposes. Social media are used to assess potential risks, or to make connections between different subjects to investigative ends.

In the context of the detection of harmful content online, including disinformation, the gathered information, if retained, can be used for further and separated law enforcement activities. The collection and retention of data from a large number of individuals in the context of content moderation by law enforcement agencies presents the risk of a shift from individualized investigations to large scale suspicions, as the data may be included in law enforcement datasets to be used in future targeted investigations.[132] In this sense, the interferences with the right to data protection imposed by measures to fight disinformation can pose broader risks to the rights and freedoms of individuals, as they can result in operations of mass surveillance that undermine the principle of the presumption of innocence for subjects sharing content on online platforms.

Law enforcement activities to tackle online disinformation can interfere with the right to privacy and data protection, as enshrined in Article 8 of the ECHR. The data protection-related concerns in the area arise from a number of elements that may characterise anti-disinformation measures. First of all, the measures to detect disinformation can include the cooperation with Internet providers handling online platforms, including the collection of data about disinformation actors from them. Secondly, measures against disinformation can

[130] European Commission, Directorate-General for Communication Networks, Content and Technology, *A multi-dimensional approach to disinformation: Report of the independent high level group on fake news and online disinformation* (Publications Office of the European Union, 2018).

[131] van Hoboken, O. Fathaigh, 'Regulating Disinformation in Europe: Implications for Speech and Privacy,' *UC Irvine Journal of International, Transnational, and Comparative Law*, 6 (2021), 9.

[132] Bloch-Wehba ,'Content moderation as surveillance,' *Berkeley Technology Law Journal*, 36 (2021), 1297.

include the collection and storage of publicly available information about individuals in law enforcement datasets. Finally, the interference with data protection can become problematic when the personal data collected with the purpose to fight disinformation pertain to special categories of personal data which are afforded an heightened protection in both EU law and the case law of the ECtHR, such as political opinions and religious or philosophical beliefs. The possibility of targeted law enforcement measures deriving from the collection of these data may have an impact not only on the right to privacy and data protection, but also on freedom of expression, as it can result in a chilling effect with regard to the expression of certain political views or other ideas.

2.3.7 The balance between the right to privacy and data protection and the need to tackle online disinformation. Lessons from the ECtHR and CJEU case law

The case law of the ECtHR offers guidelines with regard to the issues that may arise from the interference with the right to privacy and data protection in enacting measures against online disinformation.

Concerning the collection of data of a subscriber from an Internet provider by the police authorities, the ECtHR identified an infringement of Article 8 of the ECHR for a failure of law enforcement authorities to obtain a court order before accessing a subscriber's information associated with a dynamic IP address. The conclusion was drawn in the context of a case concerning Slovenian law enforcement authorities, and their request to an Internet provider to have access to personal data associated with the IP address The request followed the notification from the Swiss police to the Slovenian authorities that, according to ongoing investigations, such IP address was used to share child sexual abuse material. First of all, the Court clarified that, regardless of the online activity being illegal or not, the individual using the dynamic IP address had a right to act anonymously online. Therefore, the request of the personal data associate with the dynamic IP address amounted to an interference with the right to privacy and data protection. While according to Slovenian criminal procedure it was lawful to acquire data from an Internet provider in the context of a criminal investigation, the Court noted that the absence of a judicial order to collect such data did not guarantee an effective and independent oversight over the law enforcement activities and adequate safeguards against potential abuses of the authorities. Moreover, the lack of clarity on the extent of the retention of the collected data also represented a breach of Article 8 ECHR.[133] The judgement poses general principles that may be applied to the situation where disinformation actors conduct their online activity anonymously, and their traceability depend on the request of data to online platforms. In this case, a judicial order is considered as necessary to perform this interference with the right to privacy and data protection

The issues originating from the possibility for law enforcement authorities to collect data from digital services providers was also subject to the attention of the CJEU. While the Court did not address any matters directly concerning the balance between the rights to privacy and data protection and measures tackling online

[133] Benedik v. Slovenia, App no. 62357/14 (ECtHR 14 July 2018).

disinformation, some general principles established in this context of access requests to digital services providers may be of relevance. In the judgement An Garda Síochána[134], the Court assessed under which conditions a legislative measure adopted by a Member State can establish derogations to the general obligation of electronic communications services to ensure the confidentiality of communications, pursuant to the e-Privacy Directive.[135] The EU law in question represents a *lex specialis* with respect to the GDPR, as it refers specifically to the right to privacy and confidentiality in the electronic communications sector.[136] The e-Privacy Directive establishes a general obligation upon Member States to ensure that providers of publicly available electronic communications services only retain traffic and location data of users for as long as needed for the transmission of the communications.[137] Exceptions to this general rule may be established by Member States when measures restricting these rights constitute a necessary, appropriate and proportionate measure within a democratic society to safeguard national security, defence, public security, and the prevention, investigation, detection or prosecution of crimes.[138] In this regard, the Court recalled the already established principle according to which, in case of particularly serious offenses committed online, the access request to an IP address may be the only way to investigate such crimes. To this end, the example of child pornography offenses is mentioned by the Court.[139] According to the CJEU, retention of traffic or location data may be performed in order to pursue serious legitimate interests, such as to face serious and foreseeable threats to national security or for the investigation of serious crimes.[140] A subsequent request of access to such retained data should be linked to the legitimate interest justifying the retention.[141] However, the CJEU also specified that a police officer cannot be responsible for assessing suspicions and needs leading to the access request of certain data. In fact, a police officer would not fulfil the requirements of independence and impartiality which characterize a court or an independent administrative body. The latter should only be entrusted to perform a prior review of an access request and carefully balance all the interests and rights involved.[142] This is also in light of the necessity to ensure that such retention and access measures are accompanied by appropriate safeguards with regard to the risk of abuse for individuals.[143] These principles may be applied to the case of measures against online disinformation. The CJEU, similarly to the ECHR, confirms the necessity of a prior review of an independent body of any request to access data retained by electronic communications services providers. In the case of national legislative measures providing for the possibility to access such data with a view to trace disinformation actors, and under the condition that such legislative measures are justified by duly relevant

[134] C-140/20, G.D. v. The Commissioner of the An Garda Síochána and others [2022], ECLI:EU:C:2022:258
[135] European Union, *Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) (e-Privacy Directive)*. Available at: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32002L0058. Referred to as E-Privacy Directive as follows.
[136] E-Privacy Directive, Art. 1.
[137] E-Privacy Directive, Art. 6-9.
[138] E-Privacy Directive, Art. 15.
[139] G.D. v. The Commissioner of the An Garda Síochána (n. 120), par. 73.
[140] G.D. v. The Commissioner of the An Garda Síochána (n. 120), par. 101.
[141] G.D. v. The Commissioner of the An Garda Síochána (n. 120), par. 87.
[142] G.D. v. The Commissioner of the An Garda Síochána (n. 120), par. 107.
[143] G.D. v. The Commissioner of the An Garda Síochána (n. 120), par. 101.

national security or serious crimes-related purposes, this condition must be met. Another concern related to anti-disinformation measures may be the collection by law enforcement authorities of publicly available information. In this regard, the ECtHR found a violation of Article 8 ECHR in relation to the collection of personal data of a peaceful activist never convicted of any offenses, and its unlimited retention in an "extremism database" of the UK police. Pursuant to a request of the data subject to access his personal data as collected by law enforcement authorities, he had found that said authorities had collected data about his participation in various trade union protests as well as pro-Gaza demonstrations. The objective of the collection was to prevent disorder and crime, and such objective was not considered as unlawful by the Court. Moreover, the Court found justified the collection of data in the case, as the applicant had taken part to various protests of an association, which tended to become violent. However, the decision also stated that the unlimited retention of personal data was unnecessary and disproportionate, and the lack of provisions requiring a regular review of the dataset in question represented a breach of the guarantees under Article 8. What is also notable is that the Court identified another problem in the collection of personal data to be considered sensitive: as the data on the applicant revealed his political opinions, they would have required stronger protection compared to non-sensitive data. This protection was however not afforded, and the data was retained in spite of the applicant not being actively involved in any criminal offenses.[144]

The case shed light on the collection of publicly available data for storage purposes by police authorities, in a way similar to what could happen in the case of collection of information about individuals sharing disinformation content online. The objective to prevent disorder or crimes may in principle justify an interference with data protection in the context of measures to combat disinformation, if the provisions establishing the measures are not too vague and broad with regard to the objective pursued. However, the retention of such personal data must be justified, necessary and proportionate for the whole storage period. In the absence of convincing justifications for keeping personal data in a due dataset, law enforcement authorities are required to delete the data when they are no more useful to the initial goal of public interest.

Due to the link between disinformation campaigns and the intent to undermine elections or democratic processes at large, the possibility to collect sensitive data attaining to political views while enforcing measures against disinformation is concrete. This is why specific attention should be dedicated to the principles governing the collection of such data in the context of law enforcement activities. In this regard, the ECtHR reiterated the importance of protecting this category of data. The Court specified that, due to the particular sensitiveness of political views, the balance with other legitimate interests should be particularly accurate, as the processing of data pertaining to political opinions represents a more serious interference than the processing of other types of data. This element was highlighted as the applicant in the case was engaging in peaceful protests and trade unions' events, protected pursuant to Article 11 of the ECHR, protecting the freedom of assembly and association. In this sense, the unnecessary and disproportionate retention of data revealing

[144] Catt v. United Kingdom App no. 43514/15 (ECtHR 14 January 2019).

political opinions may have a "chilling effect" on other fundamental rights that are interlinked with the right to privacy and data protection.[145]

2.3.8 Conclusions

The ECtHR provides some guidelines in its case law to assess when a limitation of freedom of expression is acceptable pursuant to Article 10 of the ECHR. Such principles may be useful when establishing principles on how to balance the law enforcement objectives in the anti-disinformation realm and the protection of freedom of expression.

First of all, not only agreeable or inoffensive ideas are protected pursuant to Article 10 ECHR, but also those that may offend, shock or disturb the state or a sector of the population. Moreover, freedom of expression also covers the dissemination of information strongly suspected to be untruthful. However, limitations on freedom of expression are possible when they are prescribed by law and necessary and proportionate in a democratic society to pursue collectively relevant interests, such as national security, public safety and the prevention of public disorder or crimes. Therefore, in the context of measures to tackle online disinformation, the latter must be based on a domestic law accessible to the public, and any action of law enforcement authorities must be based on a legislative measure. The Court is inclined to consider a limitation of freedom of expression lawful in case of expressions aimed to incite hatred against certain sectors of the population, as such expressions would otherwise impair the rights and freedom of others. On the other hand, the Court is more stringent when assessing the lawfulness of measures based on a vague reference to public order, as the social need requiring the suppression of certain expressions should be clearly delineated. In this regard, the legitimate interest pursued by measures against disinformation should present an evident link with the measures adopted, in the sense that the interest in question cannot be achieved by alternative actions, and the interests at stake should be serious enough to justify a proportionate limitation of fundamental rights. From this point of view, the definitory effort of what disinformation is should be considered crucial: the concept of public harm caused by disinformation can play a pivotal role in defining the public interests pursued by anti-disinformation measures. Given the importance of political speech as a form of expression and its primary position in democracies, the ECtHR afford particular protection to this category of expressions. Therefore, it is especially important that the link between any false information and the caused harm to democratic processes can be verified, when anti-disinformation measures require the take-down of political opinions. The non-veracity of such information should also be carefully assessed before taking an action against it.

The CJEU did not directly address the question of the balance between measures to tackle online disinformation and freedom of expression. However, some cases addressed by the EU Court imply that it would be more inclined to justify a limitation of freedom of expression when disinformation is limited in its dissemination, rather than prohibited with consequent removal. Also, the CJEU affords particular importance

[145] *Ibidem*.

to carrying out an analysis in order to verify a link between certain forms of expression and the public harm caused by them. When measures to tackle disinformation disseminated by specific individuals are tackled, the approach of the CJEU suggests that attention should be paid to the behaviour of targeted individuals, their role in society and the use they made of media, in order to assess whether their actions entailing the dissemination of disinformation can actually cause any public harm.

The case law of the ECtHR also offers guidelines with regard to the issues that may arise from the interference with the rights to privacy and data protection in enacting measures against online disinformation. First of all, the Court provides principles guiding the collection of personal data of subscribers from Internet providers by law enforcement authorities in the context of investigations. In this context, the ECtHR established that the request of access to such data, including to a dynamic IP address, amounts to an interference with the rights to privacy and data protection. Besides the necessity of the order being based on a legal basis in domestic law, the Court also considered a judicial order as necessary in order to access the data in question. Moreover, the ECtHR sought to ensure that the extent of the retention of the collected data by law enforcement authorities is clearly defined in time, to prevent abuses. These principles are applicable in the context of an access request in order to access data related to disinformation actors.

The necessity of an oversight mechanism by an administrative independent body or a judicial authority over such personal data access requests to Internet providers was confirmed by the CJEU. in its jurisprudence, the EU Court confirmed that the obligation of electronic communications services providers to retain data for reasons connected to national security or to the prosecution of crimes may be acceptable, where such obligation is necessary and proportionate to the public interest pursued. Likewise, the request of access to such data by competent authorities is justifiable, where linked to the same legitimate interest based on which the retention of data was performed. However, the CJEU also specified that an impartial and independent authority should assess such access requests in order to balance rights and interests at stake, and a police officer does not have the necessary impartiality of independence with respect to the case.

The issues originating from the possibility for law enforcement authorities to collect data from digital services providers was also subject to the attention of the CJEU. While the Court did not address any matters directly related to the balance between the rights to privacy and data protection and measures tackling online disinformation, some general principles established in this context of access requests to digital services providers may be of relevance.

With regard to anti-disinformation measures entailing the collection by law enforcement authorities of publicly available information, the ECtHR also provided some guidelines in its case law. While the Court found that such collection may be justified based on the objectives to prevent disorder and crime, it stated that the illimited retention of such data in a police dataset and the absence of a regular review on the necessity to the retention was contrary to Article 8 ECHR, as unnecessary and disproportionate to the aim pursued. This conclusion was strengthened in relation to data revealing political opinions, which are afforded special protection due to their sensitiveness. Therefore, in the case of measures aimed to investigate disinformation actors by collecting

publicly available information, the limitation of retention policies in police datasets is particularly significant in order to deem the measures lawful. Moreover, the unnecessary and disproportionate retention of data revealing political opinions by law enforcement authorities may cause a chilling effect on other fundamental rights, such as the right to freedom of expression. Therefore, an interference with the protection of such data is considered to be particularly significant and should be given particular attention when enforcing measures against online disinformation.

3 FERMI technology convergence: Functional Requirements and Technical Specifications towards a refined Architectural Design

3.1 Technical Specifications

The derivation of technical specifications from functional and non-functional requirements is an essential step in the development of the FERMI platform and its individual components. Technical specifications define the technical details of how a platform will be implemented and are derived from the functional and non-functional requirements that describe what the foreseen platform is supposed to do (in accordance with the user requirements as identified above, which could all be transformed into functional and non-functional requirements except for a handful of requirements that all raise legal questions in line with section 2 and have therefore been discarded, see below – those mostly are Could-have requirements that are deemed less imperative than other user requirements).[146] For this purpose, the technical team of the project (INTRA, ATOS, INOV, BIGS, UCSC and ITML) started by analysing each functional requirement in detail and identifying the technical components that fulfil each one of them. Based on this analysis, a set of technical specifications that outline the technical details of each WP3 component (those are the ones that will be conceptualised in the next phase of the project and will guide the entire platform) was created along that facilitates their integration into a single platform.

[146] Unlike all other user requirements that either emphasise the need to address illegal proceedings in a disinformation context (albeit the EU officially distinguishes between disinformation campaigns and violations of the law, as clarified in section 2), address patterns that are crucial for the work of LEAs like identifying key actors, tracing back flows of information and figuring out whether a source is a real person or a bot (which are all essential for successfully investigating suspects of illegal activities) or capture platform-specific characteristics like user-friendliness, the discarded end-user requirements exceed these boundaries. Instead, they mostly aim at drawing a line between true and false allegations, which might be of interest to other target audiences but which must not be addressed by LEAs. As clarified in section 2, LEAs may only step in if they are clearly authorised by the law to examine and/or investigate false allegations but they do not have a broad mandate to help distinguish between right and wrong statements. Against this backdrop, UR 024 (The user is able to detect deepfake videos related to disinformation) and UR025 (The user is able to verify the authenticity of images and videos related to disinformation) must not be taken into consideration. The same applies to UR032 (The user is able to use a software tool to predict the likelihood of an individual sharing disinformation on social media), which even goes one step further by emphasising the need to assess whether false claims are likely to be spread and by whom such claims might be made (before they are actually made). The desire of UR023 (The user is able to measure the effectiveness of anti-disinformation campaigns) is not linked to LEA needs either and captures counter-measures against all sorts of disinformation, not just against those that fall into the LEAs' realm of competence. Lastly UR034 (The user has the ability to access personal information of social media users) is understandable in the event a social media user is subject to an investigation or at least extensive monitoring but the requirement's fairly broad language is irreconcilable with the EU's privacy and data protection rights outlined above.

Table 2 FERMI Out of Scope User Requirements List

FERMI Out of Scope User Requirements List	
URID	**User Requirement Description**
UR023	The user is able to measure the effectiveness of anti-disinformation campaigns.
UR024	The user is able to detect deepfake videos related to disinformation.
UR025	The user is able to verify the authenticity of images and videos related to disinformation.
UR032	The user is able to use a software tool to predict the likelihood of an individual sharing disinformation on social media.
UR034	The user has the ability to access personal information of social media users.

Table 3 FERMI Functional Requirements

FERMI Functional Requirements List FR001-FR021		
FR ID	**FR Description**	**Related UR ID**
FR001	***Account analysis:*** The system should have a feature that analyses the X account in question to identify patterns of behaviour that indicate whether the account is a bot or a physical actor.	UR001
FR002	***Follower-to-following ratio calculation:*** The system should be able to calculate the ratio of followers to accounts followed for the account in question.	UR002
FR003	***Profile information analysis***: The system should be able to analyse the profile information of the X account to determine whether it appears to be a bot account or not.	UR001
FR004	***Data Collection:*** The system should collect data related to relevant disinformation campaigns and actors involved in spreading them.	UR002, UR003, UR005, UR008,

	The data may include social media posts and other relevant information.	UR007, UR008, UR016, UR017-B, UR020, UR028, UR029, UR033
FR005	***Data Analysis on actors:*** The system should analyse the collected data to identify connections between different actors involved in disinformation campaigns.	UR002, UR003, UR007, UR008, UR012, UR017-B, UR020, UR021, UR028, UR029, UR033, UR037
FR006	***Data Analysis on crimes:*** The system should analyse the collected data to identify patterns and connections between different crimes and disinformation campaigns.	, UR012, UR016, UR017-A, UR017-B, UR021, UR037
FR007	***Actor Identification:*** The system should allow users to identify key actors involved in spreading disinformation campaigns, including individuals, organisations, and media outlets.	UR002, UR003, UR008
FR008	***Visualisations:*** The system should allow the user to customize and explore the investigation output via the use of interactive visualizations which analyse the data such as graphs, charts and maps. Additionally, through interfaces the system allows the user to modify the investigation process configuration.	UR002, UR007, UR008, UR011, UR012, UR013, UR020, UR029, UR033
FR009	***Export and Sharing:*** The system should allow users to export and share the identified actors and their connections with others, such as by generating reports or exporting data in various formats.	UR002, UR028
FR010	***Resource Allocation Suggestions:*** The system should provide information to the end user to support the allocation of law enforcement resources to mitigate against disinformation induced crimes.	UR004
FR011	***ML Algorithm:*** The system should apply a ML algorithm to the analysed data to estimate the most influential actor, based on various factors such as engagement level and number of followers.	UR008, UR029, UR033
FR012	***Customisation:*** The system should allow users to customise the ML algorithm and visualisation features according to their specific needs, such as by adjusting the weightings of different factors or filtering the data based on specific criteria.	UR008, UR010, UR012, UR029, UR033

FR013	***Category Classification:*** The system could allow the user to assign different categories, such as political, health-related, or other relevant categories, to the disinformation campaigns examined.	UR010
FR014	***User Interface:*** The system should provide a user interface that allows users to view and interact with the categorised posts, such as by browsing or searching for posts in specific categories.	UR010, UR017-B, UR018, UR019, UR021, UR036
FR015	***Sentiment Analysis Algorithm:*** The system should apply a sentiment analysis algorithm to the collected data to determine the emotional polarity of the posts, such as whether they express positive, negative, or neutral sentiment.	UR011
FR016	***Report Generation:*** The system should generate reports based on the analysed data, which include insights, trends, and other relevant information that can help users make informed decisions.	UR012
FR017	***Calculation and analysis of economic impact:*** The system must be able to perform analyses to determine the economic impact of violent extremism caused by disinformation and fake news	UR015, UR016, UR018
FR018	***Risk assessment:*** The system should be able to analyse the data to assess the level of risk in a given community, taking into account factors such as crime rates, socioeconomic status, and other relevant factors.	UR017-B, UR021
FR019	***Risk management:*** The system should be able to provide recommendations and strategies in order for the user to proactively manage risk.	UR017-B, UR021
FR020	***Learning:*** The system models should be able to be retrained to adjust to new incoming data.	UR031
FR021	***Crime Prediction:*** The system should be able to produce reports predicting potential offline crimes stemming from D&FN activities	UR014, UR027, UR038

Table 4 FERMI Non-Functional Requirements

FERMI Non-Functional Requirements List NFR001-NFR008		
NFR ID	**NFR Description**	**Related UR ID**

NFR001	***Performance:*** The system must be able to provide predictions in a timely manner, with minimal delay or lag.	UR017-A, UR026
NFR002	***Information Export Quality:*** Information provided to the user must be insightful, comprehensive, reliable and accurate enough to aid him/her in predicting the environment and context in which the criminal event may occur due to the D&FN	UR017-A, UR018, UR026
NFR003	***Usability:*** The AI-based tool should have a user-friendly interface that is easy to navigate and understand, even for users with limited technical expertise	UR026, UR035
NFR004	***Authentication:*** The system must ensure that only authorized individuals will have access to the platform.	ALL
NFR005	***Authorization:*** The system must ensure that users will have the appropriate permissions to access the platform system data and services.	ALL
NFR006	***Auditing/Logging:*** The system should keep track of user actions and requests and system events for easier detection and prevention on potential security breaches	ALL
NFR007	***Compliance/Privacy:*** All system properties and components must ensure compliance with relevant laws.	ALL
NFR008	***Data Protection:*** All system properties and components must ensure that all data relevant to the project's actions are protected from unauthorized access and their usage compliant with GDPR regulations.	ALL

3.2 The overall FERMI architecture and the components interaction

For the better defining of the FERMI project a blueprint of the total architecture of the different modules is essential, which has been developed and described. To that end we are using the 4+1 (ISO/IEC/IEEE 42010) view model to describe and develop the system architecture from different viewpoints. With this method we can provide a high-level and still complete overview of the architecture from multiple aspects. The 4+1 view model consists of a logical view which is concerned with system functionality, the process view which describes the system processes during the run time behaviour of the system, development view which deals with the system from the programmer's perspective and concerns software management, and the physical view which concerns the topology and connections of the system and is considered the system's engineer point of view. Finally, there is the plus one view which is the scenarios based on the use cases of the system. Additionally, each view of the 4+1 view model has a corresponding diagram. The logical view is

represented through the development of class diagrams that describe the different FERMI modules functional structures. The process view is established through the development of sequence diagrams that map the different processes within the FERMI modules in relation to time. These views were developed in collaboration with the project's technical partners. Furthermore, the system component diagram was developed, which covers the development view. And the physical view is represented in high level by the deployment diagram. Finally, the scenarios' view is satisfied from the user requirements scenarios which can be found in section 4 of this deliverable, which lays out the experimentation protocol's first draft. Through the use case scenarios architectural elements and details which are depicted in the complete architecture views can be identified.

3.2.1 Architectural diagrams

3.2.1.1 Component Diagram

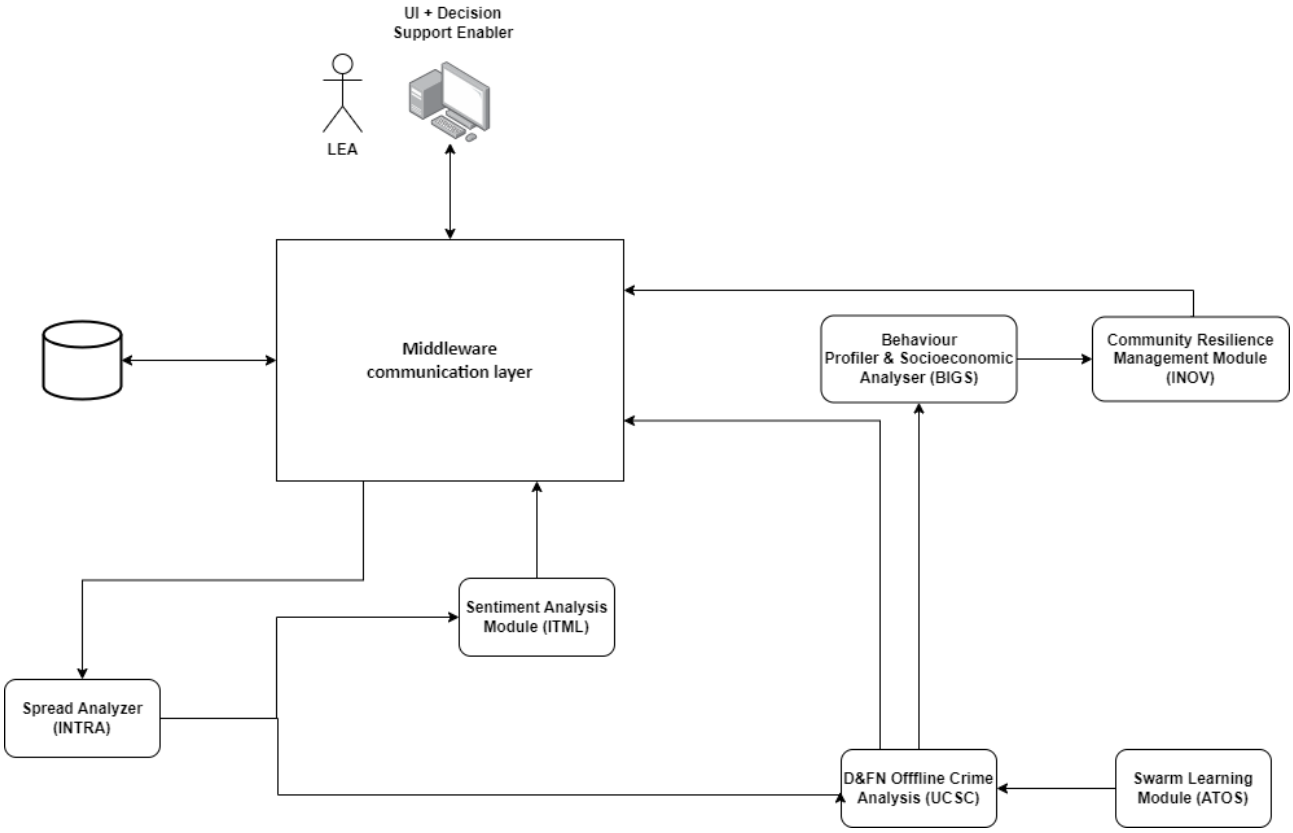


Figure 5 Components Diagram showcasing the interactions between the FERMI components

The component diagram is used to describe the collaboration between different modules of the platform. Each arrow represents the direction of movement for data from one module to another. Inside the modules (grey boxes) the different functional objects are presented. The diagram provides an insight into how components interoperate and what dependencies there are between them. This way the functionality of the modules is presented.

3.2.1.2 Class Diagrams

- UCSC T3.1 – D&FN Offline Crime Analysis

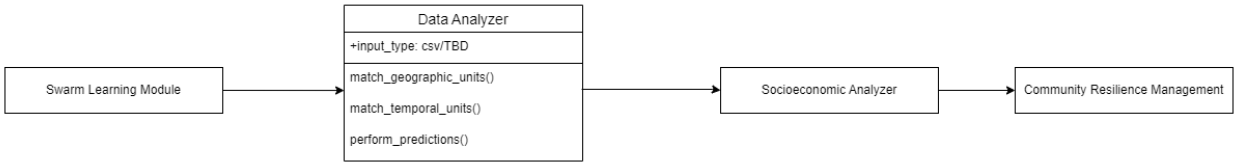


Figure 6 Class Diagram – D&FN Offline Crime Analysis

This Class Diagram represents the high-level view of the D&FN Crime Analysis module structure and internal relationships. The Data Analyser is using the Swarm Learning module to obtain timeseries information on European crimes matching the geographic and temporal units to be exploited in its own crime predictions. The produced knowledge, predictions of the most likely spatiotemporal evolution of D&FN-induced and D&FN-enabled offline crimes, is passed to the Socioeconomic Analyser, which further enriches the information and vehiculates it to the Community Resilience Management Module, and to the FERMI Decision Support Enabler and finally the system's User Interface (UIs).

- INTRA T3.2 – Disinformation, Sources & Spread Analyser

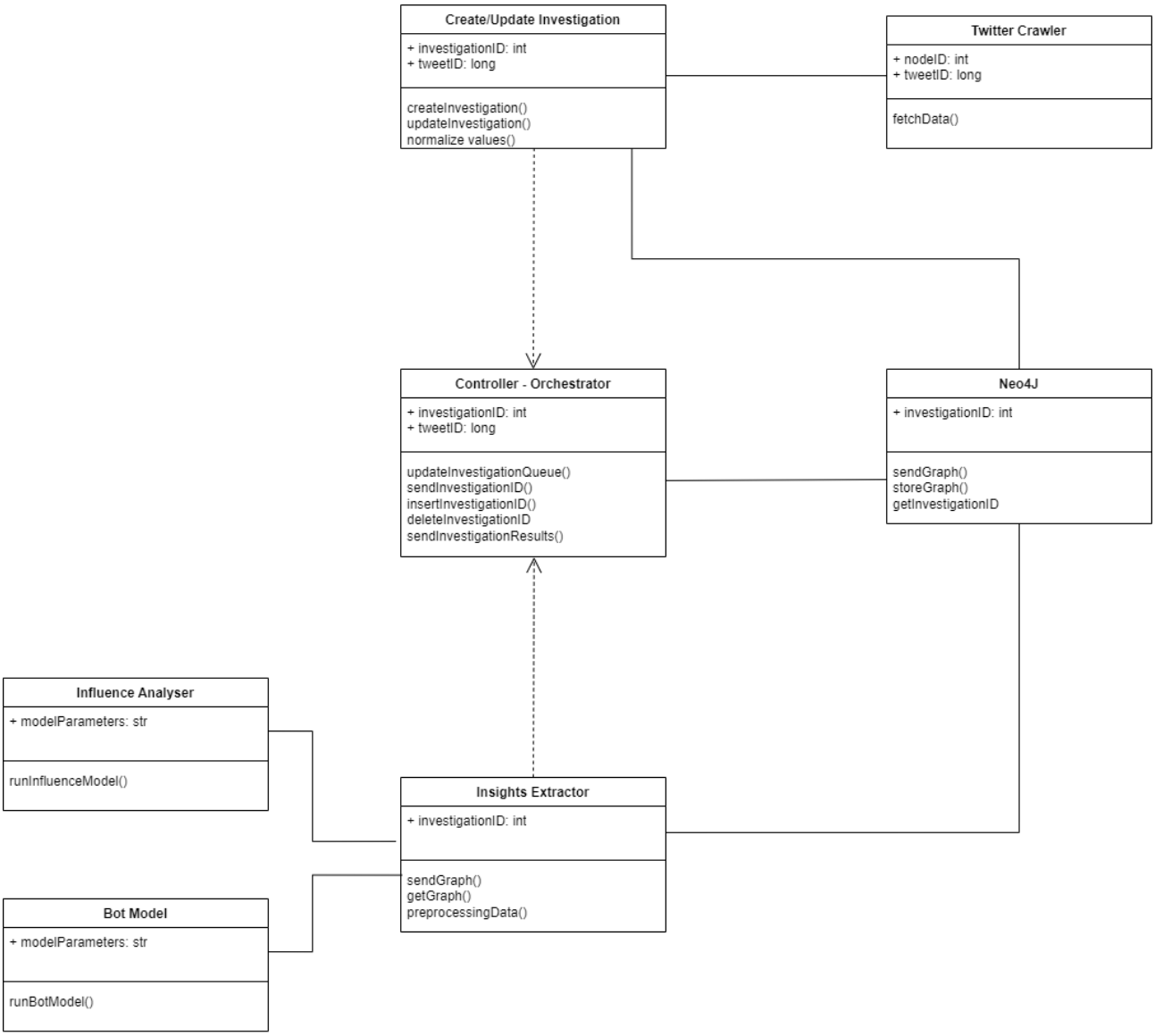


Figure 7 Class Diagram - Disinformation, Sources & Spread Analyser

The Disinformation, Sources and Spread Analyser class diagram consists of 7 main classes, all essential for the proper functioning of the component. The X/Twitter Crawler's[147] main functionality is to fetch data from X (post history, retweets, comments, mentions, e.g.) based on the input it receives. The Create/Update Investigation performs the functionality of normalising the gathered values and utilises the processed data to

[147] Albeit Twitter's name has officially been changed to X, the crawler is still being alluded to as Twitter crawler (as it was named by the time the project started) or X crawler.

the extent of tracing and mapping certain posts to the main creator. It also generates a graph that depicts the disinformation spread path, that serves as input for the Insights Extractor which is responsible for managing and fetching the required data for the Influence Analyser and Bot Model. The Influence Analyser identifies the most influential posts in the graph while the Bot Model classify the accounts as bots or humans. The Neo4J is responsible for temporarily managing and storing the graph data produced by the Create/Update Investigation during the investigation process. All the classes' communications are managed by the Controller-Orchestrator class which is also responsible for managing requests by other components.

- INOV T3.3 & T4.4 – Community Resilience Management Module

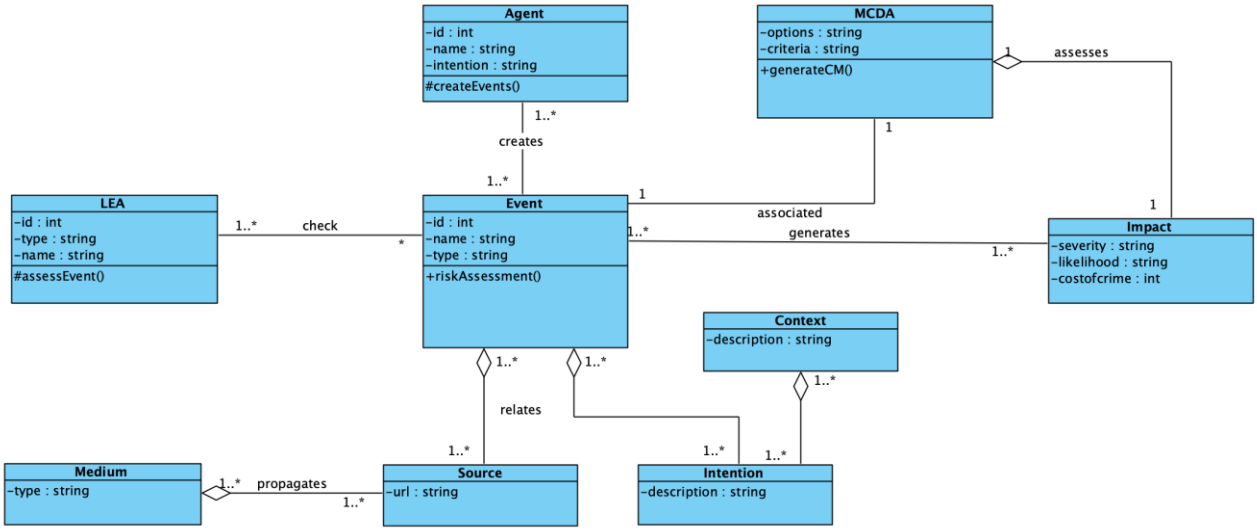


Figure 8 Class Diagram - Community Resilience Management Module

The Community Resilience Management Modeler and Disinformation watch joint module, which is represented in the above class diagram, has nine classes, four of each are considered essential for the model of resilience to work – LEA, Event, MCDA, and Impact. Without being too descriptive the LEA class is responsible for the assessment of a particular D&FN event represented within the diagram as a protected method +assessEvent(), the LEA is associated with different multiple events. The class Event has as its main function, the assessment of the risk of D&FN (represented as a public method +riskAssessment()). Each Event is responsible for the creation of at least one MCDA, and it generates at least one impact. It is important to understand that the MCDA working with a set of criterion and options will define the appropriate countermeasure(s) for each event. The method for MCDA is +generateCM(). Furthermore, for every given event and subsequent impact there is an associated control measure. The model will output a ranking of control measures related to each impact, and to each event.

Equally important to the definition of our model is the relationship of an agent of disinformation to the actual event of disinformation, each agent is responsible for the creation of one or more events. This is represented in the diagram through the association between the class Agent and Event.

Lastly, there is the aggregation relationship between the class Event and Intention. This denotes that for each event of D&FN there is a correlated source and intention between concepts and classes, though, without the existence of one class the other would still prevail. The same is true when we look at the correlation between Intention – Context and Source – Medium.

- ATOS T3.4 – Swarm Learning Module

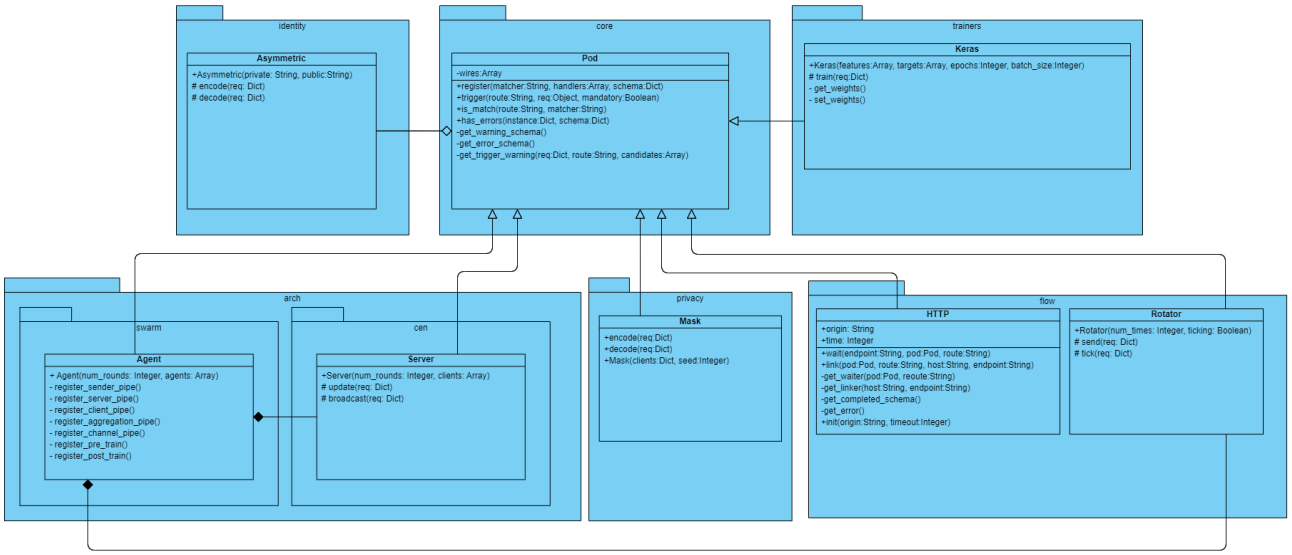


Figure 9 Class Diagram - Swarm Learning Module

This class diagram represents the structure of the Swarm Learning Module, including its classes, their attributes and the relationships between them. Its core class facilitates the Pod, the fundamental unit of the swarm that handles the operations of registering, triggering, matching and errors and warnings handling. The Agent class within swarm is the class that can act as client or server, depending on the role it takes in a specific round of the algorithm. The flow package has the HTTP class, that takes care of all the HTTP operations of the swarm, and the Rotator class that orchestrates which of the Agents has to become the Server in each round. The Keras class handles the Pods responsible for model training. In addition, the Asymmetric class introduces some security functionalities along with the Mask class, which encodes and decodes all the communications inside the swarm.

- BIGS T3.5 - Behaviour Profiler & Socioeconomic Analyser

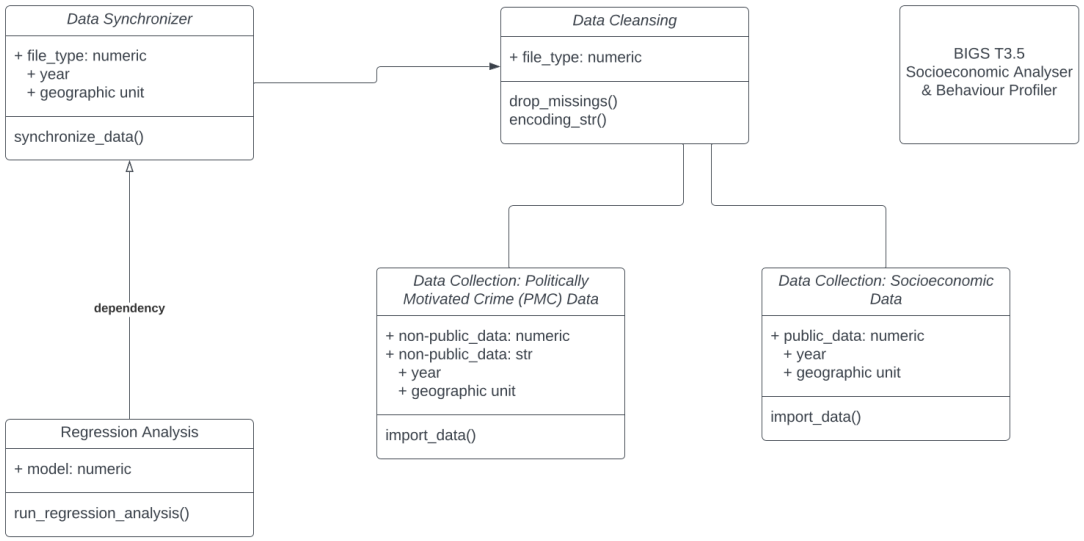


Figure 10 Class Diagram - Behaviour Profiler & Socioeconomic Analyser

The diagram above describes the Socioeconomic Analyser & Behaviour Profiler module class structure for the platform. In the data collection part of the Socioeconomic Analyser, data from publicly available databases on socioeconomic data (EUROSTAT and GENESIS - DESTATIS) is collected on the one hand. On the other hand, crime statistics delivered by the LEAs are collected. The input comes in numeric and string format. Both modules start out by importing the required data into Stata (statistic software tool). Data then is cleansed by dropping missings in the dataset and encoding string variables. Afterwards, years and geographical units are matched (Data Synchroniser). Lastly, a regression is run, for both modules respectively. For the Socioeconomic Analyser the input from the D&FN offline crime analysis is used.

Data cleansing and sorting is applied via Stata to eventually analyse the data in the regression analysis part. The socioeconomic analyser delivers an indicator measuring severity of crimes occurring in terms of economic costs. The Behavioural Profiler delivers an indicator measuring likelihood of crimes occurring. The results are in format of static datasets as well and will feed into the Community Resilience Management Module.

- ITML T3.6 – Sentiment Analysis Module

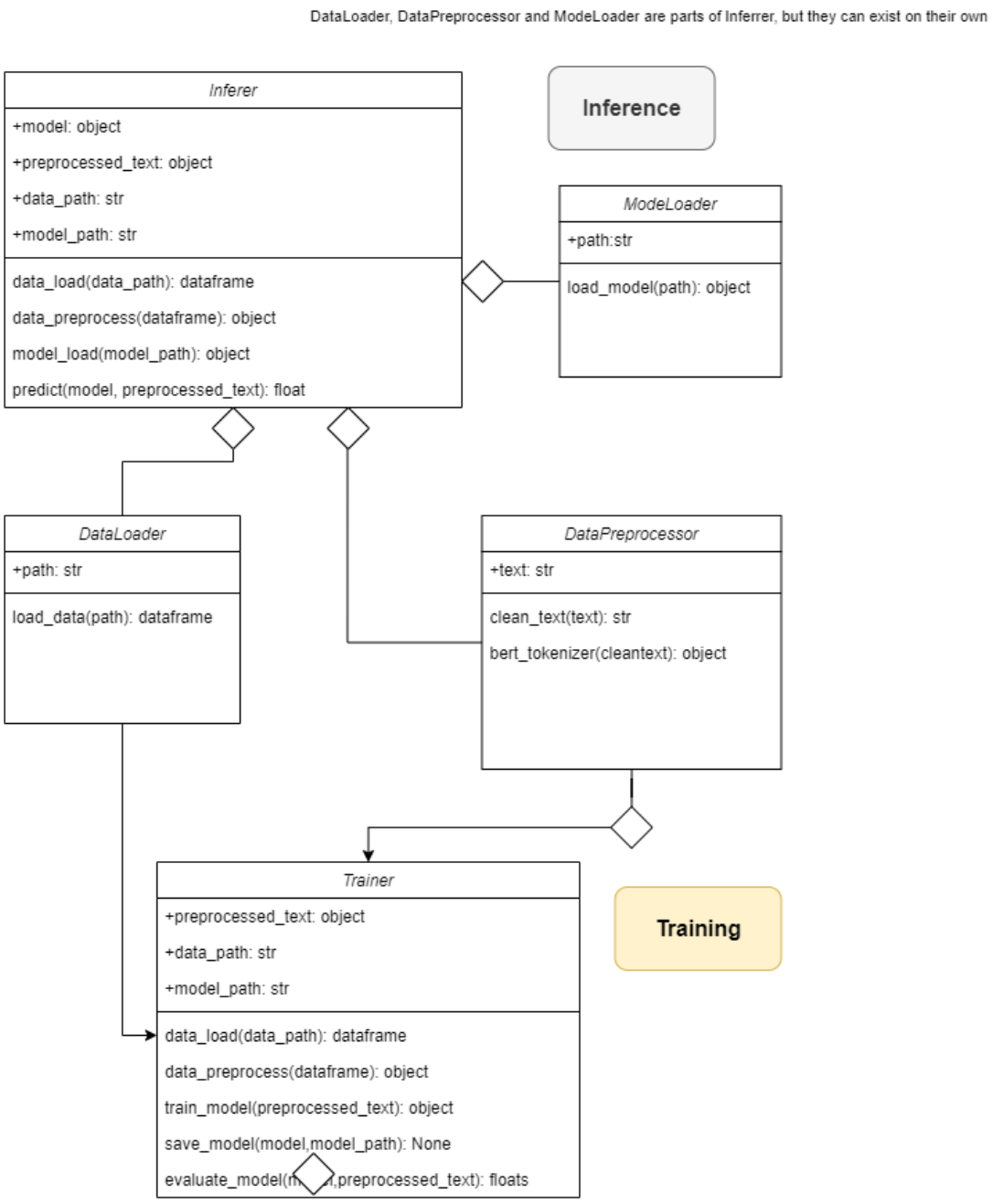


Figure 11 Class Diagram - Sentiment Analysis Module

ITML's diagram describes the Sentiment Analysis Module class structure for the platform. Two main classes implement the functionality in two distinctive roles: the Inferer(Inference) and the Trainer(Training). At the same time smaller classes are aiding the main ones: those are the DataLoader and the DataPreprocessor that have an aggregate relationship with the main classes and the ModelLoader that has an aggregate relationship

with the Inferer class. The Inferer will use data to perform the sentiment analysis based on the data produced and the model. The model used will be trained and evaluated on the available data in order to be utilised in the sentiment analysis. This process is facilitated by using smaller classes:

DataLoader: Gets the data from other components of the platform.

DataProcessor: Cleans the inputted data and tokenises the input to be used by the model for the analysis.

ModelLoader: Creates the model object to be utilised from the Inferer class for predictions.

- ITML T4.3 – Decision Support enabler

The following class diagram describes the class structure of the Decision Support enabler to be used for the initiation of an investigation.

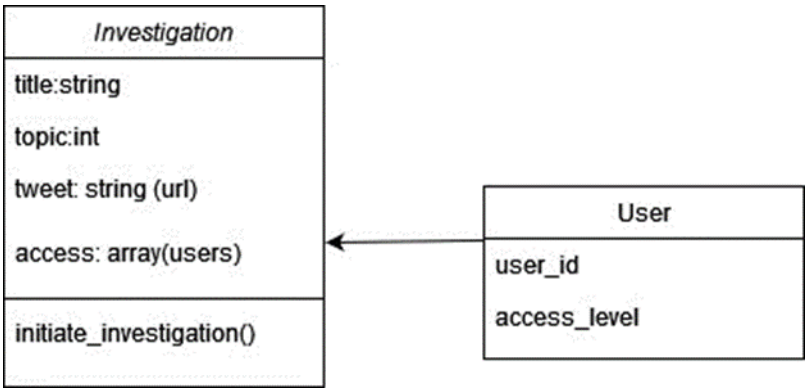


Figure 12 Class Diagram - Decision Support enabler

The investigation class describes the information provided by the user when initiating a new investigation through the FERMI UI, including information to be used by the user to identify their investigation, information required by the FERMI components and a list of users, which restricts access to this specific investigation.

Any other information which might be required as input form the user in future iterations, will also be added to the investigation class.

3.2.1.3 Sequence Diagrams

- UCSC T3.1 – D&FN Offline Crime Analysis

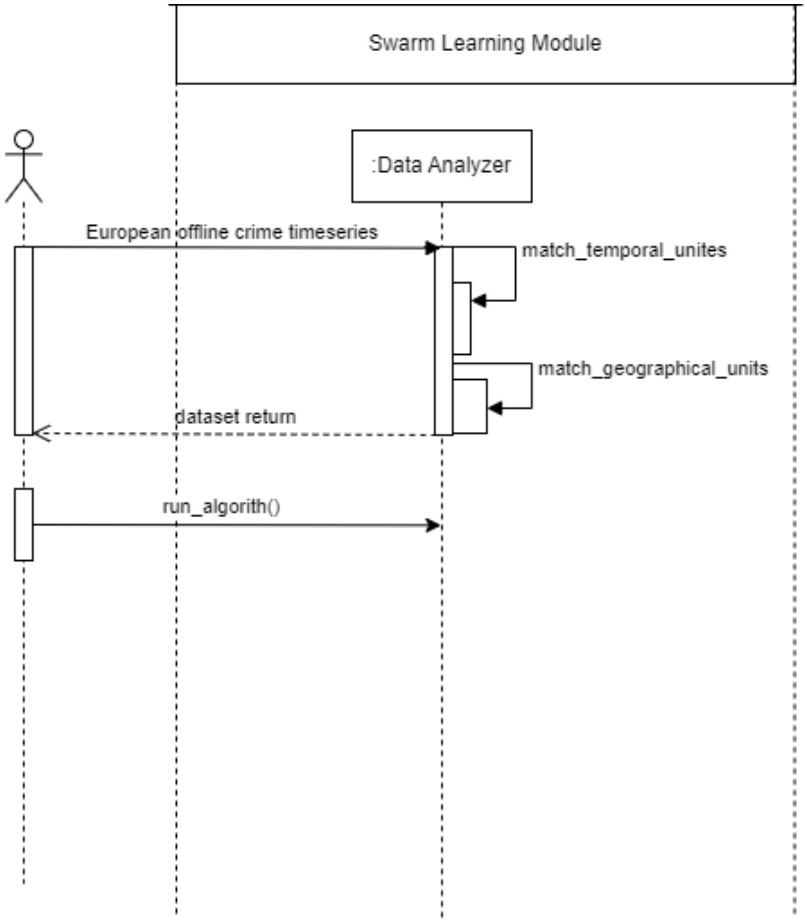


Figure 13 Sequence Diagram – D&FN Offline Crime Analysis

With the help of the sequence diagram of the D&FN Offline Crime Analysis component it is easier to identify the chronological order in which the component's internal functions take place. The diagram represents the two functionalities of the service. The sequence starts by providing an input to the Data Analyser (D&FN, criminal events, e.g.) that then creates a dataset by first matching the temporal units and then the geographical ones. Finally, the dataset is returned to the user. The system also gives the user the ability to manually start the execution of an algorithm.

Connection with other components:

- Depending on T3.4 Swarm Learning Module
- Depending on T3.2 Disinformation Sources & Spread Analysis and Impact Assessment
- Depending upon by T3.5 Socioeconomic Analyser & Behaviour Profiler

- INTRA T3.2 – Disinformation Sources Analyser

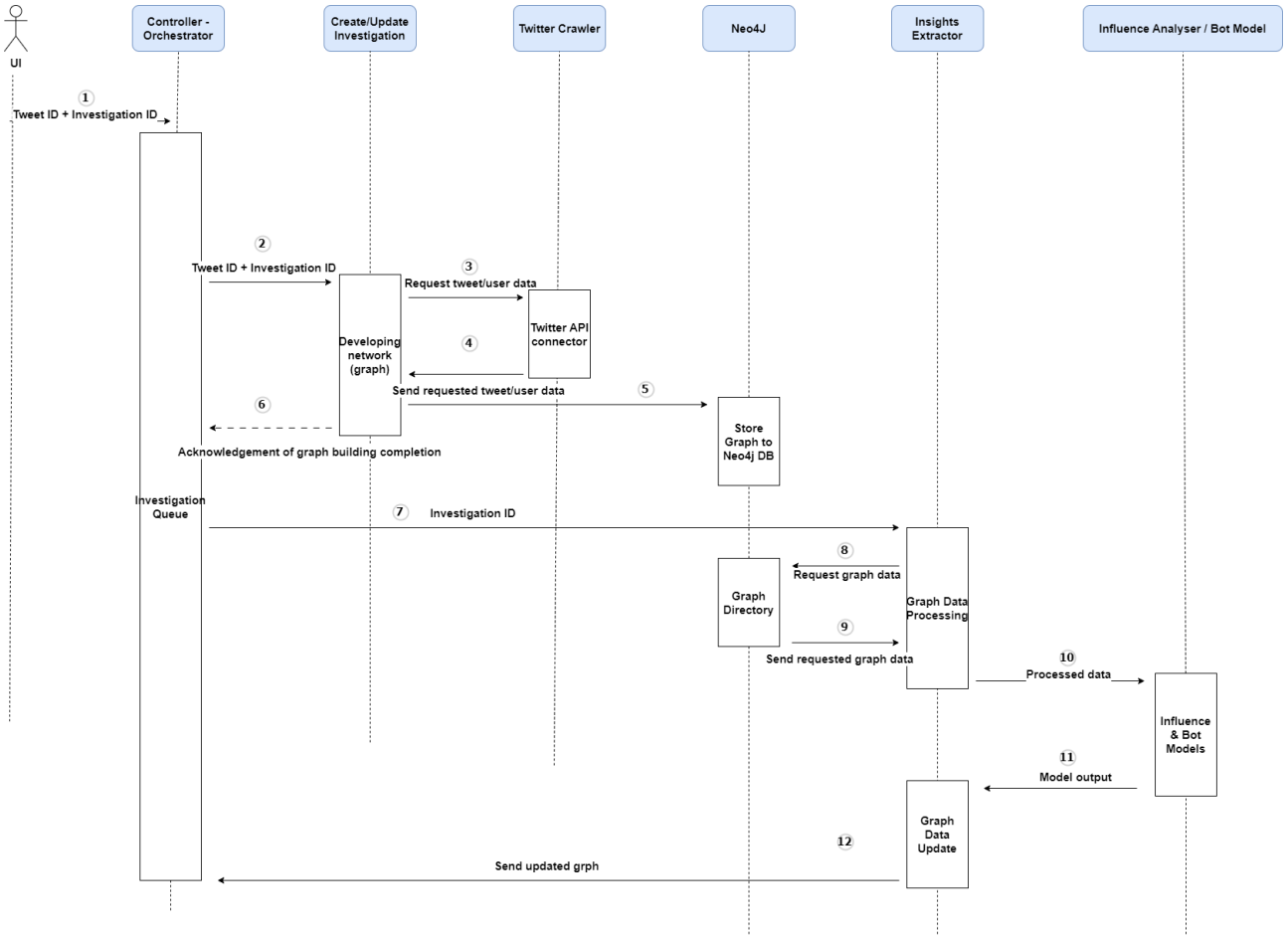


Figure 14 Sequence Diagram - Disinformation Sources Analyser

The sequence starts with the Controller-Orchestrator receiving a request from the user through the UI to start the investigation process. The request specifies a unique investigation id and a tweet ID that will be used as a starting point for fetching the corresponding data and creating a dataset using the X/Twitter Crawler. The Create/Update Investigation is processing the new datasets by removing unnecessary or unusable entries and normalising the existing values. In parallel it handles the transformation of the existing data to the end of creating a graph representing the network of tweets forming the disinformation spread cluster. This graph is stored temporarily in the Neo4J. Following this process, the Controller-Orchestrator sends the investigation id to the Insights Extractor to request the related data to the investigation graph data from the Neo4J. After receiving the data, it preprocesses them and sends them to the Influence Analyser (identify most influential posts) and Bot Model (classify accounts as bots or humans). The insights produced are returned to the Insights Extractor where the updated graph is sent back to the Controller-Orchestrator.

Connection with other components:

- Depending on T4.5 User Interfaces, Visualisation and Reporting Techniques.
- Depending upon by T3.1 D&FN-induced and D&FN-enabled offline crimes analysis and prediction.

- Depending upon by T3.6 The Sentiment Analysis module.

- INOV T3.3 & T4.4 – Community Resilience Management Module

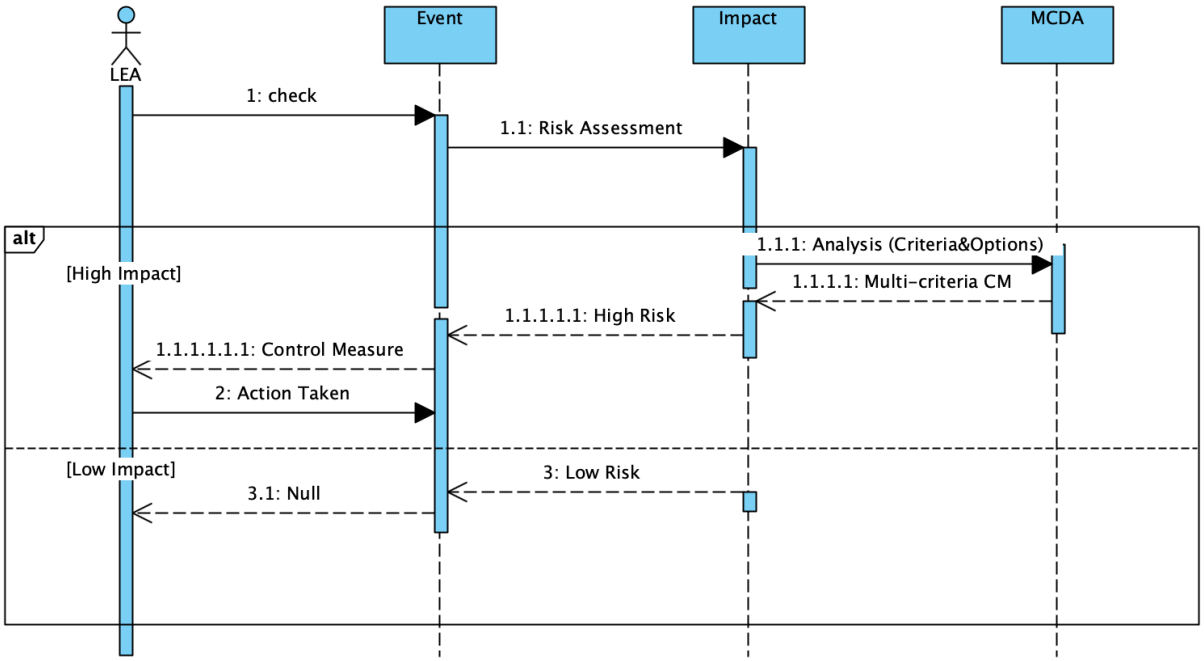


Figure 15 Sequence Diagram - Community Resilience Management Module

The sequence starts with the LEA actor that triggers the action by checking the event. The component then requires a risk assessment, essentially checking if the event is of high impact or low impact. Should an event be of high impact the system will require a multicriteria analysis. The MCDA combines a set of criteria with a set of options or alternatives to provide a ranking of the suggested countermeasures. The system will then return an indication of a high-risk impact event with a set of countermeasures that are tailored to the associated investigation. Once the return of countermeasures is returned to the LEA, action may be taken based on the set of countermeasures and the community decision-maker.

On the other hand, should the impact of an event be of low risk the system will not require a multicriteria analysis and therefore will return NULL to the LEA actor.

Connection with other components:

- Depending on T3.5 Socioeconomic Analyser & Behaviour Profiler.
- Depending upon by T4.5 User Interfaces, Visualisation and Reporting Techniques

- ATOS T3.4 – Swarm Learning Module

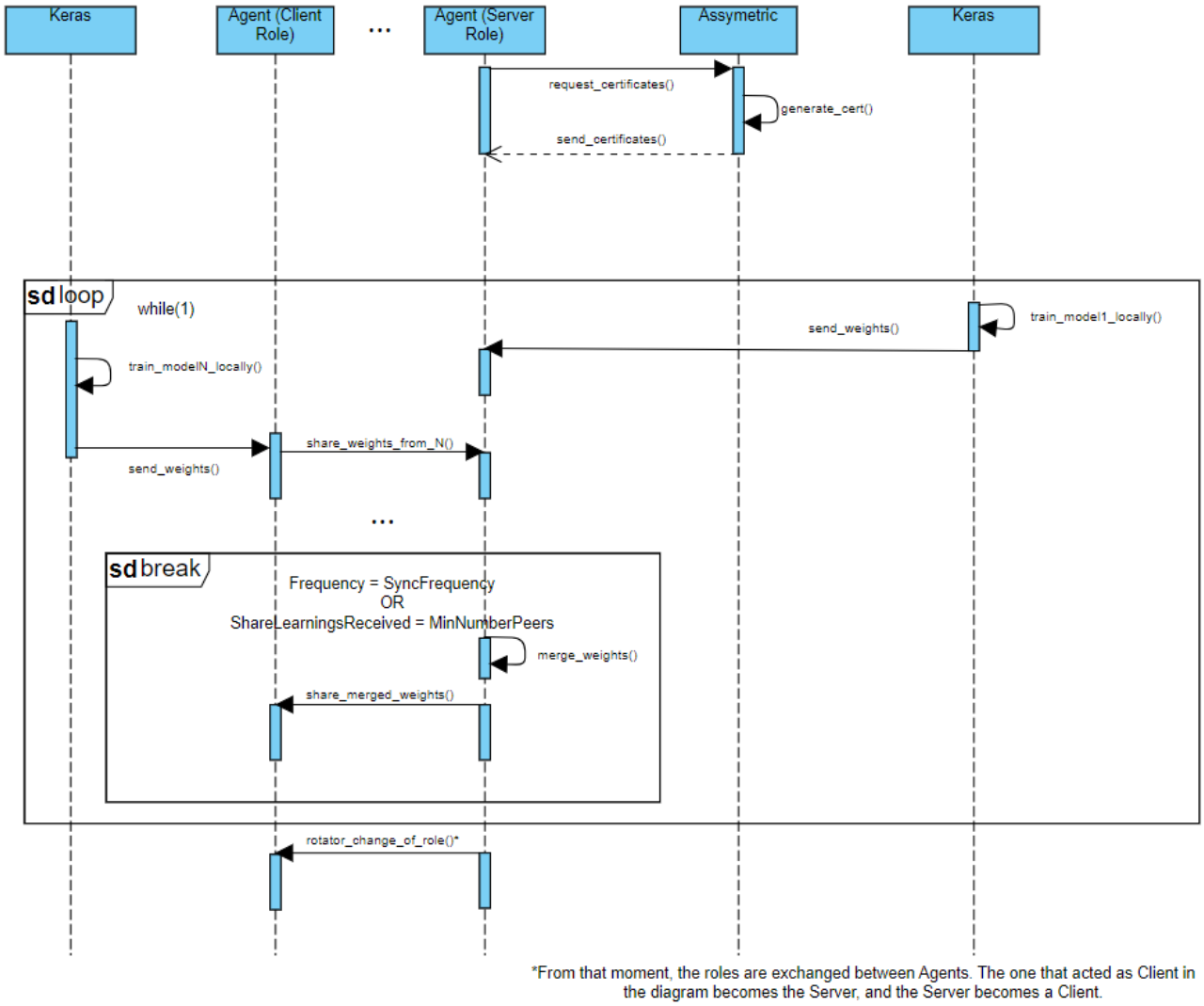


Figure 16 Sequence Diagram - Swarm Learning Module

The sequence diagram of the Swarm Learning Module indicates that the process begins with the Asymmetric, which is responsible for generating any requested certificates by the agents of the network.

One of the functionalities of the system is the training of the models. The main nodes (Agents) retrieve the weights from the Keras nodes, and when the training is over the independently calculated weights of each node are merged and shared. After each round, the Rotator (which is included inside the Agent pod) indicates the change of roles to the agents (in this methodology there is not a fixed Server node), in a way that one of the clients becomes the Server for the next round.

Connection with other components:

- Depending upon by T3.1 D&FN-induced and D&FN-enabled offline crimes analysis and prediction.

- BIGS T3.5 - Behaviour Profiler & Socioeconomic Analyser

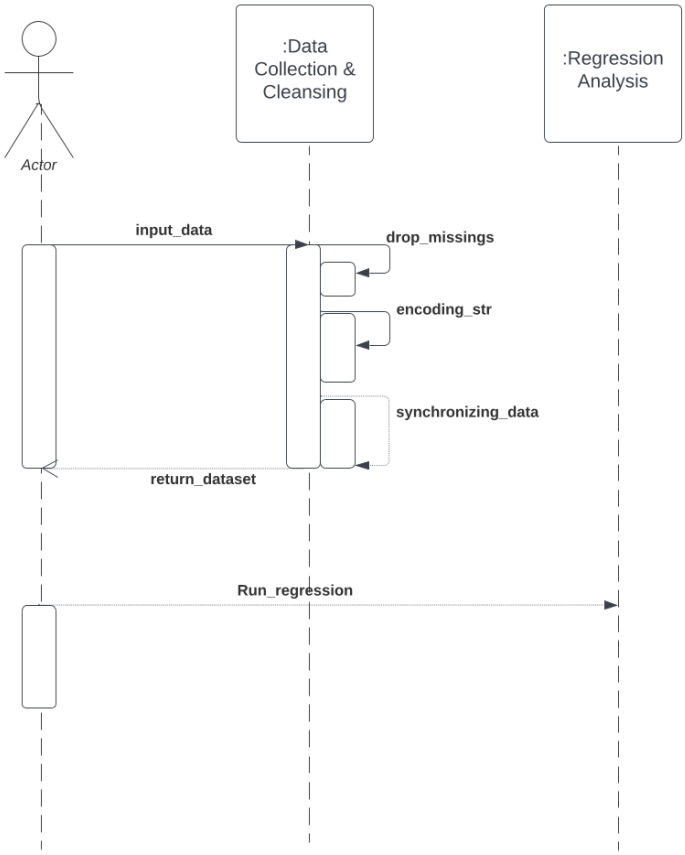


Figure 17 Sequence Diagram - Behaviour Profiler & Socioeconomic Analyser

The sequence diagram for the Socioeconomic Analyser & Behaviour Profiler shows that after inputting the collected static data missing are dropped, string variables are encoded and the data is synchronised according to year and geographical unit. The last part of the sequence diagram is the application of the regression algorithm to the dataset produced by the Data Analyser. Regarding this, the user is able to see a certain parameter produced by the regression analysis for the severity of a criminal event that has taken place.

Connection with other components:

- Depending on T3.1 D&FN-induced and D&FN-enabled offline crimes analysis and prediction.
- Depending upon by T3.3 – Community Resilience Management Module

- ITML T3.6 – Sentiment Analysis Module

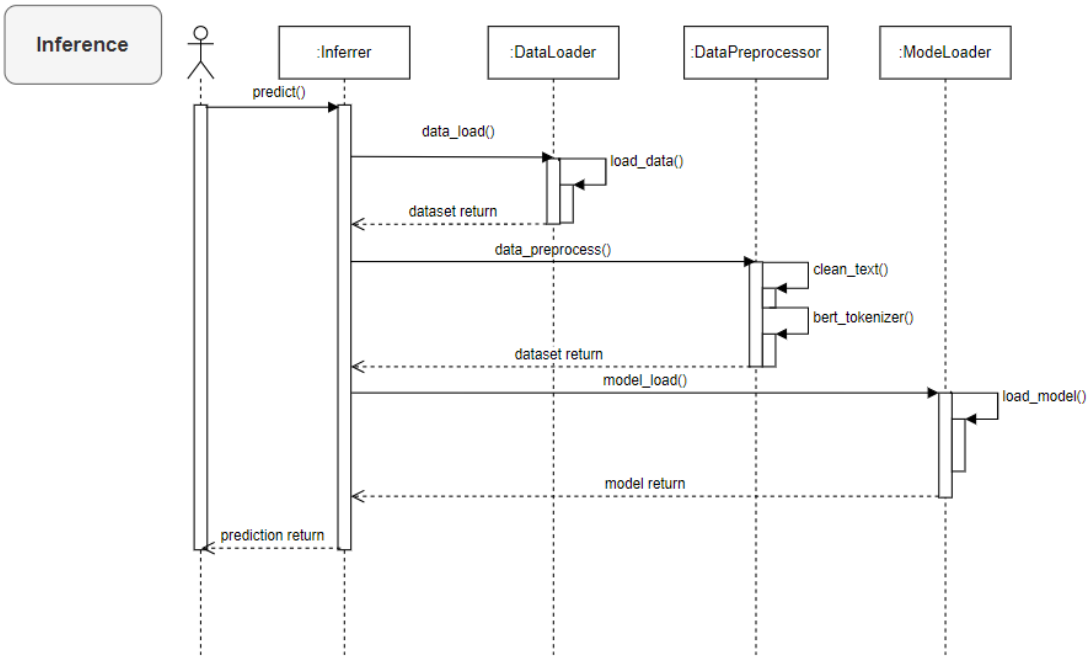


Figure 18 Sequence Diagram - Sentiment Analysis Module – Inference

The sequence of getting predictions from a trained model starts by requesting a prediction from the user. The Inferrer loads the requested data and sends them to the DataPreprocessor that is responsible for cleaning and tokenisation. The trained model is loaded from memory and makes the predictions on the cleaned data. Finally, the predictions are returned to the user.

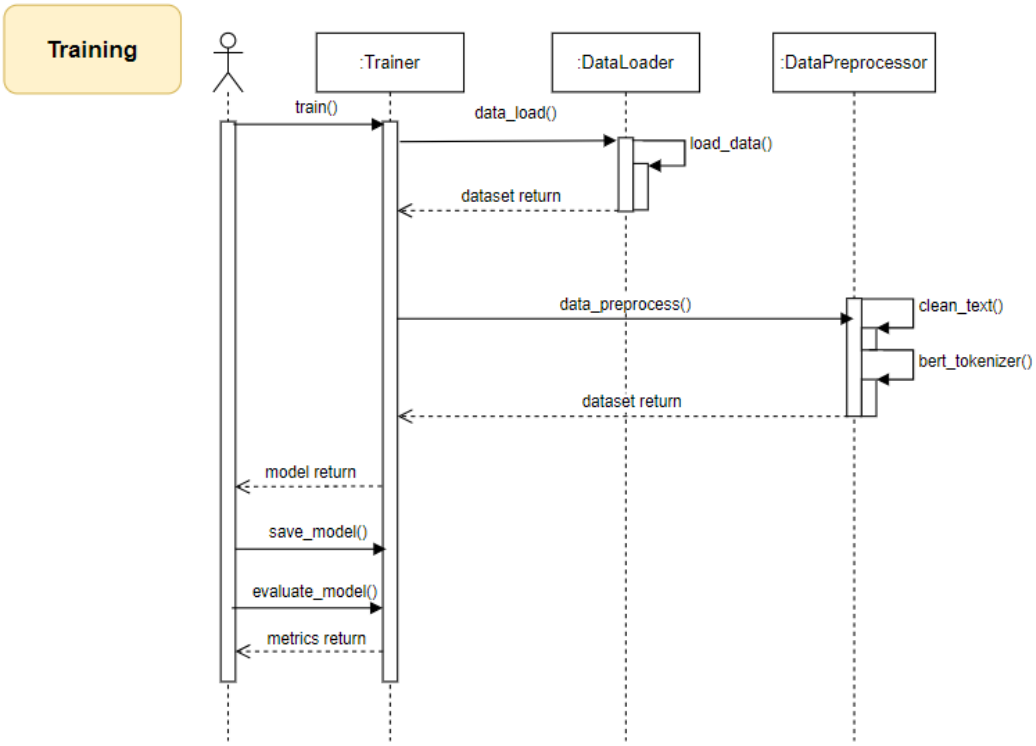


Figure 19 Sequence Diagram - Sentiment Analysis Module – Training

The sequence of training a model starts by a user train request. Data have to be retrieved from DataLoader and then cleaned and tokenised by the PreProcessor. The model then is trained, saved in memory and the metrics of its evaluations are available to the user.

Connection with other components:

- Depending on T3.2 Disinformation Sources & Spread Analysis and Impact Assessment
- Depending upon by T4.3 FERMI Decision Support enabler.
- Depending upon by T4.5 User Interfaces, Visualisation and Reporting Techniques

- ITML T4.3 Decision Support enabler

The sequence diagram of the Decision Support enabler describes the steps the user must follow to initiate an investigation.

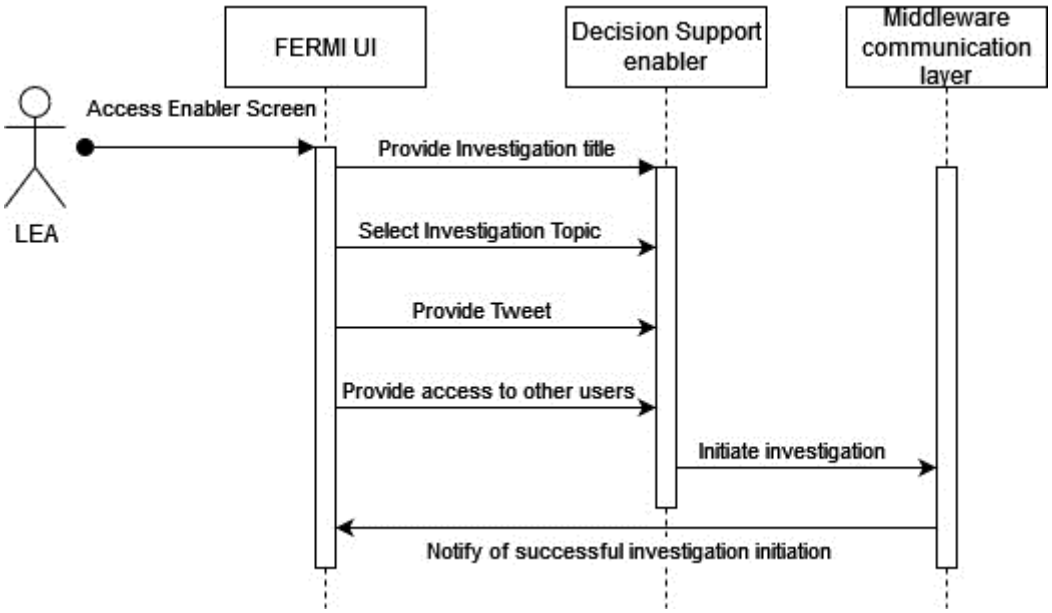


Figure 20 Sequence diagram - Decision Support enabler

The user has to select a title to identify their investigation, the topic among the use cases, provide the url of a specific tweet, and select which other users of the platform/community will have access to this investigation. Once the investigation is initiated through the Middleware communication layer, the rest of the components involved in the investigation are triggered to initiate their analysis.

Finally, it must be highlighted that the deployment view will act as the connection point between the architectural diagrams and the developed software modules. The deployment view will become the stepping stone to start developing the connectors between the actual parts of the system. The microservices described in the various diagrams will be containerised and broken down into smaller software units. Thus, the actual diagrams will act as input to T4.1 ("Continuous Integration and Delivery") of WP4 where the FERMI overall deployment view, as well as the pilot specific deployment views, are going to be developed and presented.

3.3 Potential Constrains

The FERMI project consists of several software components and tools with unique features and increased complexity. It is inevitable that potential constraints will arise from various factors such as design, hardware & OS, scheduling, documentation, and data sources. A simple design constraint may result from the complexity of integrating the various components to function cohesively, compromising the entire system functionality. Careful planning and execution are needed towards a secure architecture to ensure that the system is scalable, flexible, and meets all the users' needs.

Hardware and OS constraints can also be a major challenge. Disinformation analysis and AI-trained tools require high-performance processors, large memory capacities, storage systems, and complex data pipelines. These added limitations can significantly impact the performance of the system.

In addition to hardware and design constraints, schedule and documentation constraints can also impact the success of the project. All the components must be scheduled to be delivered on time and work together seamlessly. Detailed and comprehensive documentation should accompany every software component, including technical specifications and user manuals. This documentation must be accurate and up to date to ensure that stakeholders have the information they need to make informed decisions about the system.

Finally, data source constraints refer to the limited availability and quality of data sources. A disinformation analysis must be able to access diverse and reliable data sources to provide accurate and timely intelligence. Therefore, selecting both the data sources and the methods used to analyse them must be thoroughly planned.

Managing and overcoming all these potential failure points is of paramount importance to safeguard against a system that is difficult to use, slow and inefficient, prone to errors and not scalable. Inadequate data sources may lead to inaccurate and ineffective findings and render the system ineffective. It is critical to address all the referenced limitations to ensure that the system is delivered on time, meets functional and non-functional requirements, and provides accurate and reliable insights.

3.4 Potential Data Sources

The FERMI platform is consisted of multiple functional components with various data needs. There are several data source options to consider for the collection of data that can be used by the different components. One potential source where D&FN often spreads rapidly is X. Leveraging the X API to collect and analyse user-generated content is a possible data source worth considering. Other potential sources include government agencies, academic research institutions, and non-governmental organisations, which may provide access to reports, surveys, and datasets related to disinformation and fake news. Additionally, news outlets and media organisations could potentially provide valuable data, particularly through their archives and online databases. It's important to note that the FERMI platform benefits from data already identified as D&FN or data regarding the effects of D&FN in the community since the classification of data as a D&FN is out of the platform's scope.

4 Experimentation protocol: Use cases' and user scenarios' refinement and pathway towards FERMI validation

The UCs presented in this section are a derivative of extensive consultations and revisions of the produced outcomes – outcomes which were initially based on the Case Studies included in the GA of FERMI, which have then been specified to better match the current reality of violent extremism (a further set of revisions is due in the framework of the second experimentation protocol, which will reflect the pilot leaders' search for proper data to validate the platform. The result will be part of D5.1). In order to examine how the end-users would exploit the proposed technology and what objectives they would have (an essential component of UC definition), end-users could better envision using the FERMI tools and identifying potential objectives in a specific contextual instance (scenario). This is also fully in line with the GA. As clarified by the latter, T2.1 – amongst other things – is aimed at laying out "use-cases and scenarios, through which the results of the project will be demonstrated", whereas T2.4 stipulates that the "scenarios to be followed by each pilot will be defined, by scripting and matching all the steps to be taken in order to allow the pilots to validate the whole FERMI solution,"[148] which, besides the above-mentioned commitments includes drafting measurable KPIs so the technical output can be fully evaluated.

Each Use Case is accompanied by one scenario, thoroughly described in various fields, for better organisation of the information. The UC begins with a description of the context, the scenario, the needs and objectives of the end-users, as well as the operational steps the law enforcement officers would take. Please note that the operational steps are based on the participating LEAs' input and refer to procedures that are currently in place, with the available tools being subject to current legal provisions.

The template for the Use Cases is available in ANNEX D: Use Cases Template.

4.1 Use Cases Refinement

Accordingly, the UCs need to be defined in the first place before further specifics can be worked out. The use cases will be informed by actual events that have occurred in the host countries of the three pilots. This is an important step to ensure that the use cases and user scenarios are well-aligned with real-world events the platform will need to grasp, if and when it has been successfully completed and exploited.

With that said, the details that are provided as of now do not resemble actual proceedings, as the incorporation of such events depends on the availability of data, especially from social media providers that still remains to be obtained, depending on the data's lasting availability so it can be used in the pilots, which is a lot more predictable at a later stage, when the pilots are not that far off anymore (decreasing the odds that the rather

[148] Grant Agreement, PART B, p.35.

sensitive content of interest might be removed). Moreover, the WP2 schedule turned out to be too tight to acquire such data before the first drafting of the UCs anyway.[149] Against this backdrop, the specific cases are rather illustrative at this stage. Eventually, their implementation will be adjusted to the to-be-obtained data sets, especially from social media platforms such as X. This is totally reconcilable with the spirit of the GA, which requires the consortium to produce two experimentation reports that are meant to inform the pilot proceedings along the conceptual lines of the UCs and user scenarios, one within the framework of WP2 that lays the ground for the pilot testing and a final one within the framework of WP5, which is going to guide the exact implementation of the pilots.[150]

These constraints notwithstanding, the use cases and user scenarios are conceptualised in a way that is clearly geared towards measuring the platform's capability to meet end-user demands, as identified above with the help of a workshop and a survey. The resultant end-user requirements guide the finalisation of the user scenarios, so the latter can pave the ground for the measurement of the former. As clarified before, the FERMI platform is aimed at LEAs as end-users. Accordingly, one of the key prerequisites for adjusting the use cases and user scenarios to the measurement of end-user requirements is to ensure that those capture proceedings that are of special interest to LEAs. This implies that D&FN are to be examined in a context that requires the very active involvement of LEAs. Against this backdrop, numerous forms of false allegations are off the table. This does not only apply to D&FN that are seemingly detached from national/domestic security. Even far-reaching D&FN campaigns that are aimed at destabilising the target country such as those that have been conducted by Russia in recent years are beyond the scope of measures LEAs are responsible for.

Admittedly, Russian D&FN campaigns such as those aimed at weakening Hillary Clinton's candidacy in the US Presidential election in 2016 – amongst other things, by spreading false claims about the candidate and the election – and undermining the democratic process in Western countries are quite illustrative of the huge dangers that D&FN pose in this day and age.[151] However, as clarified in section 2 on the societal landscape, spreading false allegations is not a crime, at least not as such, unless certain boundaries are crossed and those

[149] The procurement of such data has been a little delayed by further ethics requirements that were included into the GA (in the form of WP7 and its deliverables) after the proposal's submission. In accordance with D7.2 full compliance with data protection norms and standards needed to be clarified by M3 before any personal data, including those of social media users, was allowed to be obtained (Grant Agreement, PART A, p.25). Against this backdrop, there was only a rather small window of opportunity to identify all data that might be legally and ethically procured, let alone to actually obtain such data, which – in the case of social media data – requires reaching out to social media providers and have them share the needed pieces of information. Moreover, the use cases can only be successfully concluded and reasonably guide the testing of the FERMI platform, if they cover the above-mentioned end-user requirements, which needed to be identified in the first place. However, the end-user requirements' elicitation was also delayed by the above-mentioned ethics requirements, as D7.1, which was also due by M3, obliged the consortium to document compliance with informed consent rules and norms before recruiting any research participants and collecting as well as processing any of their data (Grant Agreement, PART A, p.24). In other words, the aforementioned survey on end-user requirements, which was absolutely crucial for the elicitation thereof, needed to be postponed by a few months.

[150] Grant Agreement, PART A, p.36 and 40.

[151] Kling, Toepfl, Thurman and Fletcher, 'Mapping the website and mobile app audiences of Russia's foreign communication outlets, RT and Sputnik, across 21 countries,' *Harvard Kennedy School Misinformation Review* (2022). Available at: doi: 10.37016/mr-2020-110.

that disseminate such claims engage in illegal activities, which has also been emphasised by the LEA partners in the consortium time and again.

The same holds true for radical mind-sets, as emphasised by the GA as the UC's focal points.[152] Obviously, extremist belief systems are ideologies that can make individuals and groups particularly susceptible to buying into the logic of false allegations, if those corroborate their long-standing beliefs, which has also come up in discussions with LEA consortium partners and is further corroborated by the available literature. More specifically, numerous extremist beliefs are rooted in founding myths such as anti-Semitic resentments advocating conspiracy theories of Jewish world domination, often-times even citing fake documents such as the infamous Protocols of the Elders of Zion.[153]

Research on the subject has revealed that political extremists are particularly prone to "the belief that out-groups are engaged in secret actions to control in-group outcomes."[154] It has even been argued that violent extremists' "greatest tool is the mis-, dis-, and mal-information [...] that feeds extremist movements and ideologies."[155]

Against this backdrop, it seems highly reasonable to select D&FN use cases where the developments at stake are informed by political extremism. But again, extremist belief systems as such – just like spreading false allegations – are not subject to criminal investigations or other forms of LEA activities. LEAs are not in charge of monitoring, let alone investigating individuals or groups that have an extremist mind-set, unless there is a legal basis for surveillance in the event of a clear and present danger, which depends on the legal landscape in the given country. Whilst such beliefs appear to increase the odds that illegal activities are being committed, considering that extremist ideologies are grounded in the opposition to the EU's democratic proceedings and the rules and laws aimed at upholding those.[156]

From the standpoint of the EU's legal and policy approach, however, it is of vital importance that this threshold is actually crossed. As explained above in the section on the societal landscape, the EU's understanding of disinformation is clearly linked to efforts to undermine democracy. However, the EU has also been careful to distinguish between the illegal spread of content on the one hand and disinformation on the other, which makes it extremely difficult if not impossible to draft LEA-tailored UCs (possibly with the exception of certain

[152] Grant Agreement, PART A, p.13-15.
[153] Farinelli, 'Conspiracy theories and right-wing extremism – Insights and recommendations for P/CVE,' *Radicalisation Awareness Network (RAN)* (2021).
[154] Berger, *Extremism* (Cambridge, MA: The MIT Press Essential Knowledge series, 2018), p. 66.
[155] Mines and Hughes, 'The Fractured Threat Landscape,' *Police Chief Magazine* (2022), 36-41, p.36. Available at: https://www.policechiefmagazine.org/fractured-threat-landscape/.
[156] Amongst other things, false allegations may be used to appeal to a target audience's anti-democratic sentiments (as explained elsewhere in this deliverable, those are subject to an in-depth analysis with the help of one of the FERMI platform's tools, the Sentiment Analysis Module). Albeit a sentimental personal state is by no means a sufficient condition for the commission of crimes, highly sentimental target audiences that feel very strongly about a certain extremist cause are likely to be unusually susceptible to engage in violent and other illicit activities aimed at undermining democratic norms and rules that run counter to their extremist mind-set.

Member States, where clearly delineated forms of disinformation are illegal, but the ambition of this experimentation protocol is to present UCs that can be validated by all end-user partners, not just one or two of them). Anyways, the boundaries between extremist propaganda that does not constitute an offence and illegal activities is easily crossed if such worldviews are being aggressively promoted by agitators in a way that is apparently aimed at inciting violence.[157]

As further clarified in the societal landscape analysis above, limitations on freedom of expression are possible when they are prescribed by law and necessary and proportionate in a democratic society to pursue collectively relevant interests, such as national security, public safety and the prevention of public disorder or crimes. D&FN that start with efforts to undermine democratic norms and eventually transform into the incitement of violence are a perfect example of false allegations that are of huge interest to LEAs. The same applies to other crimes that are acknowledged EU-wide into which anti-democracy and extremist-oriented D&FN campaigns may transform such as terrorist recruitment, terrorist financing, providing and receiving training for terrorism and travelling as well as organising and otherwise facilitating travelling for the purpose of terrorism. These are all criminal offences that should be met with investigations on the part of LEAs.[158]

Accordingly, to test the FERMI platform in a way that is tailored to the needs of the envisaged end-users the use cases and user scenarios need to bridge the gap between addressing anti-democracy agitations as emphasised by the EU's definition of disinformation and criminal proceedings resulting from such activities like those alluded to above, so the LEA partners can evaluate the platform's key tools and functions in accordance with their job requirements. This is also nicely in line with the above-mentioned GA's requirement to facilitate the LEAs' fight against crime and terrorism.

While the above-mentioned set of crimes surely is not exhaustive, it gives a decent overview of criminal offences that might be committed partly because of D&FN stirring up anti-democratic sentiments if the latter are successfully shared amongst those that are susceptible to false allegations. [159]

These insights pave the road towards drafting UCs that all address a specific topic and demonstrate how FERMI will provide added value to combating the illegal spread of disinformation and fake news. The three main Use Cases defined, based on the GA, are:

[157] The EU has clearly stipulated in its Terrorism Directive "that the distribution, or otherwise making available by any means, whether online or offline, of a message to the public, with the intent to incite the commission of […] [terrorist offences], where such conduct, directly or indirectly, such as by the glorification of terrorist acts, advocates the commission of terrorist offences, thereby causing a danger that one or more such offences may be committed, is punishable as a criminal offence when committed intentionally." This article of the Terrorism Directive is called "Public provocation to commit a terrorist offence." (European Union, *Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism and replacing Council Framework Decision 2002/475/JHA and amending Council Decision 2005/671/JHA*. Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj. – Referred to as Directive on Combatting Terrorism as follows).

[158] European Union, Directive on Combatting Terrorism.

[159] Rice, 'Emotions and terrorism research: A case for a social-psychological agenda,' *Journal of Criminal Justice*, 37 (2009), 248-255.

UC1: Disinformation and Fake News related to political interference from violent extremists on the far-right
UC2: Health Crisis, riots and other forms of violence
UC3: Disinformation and Fake news leading to violence from the far-left [160]

Admittedly, these forms of violent extremism do not cover all ideological root causes of radical beliefs but considering that the FERMI project attempts to fill a void by examining the spread and ramifications of D&FN, it seems reasonable to select areas of violent extremism that are clearly linked to sharing false allegations.

Again, the choice made by the GA and the FERMI consortium is fully corroborated by informal discussions with LEA end-users within the consortium and the available literature. A timely assessment of predispositions to D&FN has concluded that such an inclination "can be observed on the left and on the right" alike.[161] A recent study of the Radicalisation Awareness Network (RAN) that was authored by Francesco Farinelli on behalf of the European Commission identifies anti-immigrant conspiracy theories, anti-Semitic conspiracy theories, anti-establishment and anti-elite conspiracy theories and conspiracy theories in the COVID-19 context as the most profound popular beliefs that are promoted by political extremists relying on false allegations.[162]

Admittedly, the report's immediate focus is on right-wing extremism, which might have informed this selection. However, further research has found out that certain right-wing "[p]arties such as the National Democratic Party of Germany[163] and The Third Way have been involved in organizing protest groups online (typically via Facebook) and stirring up anti-refugee sentiments with falsified statistics of immigrants' crimes or claims of specific events witnessed by friends and colleagues, such as incidents of rape or child abduction by refugees."[164]

Moreover, such sentiments might easily translate into organised violence, as groups "like The Third Way have also published guidebooks on how to organize large-scale protests, and have officially registered demonstrations that, in the majority of cases, devolved into violent action or took place shortly before arson attacks."[165]

[160] Grant Agreement, PART B, p.13-15.
[161] Albeit, "the right-wing political identity (or conservatism) seems to trigger motivated political reasoning more strongly and more frequently." See Baptista and Gradim, 'Who Believes in Fake News? Identification of Political (A)Symmetries,' *Social Sciences*, 11 (2022). Available at: https://doi.org/10.3390/socsci11100460.
[162] Farinelli, 'Conspiracy theories and right-wing extremism – Insights and recommendations for P/CVE,' *Radicalisation Awareness Network (RAN)* (2021).
[163] The aforementioned developments in Germany will also be taken into consideration in the sense that BPA, albeit not an LEA partner in a narrow sense, will provide a further data set on Bavaria, Germany to ensure that such proceedings can be examined in-depth.
[164] Koehler, 'Right-Wing Extremism and Terrorism in Europe. Current Developments and Issues for the Future,' *Prism: The Journal of Complex Operations*, 6 (2016). Available at: https://cco.ndu.edu/PRISM/PRISM-Volume-6-no-2/Article/839011/right-wing-extremism-and-terrorism-in-europe-current-developments-and-issues-fo/.
[165] Koehler, 'Right-Wing Extremism and Terrorism in Europe. Current Developments and Issues for the Future,' *Prism: The Journal of Complex Operations*, 6 (2016). Available at: https://cco.ndu.edu/PRISM/PRISM-Volume-6-no-2/Article/839011/right-wing-extremism-and-terrorism-in-europe-current-developments-and-issues-fo/.

The importance of violent right-wing extremism notwithstanding, anti-establishment propaganda is also at the heart of violent left-wing extremism. This is corroborated by further in-depth research that has identified "a link between [left- and right-wing] political extremism and a general susceptibility to conspiracy beliefs. Although the extreme left may sometimes endorse different conspiracy theories (e.g. about capitalism) than the extreme right (e.g. about science or immigration), both extremes share a conspiratorial mindset, as reflected in a deep-rooted distrust of societal leaders, institutions, and other groups, allied with a corresponding tendency to explain unexpected, important events through conspiracy theories."[166]

And violent COVID-related extremism might also be relatively independent without greatly overlapping with long-standing radical belief systems. The spread of COVID-related D&FN amongst violent extremists and the latter's inclination to engage in illegal activities is a more recent trend like COVID itself. Having said that, the evidence that there are widespread D&FN campaigns that all address COVID is clear and overwhelming.[167] So is the evidence of the nexus between numerous such D&FN campaigns and violence. Interestingly, "the first year of the COVID-19 pandemic did not dramatically alter the number of terrorist attacks around the world […] but individual conspiracy theory extremists were involved in an increasing number of incidents, particularly against telecom infrastructure."

More specifically, "the largest increase was in attacks committed by conspiracy theory extremists: six in 2019, versus at least 116 in 2020, in countries ranging from Australia and New Zealand to the United States, Canada, United Kingdom and Germany. Nearly all were non-lethal, and a surprising 96% were aimed at damaging telecom targets […] showing not only the influence of conspiracy theories concerning 5G and other wireless technologies – which range from causing cancer and killing animals and plants to causing the coronavirus outbreak – but also how perpetrators in the United States and western Europe are largely acting as part of loose ideological movements, not in concert with organizations."[168]

Accordingly, the FERMI project will carry out three pilots along the lines of the GA's focal points. UC1 is led by FMI,[169] UC2 is led by BFP and UC3 is led by BPA.[170]

The use case descriptions are informed by the GA's priorities but – again – remain to be adjusted to the available data sets once such data has been acquired in the context of the final experimentation protocol's drafting, as

[166] van Prooijen, 'Voters on the extreme left and right are far more likely to believe in conspiracy theories,' *EUROPP – European Politics and Policy at LSE blog*. Available at: http://bit.ly/1zS8hW3.
[167] Lynas, 'COVID: Top 10 current conspiracy theories.' *Alliance for Science*, 20 April 2020. Available at: https://allianceforscience.org/blog/2020/04/covid-top-10-current-conspiracy-theories/.
[168] Farrell, 'UMD Report: Conspiracy theories fueled more terror attacks in 2020,' *National Consortium for the Study of Terrorism and Responses to Terrorism* (START, 7 July, 2022). Available at: https://www.start.umd.edu/news/umd-report-conspiracy-theories-fueled-more-terror-attacks-2020.
[169] Like other European countries, Finland has witnessed arson attacks on refugee shelters (Koehler, 'Right-Wing Extremism and Terrorism in Europe. Current Developments and Issues for the Future,' *Prism: The Journal of Complex Operations*, 6 (2016). Available at: https://cco.ndu.edu/PRISM/PRISM-Volume-6-no-2/Article/839011/right-wing-extremism-and-terrorism-in-europe-current-developments-and-issues-fo/), which speaks for the selection of Finland as a country where the corresponding pilot will be organised (by FMI).
[170] Grant Agreement, PART A, p.13.

explained above. For the time being, the use cases remain largely illustrative. More definitive planning can be done with respect to the user scenarios that capture the different steps that are to be taken by the pilot participants, which are structured in a manner that resembles the above-mentioned need for LEA involvement starting with a violation of the law (in the context of anti-democratic violent extremism-induced D&FN) requiring an investigation.

Accordingly, all pilots include one user scenario, each scenario (investigation scenario, threat assessment scenario and community resilience scenario) addresses a different cornerstone of an LEA's reaction to illegal D&FN campaigns.

4.2 Preliminary Use Cases and User Scenarios

4.2.1 UC1: Disinformation and fake news related to political interference from violent extremists on the far-right

Partner	FMI
Use Case number	UC1 RIGHT- WING EXTREMISM (Investigation)
Use Case Description	In Europe, including Finland, the spread of D&FN on social media by violent right-wing extremist groups targeting migrants and refugees has been causing huge unease among law enforcement agencies. Some of those Tweets link migrants and refugees to crime using false statistics and inflammatory language suggesting that migrants and refugees pose a danger to public safety. The rhetoric may turn out to be so aggressive that the EU's definition of a "Public provocation to commit a terrorist offence" is met. The Tweets have been shared widely on social media, contributing to a wave of xenophobic comments and threats against migrants and refugees undermining public safety. Accordingly, an investigation into the Tweets' origin and the individuals or groups behind them is being launched. The investigation is a challenging task, as the Tweets have been disseminated so broadly. The investigation involves tracking down the original source of the Tweets, including the attempt to identify whether individuals or groups are behind them or just bots and collecting proper evidence.
Investigation Scenario	**Investigating the incident – a three-tier technical approach towards investigating the account and collecting (further) evidence** Step 1: Launching the investigation Amidst launching the investigation, the messages are pre-categorised in accordance with the kind of (violent) belief system they support, in this case: right-wing extremism. Step 2: Investigating the account The investigators need to uncover whether the accounts that are used to spread such illegal D&FN are operated by human beings or happen to be just bots. If the former is correct, an investigation can be launched against the alleged perpetrator(s) behind the account. Step 3: Collecting further evidence – investigate the spread of criminal D&FN

	In the event the account has been verified to be run by an actual person or group of persons, the analysis has proceeded quite significantly in the sense that a formal investigation against the alleged perpetrator(s) behind the account can be launched. That being said, further evidence can be collected. More specifically, the relevant message's popularity might be examined. Considering that the illegal activity at stake is at least indirectly aimed at compelling others to engage in illegal activities, namely terrorist proceedings, the scope of dissemination might further inform the to-be-made case against the alleged perpetrator(s). The more popular illegal messages are, the more likely it is that such messages compel others to heed the alleged perpetrator(s) call to action/follow in the alleged perpetrator(s) footsteps.
Usability Evaluation	**The Usability Evaluation** LEAs should not only evaluate the effectiveness of their actions but also assess whether any necessary adjustments or improvements to the platform's essential tools are necessary to ensure their usability in an investigation context.
Factors	
Actors	Not immediately involved in the pilot but of huge importance for its success • Perpetrator(s) of the illegal D&FN campaign - the source of the D&FN message and the (main) actor(s) responsible for spreading D&FN (data to be acquired) • X - the platform where the D&FN is shared and spread, where law enforcement agencies look for further information. Actual pilot participants/supporters • LEAs - the main actors responsible for investigating the D&FN Tweet. • FERMI technical partners – provide advice and help wherever necessary to facilitate the testing of the FERMI platform's relevant tool (see below).
Technologies currently available (Users)	The technologies currently available to users in this use case include social media platforms, digital monitoring and analysis tools, and

	traditional law enforcement technologies such as surveillance cameras and crowd control equipment.
Technologies desired by the PROJECT platform	The officers will utilise the disinformation sources, spread and impact analyser to detect bots and gather evidence of the spread of disinformation (including the influence of accounts).
To-be-examined End-User Requirements	**Investigating the incident – a three-tier technical approach towards investigating the account and collecting (further) evidence** The pre-investigative step of categorising the message at stake along the lines of its ideological roots is embedded in UR010 (The user through the platform is able to classify disinformation posts by category (e.g., political, health-related)). Investigating the specific account captures UR001 (The user is able to identify whether the X account spreading fake news online is a physical actor or a bot), UR002 (The user is able to assess the origin of the disinformation with accuracy more than 80%) and UR020 (The user is able to track down the origin and distribution of disinformation campaigns related to violent extremism (right-wing extremism, left-wing extremism, health-related extremism)). Collecting evidence on the D&FN spread captures UR003 (The user is able to identify key actors involved in spreading illegal disinformation campaigns), UR005 (The user is able to grasp the social media interactions of those who are actively promoting D&FN), UR007 (The user is able to use graph data for analysis, based on fetching and transformation of all the responses, likes, and retweets of a disinformation post), UR033 (The user is able to measure the reach and impact of illegal disinformation campaigns on social media) and UR008 (The user is able to estimate the most influential actor in the graph (social media account post) spreading D&FN). **The Usability Evaluation** The platform's evaluation captures the user-oriented dimension of UR012 (The reports should be customisable based on the user's needs and should be easy to understand and interpret), UR013 (The user is able to have access to interactive visualisations and dashboards generated by the platform to help law enforcement officers understand complex data

	patterns and trends), UR035 (The user is able to use the platform in a user-friendly way), UR036 (The user complies with relevant data protection and privacy regulations while using the platform) and UR037 (The user is able to process and analyse large volumes of data from various sources, including social media platforms through the utilisation of the FERMI platform).

Goals and Objectives	The pilot-specific goals and objectives are to enable LEAs to investigate right-wing terrorists that use D&FN to incite violence, to grasp the spread of such D&FN on social media, especially the scope of right-wing terrorist incitement. More specifically, the pilot aspires to
	• Enable LEA officers to use the FERMI platform, incl. the capability to
	  o distinguish between human beings and bots as account operators,
	  o capture the spread of to-be-investigated Tweets, and
	  o grasp the influence thereof.

4.2.2 UC2: Health Crisis, riots and forms of violence

Partner	BFP
Use Case number	UC2 COVID-RELATED EXTREMISM (Threat Assessment)
Use Case Description	A police officer in the Belgian Federal Police's open-source intelligence (OSINT) unit notices that numerous X accounts spread D&FN related to the COVID-19 pandemic. These claims have been shared with a large following and have already garnered numerous likes and shares. They are supplemented by calling on civilians to exercise violent protests and attacks on vaccination centres. Considering that the Tweets include a clear violation of the law, namely a "Public provocation to commit a terrorist offence," an investigation into the subject matter has been launched already. However, the OSINT unit is more concerned about the large following of the relevant X account, which implies that violent activities might be conducted, possibly up to a point where public safety might be undermined. Accordingly, the OSINT unit needs to do a threat assessment so they can get a better idea of the scope of further crimes that might still unfold. In this regard, law enforcement agencies should communicate with other relevant agencies to gather additional information or support. The interactions with fellow LEAs can be hugely facilitated by using advanced data analysis tools and techniques such as machine learning algorithms and network analysis. To obtain an artificial intelligence model as powerful as possible, collaboration between different LEAs is required enabling all players to use sensitive data from different law enforcement agencies across Europe, whilst ensuring privacy and confidentiality.
Threat Assessment Scenario	**Assessing the incident's immediate ramifications – grasping the danger of escalations and predicting what crimes are likely to be committed** Step 1: Grasping the sentiment of the social media messages at stake A basic threat assessment will be carried out by grasping the sentiments that drive the relevant social media messages. Those are a first indication that can cast light on the likelihood of tensions and escalations. Overly negative messaging would surely be a growing cause of concern, whereas more nuanced messaging would imply a lesser scope of threat. Step 2: Crime prediction

	Making an estimate of the types, times and areas of crimes that are likely to be influenced by the illegal D&FN campaign. In this manner, the FERMI's platform supports the allocation of police resources. Throughout this process data is being exchanged with fellow LEAs.
Usability Evaluation	**Usability Evaluation** LEAs should not only evaluate the effectiveness of their actions but also assess whether any necessary adjustments or improvements to the platform's essential tools are necessary to ensure their usability is guaranteed to properly carry out threat assessments.
Factors	
Actors	Not immediately involved in the pilot but of huge importance for its success • Perpetrator(s) of the illegal D&FN campaign - the source of the D&FN message and the main actor(s) responsible for spreading D&FN (data to be acquired) • X - the platform where the D&FN is shared and spread, where law enforcement agencies look for further information. • Social media and Internet users - the general public who are exposed to the D&FN campaign and whose opinions and sentiments may be influenced by it (data to be acquired) Actual pilot participants/supporters • LEAs - responsible for doing the threat assessment. • FERMI's technical partners – provide advice and help wherever necessary to facilitate the testing of the FERMI platform and its tools.
Technologies currently available (Users)	The technologies currently available to users in this use case include social media platforms, digital monitoring and analysis tools, and traditional law enforcement technologies such as surveillance cameras and crowd control equipment.
Technologies desired by the PROJECT platform	The sentiment analysis module is to be used to grasp the sentimental state of the relevant players that engage in spreading the D&FN messages with the possibly destabilising impact. The D&FN Offline Crime Analysis will be used to predict potential future criminal events related to the spread of the D&FN proceedings at stake.

	The Swarm Learning module will provide an estimation of the number of crimes in different areas and periods of time, easing the collaboration between LEAs and ensuring the privacy and confidentiality of the data. Moreover, the tool will facilitate data analysis by different LEAs. More specifically, it allows a federated training of machine learning and deep learning models by using data from different LEAs in a private and confidential manner.
To-be-examined End-User Requirements	**Assessing the incident's immediate ramifications – grasping the danger of escalations and predicting what crimes are likely to be committed** Conducting a general threat analysis and assessment captures UR021 (The user will be able to identify potential threats to public safety) and UR029 (The user should be able to evaluate the impact of illegal disinformation campaigns on public opinion). A particular sub-element of this rather general threat assessment includes grasping the emotional state of the relevant players, which captures UR011 (The user is able to analyse the emotional polarity of social media posts related to disinformation). Predicting potential future criminal events captures UR027 (The user should be able to predict which kind of crimes the D&FN will eventually lead to), UR014 (The user is able to predict who are the potential victims and targets of crimes related to D&FN) and UR017A (The user can identify the environment and context in which the criminal event may occur due to the D&FN). The ensuing capability to alert LEAs to areas of risk captures UR004 (The user is able to contribute to the better allocation of law enforcement resources to prevent and respond to disinformation-induced crimes) and UR038 (The user is able to provide near real-time alerts and notifications to law enforcement officers when new threats are detected. The alerts should be customised based on the user's preferences and job responsibilities). Coordinating with fellow LEAs amidst the effort to predict criminal events captures UR019 (The user could be able to collaborate with other law enforcement agencies to combat the illegal ramifications of disinformation campaigns without the need of sharing the data outside of its facilities) and UR031 (The user should be able to access accurate

	information regarding offline crimes stemming from D&FN campaigns, improved through incoming data collected from different LEAs/sources). **The Usability Evaluation** The evaluation of the platform's usability in a threat assessment context captures UR026 (The user should be able to easily handle an AI-based tool to reliably predict the scope of disinformation-induced crimes) and the user-oriented part of UR038 (near real-time alerts […] should be customised based on the user's preferences and job responsibilities.). More general usability notions include the user-oriented dimension of UR012 (The reports should be customisable based on the user's needs and should be easy to understand and interpret), UR013 (The user is able to have access to interactive visualisations and dashboards generated by the platform to help law enforcement officers understand complex data patterns and trends), UR035 (The user is able to use the platform in a user-friendly way), UR036 (The user complies with relevant data protection and privacy regulations while using the platform) and UR037 (The user is able to process and analyse large volumes of data from various sources, including social media platforms through the utilisation of the FERMI platform).
Goals and Objectives	The pilot-specific goals and objectives are to empower LEAs to understand the scope of the newfound threat posed by D&FN concerning health-related issues and to draw proper conclusions in terms of operational reactions. More specifically, this involves the use of the FERMI platform to analyse the sentiments of the relevant Tweet landscape and to predict the types, times and areas of crimes.

4.2.3 UC3: Disinformation and Fake news leading to violence from the far-left

Partner	BPA
Use Case number	UC3 LEFT-WING EXTREMISM (Community Resilience)
Use Case Description	Violent left-wing extremist groups have spread D&FN about alleged government repression online and offline causing concern among law enforcement agencies. Often-times such Tweets include false information and use sensational language to incite fear and anger among the public and portray the country's current form of democracy as kleptocracy benefitting the wealthy and the powerful whilst oppressing the vulnerable. Some Tweets call on the public to join unauthorised demonstrations and to use violence against security forces. After having launched an investigation into the matter and doing a threat assessment the police – in coordination with the Ministry of the Interior – decides to analyse the likelihood and severity of potential further crimes, so an overall cost estimate can be made. If required by the D&FN campaign's scope, possible counter-measures are to be weighed to strengthen community resilience. Moreover, an evaluation of the overall response to the threat is conducted in the sense of analysing the synchronisation and reliability of the platform and its specific tools aimed at facilitating the above-mentioned crucial steps including investigation, threat assessment and community resilience.
Community Resilience Scenario	**Assessing the incident's broader ramifications – coming to grips with the cost of D&FN induced extremism and proper counter-measures** Step 1: Assess the likelihood of successful pushback – measuring community resilience The officers will assess community resilience and perform risk management. More specifically, they will measure community resilience trough the detailed analysis of the likelihood that further crimes take place and the ensuing socioeconomic cost thereof. Step 2: Assess the situation from an LEA standpoint – identify the decisions that need to be made The FERMI platform can assess the risk posed by D&FN campaigns and help LEA decision-making, if necessary. More specifically, the platform can be consulted to identify the best course of action. Admittedly, this does not relief LEA officers of the burden of making their own decisions

	but it can cast light on the available options and facilitate such decision-making by making proper suggestions that are informed by a wealth of evidence.
Usability Evaluation	**Usability Evaluation** LEAs should not only evaluate the effectiveness of their actions but also assess whether any necessary adjustments or improvements to the platform's essential tools are necessary to ensure their usability is guaranteed and community resilience can be properly strengthened.
Factors	
Actors	Not immediately involved in the pilot but of huge importance for its success • Perpetrator(s) of the illegal D&FN campaign - the source of the D&FN message and the main actor(s) responsible for spreading D&FN (data to be acquired) • X - the platform where the D&FN is shared and spread, where law enforcement agencies look for further information. • Social media and Internet users - the general public who are exposed to the D&FN campaign and whose actions may be influenced by it (data to be acquired) Actual pilot participants/supporters • LEAs - the main actors responsible for doing the community resilience analysis and responding to the D&FN campaign, if necessary. • FERMI technical partners – provide advice and help wherever necessary to facilitate the testing of the FERMI platform and its tools.
Technologies currently available (Users)	The technologies currently available to users in this use case include social media platforms, digital monitoring and analysis tools, and traditional law enforcement technologies such as surveillance cameras and crowd control equipment.
Technologies desired by the PROJECT platform	The officers will use the Community resilience management modeler to assess community resilience and manage risk based on behavioural profiles and socioeconomic analysis, which includes the Behaviour profiler and the Socioeconomic analyser. The tool will assist LEA with

	ranked countermeasures provided through the employment of a MCDM analysis. Further tools used for validating the platform's synchronisation are listed as follows: FERMI platform tools to be used amidst the investigation - The officers will utilise the disinformation sources, spread and impact analyser to check if the sources of disinformation are physical actors or a bots. - The disinformation sources, spread and impact analyser will also be used to analyse the spread of the relevant pieces of illegal D&FN in the sense of quantifying the level of D&FN spread and the influence thereof. - The sentiment analysis module is to be used to grasp the sentimental state of the relevant players that engage in spreading the D&FN messages with a possibly destabilising impact. - AI-based predictive tools such as the D&FN Offline Crime Analysis, along with federated learning tools, such as the Swarm Learning module, will be used to identify and prevent potential violent incidents. - The Swarm Learning module will facilitate data analysis by different LEAs. More specifically, it allows a federated training of machine learning and deep learning models by using data from different LEAs in a private and confidential manner.
To-be-examined End-User Requirements	**Assessing the incident's broader ramifications – coming to grips with the cost of D&FN induced extremism and proper counter-measures** The attempt to assess community resilience and manage risk captures UR028 (The user is able to assess community resilience based on community behavioural profiles and socioeconomic analysis) and UR017B (The user is able to manage risk based on community behavioural profiles and socioeconomic analysis). Risk management amidst estimating the likely costs captures UR016 (The user is able to quantify the economic impact by making an approximation on the costs of violent extremism caused by disinformation and fake news), UR018 (The user is able to determine the economic factors that

	play into the ramifications of disinformation) and UR015 (The user is able to increase his/her knowledge about the socioeconomic and cultural aspects and the perception of illegal disinformation among citizens). Identifying possible courses of actions captures UR012 (The user is able to have access to detailed reports, generated based on the data analysed. The reports should be customisable based on the user's needs and should be easy to understand and interpret). **The Usability Evaluation** The platform's usability captures the user-oriented dimension of UR012 (The reports should be customisable based on the user's needs and should be easy to understand and interpret), UR013 (The user is able to have access to interactive visualisations and dashboards generated by the platform to help law enforcement officers understand complex data patterns and trends), UR035 (The user is able to use the platform in a user-friendly way), UR036 (The user complies with relevant data protection and privacy regulations while using the platform) and UR037 (The user is able to process and analyse large volumes of data from various sources, including social media platforms through the utilisation of the FERMI platform).
Goals and Objectives	The pilot-specific goals and objectives are to successfully estimate the economic cost of violent left-wing extremism, including welfare losses, opportunity cost, and externalities. Besides that, the risk posed by such D&FN campaigns is to be assessed and proper counter-measures are to be produced. Moreover, an in-depth testing and evaluation of the entire FERMI platform's synchronisation and the tools' interaction should demonstrate the advanced stage and reliability thereof.

4.3 Evaluation Strategy

After having elicited the essential requirements as reported by end-users and having defined use cases to trace the fulfilment of requirements and KPIs, a brief overview of the outstanding steps is given here.

As explained above, this includes

- KPI definition – This step aims at identifying key performance indicators that can be used to further measure the use cases and user scenarios' successful implementation in the sense of grasping whether the key expectations of end-users as summarised in the end-user requirements have been met.

- Fine-tuning the use cases – in view of data availability by LEA stakeholders, especially use case leaders in D5.1, when the likelihood of social media content being removed is a little smaller because the pilot dates are closer and very recent LEA needs can be taken into consideration. Some use cases will also be further defined with the contribution of technical partners, especially the ones to evaluate the requirements and KPIs that are more technical.
- Traceability matrix definition – This task is about making a traceability matrix for WP5 to prove a holistic view that all requirements have been properly addressed by the use cases and user scenarios. This matrix will be done by the task 5.6 ("Pilot Evaluation & Assessment") leader.
- Use cases and user scenarios execution – In this step, all user scenarios will be conducted throughout the use case pilots that will be implemented within WP5. Each use case leader will be in charge of coordinating the execution of their use cases and user scenarios. Feedback will be collected and evaluated by IANUS in accordance with T5.6.
- Use cases and user scenarios results evaluation and analysis – Involves the evaluation of results from the use cases and scenarios and other collected data in the process of their execution in the framework of WP5. To do this, the Likert scale[171] will be used.

Figure 21 Validation system

Accordingly, the one outstanding task that still remains to be fully addressed in the framework of WP2 is the specification of KPIs. As explained above, in FERMI, the measurement of the performance and the progress for the proposed system, in terms of technical components, as well as a whole is done via the usage of KPIs to capture the fulfilment of requirements defined in WP2. Again, both the KPIs and requirements will be checked with the help of use cases and scenarios such as the above-mentioned ones. Test results will be benchmarked then against a defined set of KPIs, validating them accordingly.

Questionnaires will serve as an essential component as part of a comprehensive evaluation strategy, designed to assess the effectiveness and impact of utilising FERMI's tools. Grounded on the foundational insights of D2.1, this ensures that all relevant information regarding each pilot scenario is methodically presented. Administered to external evaluators (LEAs), the questionnaires guarantee an impartial assessment, vital for garnering objective feedback. Moreover, they help link the evaluation of activities with the project's exploitation and potential commercialisation endeavours of WP6.

[171] Joshi, Kale, Chandel and Pal, 'Likert Scale: Explored and Explained,' *British Journal of Applied Science & Technology*, 7 (2015). Available at: doi: 10.9734/bjast/2015/14975.

More specifically, the questionnaires will enable the FERMI consortium to measure to-be-defined KPIs in the field of end-user satisfaction. Technically, a KPI is a type of performance measurement, which is done against a predefined set of values, called indicators. According to Parmenter "Key performance indicators […] represent a set of measures that focus on the aspects of organizational performance that are the most critical for the current and future *success of the organization. The identification of KPIs is crucial as it provides* a way to quantify the outcomes of a demo and assess the performance of the demonstrated solutions."[172]

Such KPIs can be derived from the end-user requirements. Considering that the Won't-have end-user requirements are irrelevant, they can be ignored. Obviously, the same applies to those that are beyond the scope of the project as clarified in section 3. The achievement of the other requirements can be measured along the lines of end-user satisfaction based on the level of approval they receive (in this regard, a typical 5-point Likert scale comes in handy, as it allows for the easy distinction between approval (in the form of the "strongly approve/agree" and "approve/agree" categories) and further feedback expressing non-approval ("strongly disapprove/disagree" and "disapprove/disagree") or neutrality ("neither approve/agree nor disapprove/disagree").

Accordingly, there are three remaining categories of relevance, the Could-haves, Should-haves and Must-haves. If one assumes that even the Could-haves should pass a litmus test of garnering more than 50% end-user satisfaction, which would validate their technical development and indicate that they can be exploited in good conscience, the only outstanding distinction to be made would be the one between Should have and Must-have end-user requirements. In both cases, end-user satisfaction should clearly be above the 50% threshold. In the latter case, it should be significantly higher. Whilst all such distinctions are somewhat arbitrary, the highest threshold assigned to an end-user requirement in the survey is 80%.[173] This is a highly ambitious KPI for the Must-haves, which the FERMI consortium embarks on anyway. As far as the Should-haves are concerned, a line can be drawn in the middle of both ends, which would be the 65% mark.

To sum up, all MoSCoW user-requirements are to be transformed into KPIs as follows:

- Won't-haves: irrelevant
- Could-haves: at least 50% end-user satisfaction
- Should haves: at least 65% end-user satisfaction
- Must-haves: at least 80% end-user satisfaction

[172] Parmenter, 'Background to the Winning KPI Methodology,' *Key Performance Indicators* (2019). Available at: doi: 10.1002/9781119620785.ch3.
[173] UR002 ("The user is able to assess the origin of the disinformation with accuracy more than 80%", see below).

Table 5 FERMI User Requirements and KPIs

FERMI Requirements List UR001-UR038				
UR ID		**Title**	**Priority**	**KPI**
UR001		The user is able to identify whether the X account spreading fake news online is a physical actor or a bot.	Must	>80% end-user satisfaction
UR002		The user is able to assess the origin of the disinformation with accuracy more than 80%.	Should	>65% end-user satisfaction
UR003		The user is able to identify key actors involved in spreading disinformation campaigns.	Should	>65% end-user satisfaction
UR004		The user is able to contribute to the better allocation of law enforcement resources to prevent and respond to disinformation-induced crimes.	Should	>65% end-user satisfaction
UR005		The user is able to grasp the social media interactions of those who are actively promoting D&FN.	Should	>65% end-user satisfaction
UR007		The user is able to use graph data for analysis, based on fetching and transformation of all the responses, likes, and retweets of a disinformation post.	Must	>80% end-user satisfaction
UR008		The user is able to estimate the most influential actor in the graph (social media account post) spreading D&FN.	Should	>65% end-user satisfaction
UR010		The user through the platform is able to classify disinformation posts by category (e.g., political, health-related).	Could	>50% end-user satisfaction
UR011		The user is able to analyse the sentiment polarity of social media posts related to disinformation.	Must	>80% end-user satisfaction
UR012		The user is able to have access to detailed reports, generated based on the data analysed. The reports should be customisable based on	Must	>80% end-user satisfaction

		the user's needs and should be easy to understand and interpret.		
UR013		The user is able to have access to interactive visualisations and dashboards generated by the platform to help law enforcement officers understand complex data patterns and trends.	Should	>65% end-user satisfaction
UR014		The user is able to predict who are the potential victims of crimes related to D&FN.	Must	>80% end-user satisfaction
UR015		The citizen is able to increase his/her knowledge about the socioeconomic and cultural aspects and the perception of disinformation among citizens.	Should	>65% end-user satisfaction
UR016		The user is able to quantify the economic impact by making an approximation on the costs of violent extremism caused by disinformation and fake news.	Could	>50% end-user satisfaction
UR017	A	The user can identify the geographical unit in which the criminal event may more likely occur due to the D&FN	Should	>65% end-user satisfaction
	B	The user is able to manage risk based on community behavioural profiles and socioeconomic analysis.	Should	>65% end-user satisfaction
UR018		The user is able to determine the economic factors that play a role in the ramifications of disinformation.	Should	>65% end-user satisfaction
UR019		The user is able to collaborate with other law enforcement agencies to combat the illegal ramifications of disinformation campaigns without the need of sharing the data outside of its facilities.	Should	>65% end-user satisfaction
UR020		The user is able to track down the origin and distribution of disinformation campaigns related to violent right-wing extremism.	Must	>80% end-user satisfaction
UR021		The user is able to identify potential threats to public safety.	Should	>65% end-user satisfaction

UR026		The user is able to easily handle an AI-based tool to reliably predict the scope of disinformation-induced crimes.	Should	>65% end-user satisfaction
UR027		The user is able to predict which kind of crimes the D&FN will eventually lead to.	Should	>65% end-user satisfaction
UR028		The user is able to assess community resilience based on community behavioural profiles and socioeconomic analysis.	Should	>65% end-user satisfaction
UR029		The user is able to evaluate the impact of disinformation campaigns on public opinion.	Should	>65% end-user satisfaction
UR031		The user should be able to access accurate information regarding offline crimes stemming from D&FN campaigns, improved through incoming data collected from different LEAs/sources.	Should	>65% end-user satisfaction
UR033		The user is able to measure the reach and impact of disinformation campaigns on social media (i.e., X).	Must	>80% end-user satisfaction
UR035		The user is able to use the platform in a user-friendly way.	Must	>80% end-user satisfaction
UR036		The user complies with relevant data protection and privacy regulations while using the platform.	Must	>80% end-user satisfaction
UR037		The user is able to process and analyse large volumes of data from various sources, including social media platforms through the utilisation of the FERMI platform	Must	>80% end-user satisfaction
UR038		The user is able to provide near real-time alerts and notifications to law enforcement officers when new threats are detected. The alerts should be customised based on the user's preferences and job responsibilities.	Should	>65% end-user satisfaction

5 Conclusion

This deliverable summarises the starting point of the FERMI project. It started with the description of the extracted user needs and their requirements in order to combat the ramifications of D&FN. The methodology followed along with the analysis of the user requirements is also described in Section 1. Moreover, the societal landscape is described in the sense of finding a fair balance between law enforcement objectives and the protection of fundamental rights and democratic values. In this context, three common elements that will guide FERMI's understanding of disinformation have been identified: 1) factual or misleading nature of the information; 2) intention of the actors to spread such information they know to be false to obtain economic gain or deceive the public; 3) public harm.

In Section 3, an extensive technical analysis is presented driving the architectural design of the project based on the derived functional requirements. Lastly, the use cases and scenarios are presented in Section 4, accompanied by a concrete description of the KPIs. The purpose of this deliverable is also to serve as the guide for the development activities in WP3, the platform integration process in WP4 and the pilot validation campaigns in WP5. Any updates on the topics of this deliverable, will be documented in the deliverables of the respective WPs, especially in D5.1, which will include an extended and revised experimentation protocol.

References

8 T-262/15, Kiselev v. Council [2017] ECLI:EU:T:2017:392.

Baldassi and Others v. France App no. 15271/16, 15280/16, 15282/16 et al., (ECtHR 11 June 2020).

Baptista and Gradim, 'Who Believes in Fake News? Identification of Political (A)Symmetries,' *Social Sciences*, 11 (2022). Available at: https://doi.org/10.3390/socsci11100460.

Bayer et al., *The fight against disinformation and the right to freedom of expression* (Policy Department for Citizens' Rights and Constitutional Affairs Directorate-General for Internal Policies, European Parliament, 2021).

Benedik v. Slovenia, App no. 62357/14 (ECtHR 14 July 2018).

Berger, *Extremism* (The MIT Press Essential Knowledge series, 2018).

Bloch-Wehba, 'Content moderation as surveillance,' *Berkeley Technology Law Journal*, 36 (2021).

Bradshaw, Howard, *The global disinformation order: 2019 global inventory of organised social media manipulation* (n.2 Working Paper 2019: Project on Computational Propaganda, 2019).

Brzeziński v. Poland App no. 47542/07 (ECtHR 25 July 2019).

Buckingham, 'Teaching media in a 'post-truth' age: fake news, media bias and the challenge for media/digital literacy education*,' Culture and Education*, 31 (2019), 213-231.

C-140/20, G.D. v. The Commissioner of the An Garda Síochána and others [2022], ECLI:EU:C:2022:258.

C-622/17, Baltic Media Alliance v. Lietuvos radijo [2019] ECLI:EU:C:2019:566.

Castets-Renard, 'Algorithmic Content Moderation on Social Media in EU Law: Illusion of Perfect Enforcement,' *University of Illinois Journal of Law, Technology & Policy*, 283 (2020).

Catt v. United Kingdom App no. 43514/15 (ECtHR 14 January 2019).

De Streel et al., *Online Platforms' Moderation of Illegal Content Online: Law, Practices and Options for Reform* (Policy Department for Economic, Scientific and Quality of Life Policies Directorate-General for Internal Policies, European Parliament, 2020).

European Commission, *Action Plan against Disinformation* (Joint Communication to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, 2018).

EU Commission, *Funding and Tender Opportunities, Disinformation and fake news are combated and trust in the digital world is raised* (TOPIC ID: HORIZON-CL3-2021-FCT-01-03) (2022). Available at: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-cl3-2021-fct-01-03.

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

European Commission, Directorate-General for Communication Networks, Content and Technology, *A multi-dimensional approach to disinformation: Report of the independent high level group on fake news and online disinformation* (Publications Office of the European Union, 2018).

European Union, *Charter of Fundamental Rights of the European Union*.

European Union, *Council Framework Decision 2008/913/JHA of 28 November 2008 on combating certain forms and expressions of racism and xenophobia by means of criminal law*.

European Union, *Council Regulation (EU) No 269/2014 concerning restrictive measures in respect of actions undermining or threatening the territorial integrity, sovereignty and independence of Ukraine* (2014).

European Union, *Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications) (e-Privacy Directive)*. Available at: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A32002L0058.

European Union, *Directive 2010/13/EU of the European Parliament and of the Council of 10 March 2010 on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (AVMSD), OJ L 95 of 15 April 2010*.

European Union, *Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA* (Law Enforcement Directive), (n. 55), Art. 2. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016L0680.

European Union, *Directive (EU) 2017/541 of the European Parliament and of the Council of 15 March 2017 on combating terrorism and replacing Council Framework Decision 2002/475/JHA and amending Council Decision 2005/671/JHA*. Available at: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

European Union, *EU Code of Practice on Disinformation* (European Union, 2018). Available at: https://digital-strategy.ec.europa.eu/en/library/2018-code-practice-disinformation.

European Union, *European Convention of Human Rights*.

European Union, *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC*, (n. 54), Art. 4(1). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32016L0680.

European Union, *Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online*.

European Union, *The Strengthened Code of Practice on Disinformation* (European Union, 2022). Available at: https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation.

European Union, *The Treaty on the Functioning of the European Union*.

European Union External Action Service, *1st EEAS Report on Foreign Information Manipulation and Interference Threats. Towards a framework for networked defence* (European Union, 2023).

Farinelli, 'Conspiracy theories and right-wing extremism – Insights and recommendations for P/CVE,' *Radicalisation Awareness Network (RAN)* (2021).

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

Gant Agreement.

Garaudy v. France App no. 65831/01 (ECtHR 24 June 2003).

Gunatilleke, 'Justifying limitations on the freedom of expression,' *Human Rights Review*, 22 (2021), 91-108.

Handyside v United Kingdom App no. 5493/72 (ECtHR 7 December 1976).

Hughes, Waismel-Manor, 'The Macedonian Fake News Industry and the 2016 US Election,' *PS: Political Science & Politics*, 54 (2021), 19-23.

International Institute of Business Analysis, *BABOK: A guide to the business analysis body of knowledge®* (2015). Available at: https://www.iiba.org/career-resources/a-business-analysis-professionals-foundation-for-success/babok/.

Johnson, Marcellino. *Bad Actors in News Reporting: Tracking News Manipulation by State Actors* (RAND Corporation, 2021).

Joshi, Kale, Chandel and Pal, 'Likert Scale: Explored and Explained,' *British Journal of Applied Science & Technology*, 7 (2015). Available at: doi: 10.9734/bjast/2015/14975.

Kling, Toepfl, Thurman and Fletcher, 'Mapping the website and mobile app audiences of Russia's foreign communication outlets, RT and Sputnik, across 21 countries,' *Harvard Kennedy School Misinformation Review* (2022). Available at: doi: 10.37016/mr-2020-110.

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

Kravchenko, Bogdanova, and Shevgunov, 'Ranking Requirements Using MoSCoW Methodology in Practice,' *Lecture Notes in Networks and Systems* (2022). Available at: doi: 10.1007/978-3-031-09073-8_18.

Kuczerawy, 'Fighting online disinformation: did the EU Code of Practice forget about freedom of expression?,' *Disinformation and Digital Media as a Challenge for Democracy: European Integration and Democracy Series*, 6 (2019).

Kuczerawy, *The proposed Regulation on preventing the dissemination of terrorist content online: safeguards and risks for freedom of expression* (Center for Democracy and Technology, 2018).

Levush, *Government Responses to Disinformation on Social Media Platforms: Argentina, Australia, Canada, China, Denmark, Egypt, European Union, France, Germany, India, Israel, Mexico, Russian Federation, Sweden, United Arab Emirates, United Kingdom* (The Law Library of Congress, Global Legal Research Directorate, 2019).

Lynas, 'COVID: Top 10 current conspiracy theories,' *Alliance for Science*, 20 April 2020. Available at: https://allianceforscience.org/blog/2020/04/covid-top-10-current-conspiracy-theories/.

Madsen, 'How To Prioritise Requirements With The MoSCoW Technique,' *Knowledgehut* (12 April, 2023). Available at: https://www.knowledgehut.com/blog/agile/how-to-prioritise-requirements-with-the-moscow-technique.

Marsden, Meyer, Brown, 'Platform values and democratic elections: How can the law regulate digital disinformation?,' *Computer law & security review*, 36 (2020).

Mines and Hughes, 'The Fractured Threat Landscape,' *Police Chief Magazine* (2022), 36-41. Available at: https://www.policechiefmagazine.org/fractured-threat-landscape/.

Monti, *The EU Code of Practice on Disinformation and the Risk of the Privatisation of Censorship, in Democracy and Fake News* (*Routledge*, 2020), 214-225.

Navarro, 'Free Speech: A Right in Crisis as Turkish Parliament Passes New "Disinformation" Bill,' *CICLR Online, 64* (2023), Available at: https://larc.cardozo.yu.edu/ciclr-online/64/.

NIT S.R.L. v. Republic of Moldova App no. 28470/12 (ECtHR 5 April 2022).

Ó Fathaigh, Helberger, Appelman, 'The perils of legally defining disinformation,' *Internet policy review*, 10 (2021), 2022-2040.

Parmenter, 'Background to the Winning KPI Methodology,' *Key Performance Indicators* (2019). Available at: doi: 10.1002/9781119620785.ch3.

Pech, *Concept of Chilling Effect: Its Untapped Potential to Better Protect Democracy, the Rule of Law, and Fundamental Rights in the EU* (Open Society Foundations, 2021).

Perinçek v. Switzerland App no. 27510/08 (ECtHR 15 October 2015).

Pielemeier, 'Disentangling Disinformation: What Makes Regulating Disinformation So Difficult?,' *Utah Law Review*, 917 (2020).

Plasilova et al., *Study for the assessment of the implementation of the Code of Practice on Disinformation* (European Commission, 2020). Available at: https://digital-strategy.ec.europa.eu/en/library/study-assessment-implementation-code-practice-disinformation.

Pustorino, *Introduction to International Human Rights Law* (Springer Nature, 2023).

Quinn and Malgieri, 'The Difficulty of Defining Sensitive Data—The Concept of Sensitive Data in the EU Data Protection Framework,' *German Law Journal*, 22 (2021), 1583-1612.

Rapp, Salovich, 'Can't We Just Disregard Fake News? The Consequences of Exposure to Inaccurate Information,' *Policy Insights from the Behavioral and Brain Sciences*, 5 (2019), 232–239.

Rice, 'Emotions and terrorism research: A case for a social-psychological agenda,' *Journal of Criminal Justice*, 37 (2009), 248-255.

Rotaru v. Romania App no. 28341/95 (ECtHR 4 May 2000).

Salov v. Ukraine App no. 65518/01 (ECtHR 6 September 2005).

Sander, 'Freedom of Expression in the Age of Online Platforms: The Promise and Pitfalls of a Human Rights-Based Approach to Content Moderation,' *Fordham International Law Journal*, 43 (2020).

Tenove, 'Protecting democracy from disinformation: Normative threats and policy responses,' *The International Journal of Press/Politics*, 25 (2020), 517-537.

The Observer and The Guardian v. United Kingdom App no. 13585/88, (ECtHR 26 November 1991).

United Nations General Assembly, *Countering disinformation for the promotion and protection of human rights and fundamental freedoms* (Report of the Secretary-General, 2022)

van Hoboken, O. Fathaigh, 'Regulating Disinformation in Europe: Implications for Speech and Privacy,' *UC Irvine Journal of International, Transnational, and Comparative Law,* 6 (2021).

van Prooijen, 'Voters on the extreme left and right are far more likely to believe in conspiracy theories,' *EUROPP – European Politics and Policy at LSE blog*. Available at: http://bit.ly/1zS8hW3.

Vese, 'Governing fake news: the regulation of social media and the right to freedom of expression in the era of emergency,' *European Journal of Risk regulation*, 13 (2022), 477-513.

Wardle, *Information disorder: Toward an interdisciplinary framework for research and policy making* (Council of Europe, 2017).

Yildirim, 'Silenced, Chilled, and Jailed,' *Verfassungsblog on matters constitutional* (2022). Available at: https://verfassungsblog.de/silenced-chilled-and-jailed/.

Annex A Questionnaire to the LEAs

FERMI PROJECT T2.1 End-user requirements elicitation

Fields marked with * are mandatory.

A WELCOME

Information for survey participants

Dear participant,
You are invited to take part in an Online Survey carried out as part of the Fake nEws Risk (https://fighting-fake-news.eu/) (101073980) project, a 3-year HORIZON EUROPE funded project, which started on 1st October 2022.
The following Survey aims at gathering user requirements of the FERMI Framework, concerning the various tools and solutions FERMI envisions to develop, in order to enhance the capacity security authorities across Europe in fighting against disinformation and fake news and the related crimes that may arise due to them.
Before you decide to provide us with your replies, please, be informed of the following details and, if you wish, consent to your participation by clicking the respective boxes in the EU Survey platform.

What is FERMI about?
Online social networks, news media and web platforms are the way contemporary societies operate for communication, information exchange, business, co-creation, learning and knowledge acquisition. However, the veracity of information circulating in the digital world is often in dispute. Indeed, disinformation and fake news (D&FN) increasingly affect and distort public opinion. National governments and supranational institutions recognize the spread of D&FN as a pernicious social problem. Indeed, the diffusion online of D&FN may have severe consequences. First, the spread of D&FN might infuse uncertainty and fear, intensify the crisis situations, and weaken the European societies aggravating their divisions. In turn, the increase in divisions and fear leads to episodes of physical violence offline and other hate crimes. As such, D&FN have the power to polarise public debates and put the health, security, and environment of EU citizens at risk. Finally, the use of fake accounts, the involvement of AI-generated fake content and the use of bots that can spread D&FN at scale pose additional problems.

FERMI will exploit a holistic and cross-disciplinary methodology towards a framework that will thoroughly analyse D&FN and their sources, in combination with all the socioeconomic factors that may affect both the spreading of such incidents and their effects on multiple dimensions of society. Comprising a set of innovative technological developments, FERMI will facilitate EU Police Authorities to detect and monitor the way that D&FN spread, both in terms of locations and within different segments of the society, and to put in place relevant security countermeasures; it will produce and diffuse tailormade training material designed for:

1. European Police Authorities,
2. Other professionals and stakeholders,

3. EU citizens for combating the spread and limiting the impact of D&FN and increasing digital trust.

The Consortium of this project is consisted of: Hochschule fur den offentlichen dienst in Bayern (BPA), ATOS IT solutions and services Iberia SL (ATOS), Intrasoft International SA (INTRA), Information technology for market leadership (ITML), INOV instituto de engenharia de sistemas e computadores-inovacao (INOV), Brandenburgisches institut fur gesellschaft und sicherheit ggmbh (BIGS), universita cattolica del sacro cuore Italy (UCSC), Ianus Consulting LTD (IANUS), The Lisbon Council for economic competitiveness asbl (LU), Convergence (CONV), Vrije Universiteit Brussel (VUB), Katholieke Universiteit Leuven (KUL), Poliisiammattikorkeakoulu (PUCF), Ministry of the interior of Finland (FMI), Police Federale Belge (BFP), Ministere de l'interieur (DMIA), Polismyndigheten swedish police authority see (SP).

What will you need to do?
For the purposes of the current research activity, the FERMI partners have prepared a questionnaire which you are asked to complete. In particular,
- You are called to answer a set of questions. Please be specific and short as possible.
- Please take into consideration that during the analysis of the results, all personal information will be anonymized.
- If some information is confidential, please mention that in your reply.

Who is the contact person?
For more information on this survey, you should contact Eleni Papargyri, Researcher at IANUS Consulting, e.papargyri@ianus-consulting.com, +357 24723131.

B Survey Information and Consent Form

B.1 Information Sheet concerning the participation to this Survey

Info_Sheet_T2.1_questionnaire.docx

B.2 Informed Consent Form concerning the participation in this Survey
Informed_Consent_T2.1_questionnaire.docx

Giving my consent, I undersign that:
a. I have carefully read and understood the attached Information Sheet.
b. I have carefully read and understood the attached Informed Consent and I agree with the terms and conditions.
c. I am fully aware of all my rights and especially my right to withdraw this consent, at any time, without consequences, by sending an e-mail to the Data Protection Officer of IANUS.

* B.3 Hereby I, freely and voluntarily consent to participate in the questionnaire distributed via EU Survey under the conditions set out in the Information Sheet.

- (●) Yes
- ( ) No

* B.4 Hereby, I agree to the participation in the FERMI project and to the use of my personal data that follows there out.

- (●) Yes
- ( ) No

* B.5 Name and Surname?

Please note that your personal data will not be publically available and the only person that will have access is the authorized Data Protection Officer from Ianus Consulting. This question aims to verify that you agree to the terms mentioned in the info-sheet and informed consent.

p['[

C End-User related questions

* C.1 What is your current LEA-related role?

- ( ) Active Duty
- ( ) Retired Active Duty
- ( ) Non-active duty LEA personnel
- ( ) LEA-affiliated (e.g., Police College personnel)
- ( ) LEA adviser
- ( ) LEA partner
- ( ) Other

C.3 Could you provide a brief description of your duties at the agency? (no more than 100 words, if possible)

* C.4 Is combating disinformation and fake news, as well as the crimes caused by them under your responsibilities?

- ( ) Yes
- ( ) No

* C.5 How often do you think the spread of disinformation and fake news leads to online or offline crimes?

- ( ) Never
- ( ) Rarely
- ( ) Sometimes
- ( ) Often
- ( ) Always

D End-Users' Requirements Questions

* D.1 According to your knowledge, does your agency have specific tools to examine if an account that is spreading fake news is a physical actor or a bot?

- ( ) Yes

- No

* D.2 Do you believe that an AI-based predictive tool capable of forecasting most likely offline and online crimes induced by identified D&FN can lead to better allocation of law enforcement recourses?
  - Yes
  - No

* D.3 What is the minimum accuracy level that is acceptable on the assessments of the origin of D&FN?
  - >50%
  - >60%
  - >70%
  - >80%

D.4 According to your knowledge, how important is it to examine if an account that is spreading fake news is a physical actor or a bot?

D.5 According to your knowledge, how important is it to:

	Not important	Slightly Important	Important	Fairly Important	Very Important
* examine if an account that is spreading fake news is a physical actor or a bot?					
* predict which kind of crimes the D&FN will eventually lead to?					
* constantly update one's insights into disinformation and fake news by processing textual content from websites and social media channels through machine learning					
* predict who are the potential victims and targets of crimes related to D&FN?					
* predict the environment and context in which the criminal event may occur due to the D&FN?					
* estimate the level of threat and risk posed by disinformation and fake news?					

* increase the knowledge of law enforcement and security officers about cultural aspects and the perception of disinformation among citizens?					
* create technological tools to enhance the capabilities of security officers to analyze in near-real-time large volumes of data to combat D&FN?					
* design and deliver mechanisms that determine economic and social factors that could influence the discourse outcomes and lead to potential polarization cases?					
* quantify the costs of violent extremism caused by disinformation and fake news?					
* grasp a community's resilience to disinformation and fake news?					
* develop a behavioral profiler to identify who is driving a campaign and for what purpose (e.g., for short-term disruption, long-term influence, economic damage, etc.)?					
* suggest countermeasures that can be taken proportionally to minimize the impact and the risk?					
* give an overview of the key proceedings in a user-friendly interface?					
* examine the causes, circumstances and consequences of disinformation and fake news rooted in right-wing extremism?					
* examine causes, circumstances and consequences of disinformation and fake news rooted in left-wing extremism?					
* causes, circumstances and consequences of disinformation and fake news rooted in health-related myths (i.e. false allegations concerning Covid-19)?					

D.6 If you have any further comments on what user requirements a platform capturing the causes, circumstances and consequences of disinformation and fake news should meet, please mention these below.

Contact

Contact Form

Annex B Information Sheet

Information Sheet

INFORMATION SHEET

FOR PARTICIPANTS PARTAKING IN A SURVEY TO IDENTIFY USER REQUIREMENTS OF A PLATFORM EXAMINING DISINFORMATION AND FAKE NEWS (D&FN) AS WELL AS PROPER USE CASES/SCENARIOS IN SUPPORT OF THE HORIZON EUROPE FERMI PROJECT (Project 101073980)

17.01.2023

Introduction

You have been invited to be a participant in a research activity carried out in the scope of the EU-funded Horizon Europe project FERMI (Fake nEws Risk MItigator). This document intends to provide you with detailed information concerning the research, in order for you to be able to take an informed decision on whether or not to participate.

Before making a decision, it is important that you clearly understand the purpose of the research and what it would imply for you. As such, please take time to read this document carefully and ask all the questions you may have so you can be completely sure that you understand all the proceedings of the research, including any potential risks and benefits it may entail.

This information document may include terms or concepts that you do not fully understand. If this is the case, please ask the contact person or any other member of the team coordinating the activity to fully explain it or clarify pieces of information.

At all times during the foreseen activities and afterwards, the FERMI consortium assures the compliance with the relevant national and European legislation.

Project and research ambitions

FERMI is an Horizon Europe project that studies and attempts to counter the root causes, spread and ramifications of D&FN. In this context, we are interested in identifying the proper user requirements of a platform that examines such proceedings as well as use cases and user scenarios that might be studied.

This information sheet informs you about the details and implications of participating in giving feedback on what user requirements, use cases and user scenarios you deem adequate.

Your participation

Your participation in the research activity described below would be highly valuable to meet the project's research ambitions. Your participation is only possible in case you freely consent to participating, which is the legal basis for taking part in the activity and for collecting and processing the data you and other research participants will be kindly asked to provide.

Taking part in this research project is voluntary

You are free to withdraw your participation from this activity at any time without providing justification. No consequences will follow your withdrawal. Any personal data of yours that might still be attributable to you will be immediately deleted by removing such data from the pool of to-be-analysed information. Any relevant files and hard copies will be destroyed.

Funded by the European Union

FERMI

The activity

Participants will be provided with this information sheet and asked to sign a consent form expressing their consent to participating and to sharing some personal data by filling in a questionnaire.

Data Collection and Processing

As mentioned above, this research is carried out under the HORIZON Europe project FERMI which processes personal data for the purposes of examining and countering the root causes, spread and implications of D&FN. Again, data processing takes place amidst the effort to identify the proper user requirements of a platform that examines such proceedings as well as use cases and user scenarios that might be studied. The FERMI research activities will take place between 01 October 2022 and 30 September 2025.

We do not plan to collect personal data other than the participants' views on the user requirements, use cases and user scenarios as well as their affiliation with LEAs (in the role of an active-duty LEA, former active-duty LEA, other LEA employee, LEA adviser/collaborator).

The data collected by asking participants to fill in the above-mentioned questionnaire is to be stored in a highly aggregated form.

Information that could identify participants will be removed from the data set prior to being analysed.

Again, we will not process any of your information if you do not give your consent; as explained above, if you change your mind, you can withdraw your consent at any time during or after the activity by reaching out to the responsible research coordinator(s) and/or the Data Protection Officer (contact details are given at the end of this document).

We will only collect and process the minimum amount of information that is required to achieve the purpose of this research activity. We will apply appropriate security measures during all processing and storage of your data (see below). All data collection and processing take place in strict accordance with the EU General Data Protection Regulation 2016/679. You'll be able to exercise all of your rights as a research participant in accordance with the EU General Data Protection Regulation 2016/679, which we fully respect and comply with; these are also described in this sheet.

No automated decision-making or profiling in accordance with Articles 22(1) and (4) of the General Data Protection Regulation 2016/679 is going to take place.

Will personal data be used for future research or shared with others?

We do not plan to share any personal data with anybody other than the research partners with a need to know (for an overview of the consortium, please consult the project website at https://fighting-fake-news.eu/) or use it for further research. If otherwise, the information will not be shared without your consent.

Will personal data be stored?

Access to the systems is limited to the research coordinator(s) (and their co-workers) in accordance with current security standards.

The data may be stored for the purpose of conducting a possible audit for five more years after the final payment. Thereafter, the data will be deleted from all storage media. No further processing will be conducted.

FERMI

Who may access my information?

We will disclose your information if the European Commission, who is funding this research, requests information for auditing purposes or to evaluate our procedures.

Within the consortium partners, your information will only be accessible on a need-to-know basis.

Your key data protection rights

In accordance with principles of research ethics and the EU data protection framework, you have rights regarding how your personal data is processed. Here are your rights and how we can fulfil them:

- Right to access personal data processed about you, and the right for these data to be in a portable form – If you request access to personal data that we hold about you, we will provide you with these data in an easily accessible format.
- Right to rectify personal data held about you – If you think the personal data that we hold about you is inaccurate or incomplete, you can request us to correct or complete your personal data.
- Right to restrict the processing of your personal data – If you want to restrict the way we process your personal data, you can request that we do so.
- Right to request that your personal data is erased – If you want us to delete your personal data from our systems, you can request that we do so.
- Right to leave the research activity – If you wish to withdraw from participating in this survey, you can do so at any time without negative consequences and your personal data will not be processed.
- Right to be informed about the purpose of the data collection, the use of the data and storage of the provided data.
- Right to request that we transfer your personal data to another organisation.
- Right to complain to a supervisory authority – If you feel we have not adequately dealt with your requests, you can complain to the national data protection authority. This research activity is managed by IANUS, and you can find information on the national data protection agency as follows: Independent advisory authority for the protection of the individual (commissionerdataprotection.gov.cy)
- Right to lodge a complaint to the project coordinator and/or the DPO (contact below).

Requests to exercise these rights will be handled without undue delay. If requests to exercise these rights are excessive, malicious, impossible to fulfil, or require a disproportionate effort, we may reject some requests in accordance with European and national data protection legislation.

What risks will I face by taking part in this study? What will be done to protect me against such risks?

Risks of discomfort resulting from this research may include the confrontation with controversial political proceedings. In principle, risk of breach of confidentiality in the handling of your personal data may lead to personal data breaches. Because this study collects information about you, one of the risks of this research is a loss of confidentiality.

Measures to protect against such risks:

- Voluntary participation
- Acknowledgement of the legal and ethical framework by partners
- Accountability and documentation
- Security of retention (appropriate security measures implemented in storing the data)
- Proper explanations of the policy context

FERMI

- Proper explanation as to how to use the technological tools
- The presence (online or offline) of technical experts that can be consulted

What are the benefits of the participation?

None of the participants will be paid for their participation. However, you may be able to learn about a steadily growing threat to domestic security, namely D&FN, timely efforts to examine and to counter this threat and – possibly – to connect with experts and observers with an interest in D&FN.

Confidentiality of the research/pilot study

All data that will be collected and processed will be treated highly confidentially. As explained above, no sharing will take place other than with consortium members that have a need-to-know (possibly except for EU requests, as explained above). Moreover, the data will be stored safely on servers to which only employees of the research partners have access. Appropriate security measures will be implemented to further ensure the confidentiality of data.

Contact details

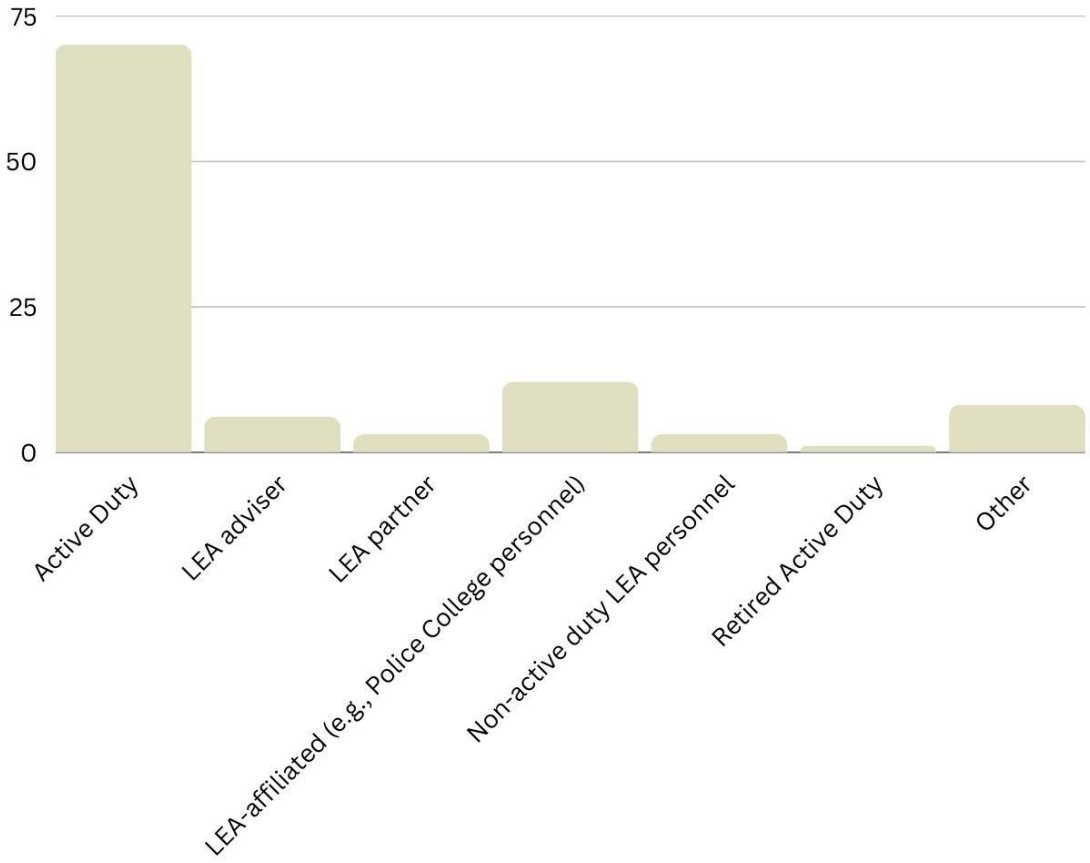
FERMI Project Coordinator: Dr. Holger Nitsch (holger.nitsch@pol.hfoed.bayern.de)

Data Protection Officer for this activity: Eleni Papargyri (e.papargyri@ianus-consulting.com

Research coordinator for this activity: Eleni Papargyri (e.papargyri@ianus-consulting.com

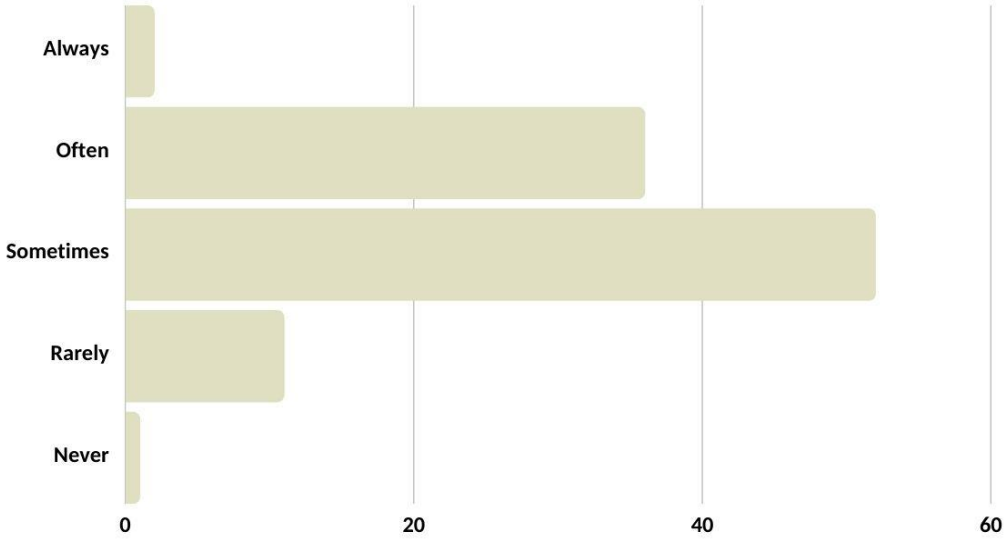
Annex C Results from End-Users Questionnaire

1.What is your current LEA-related role?

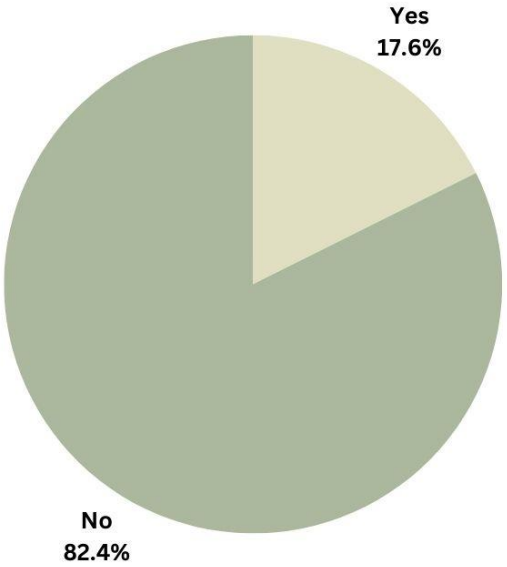


2. Is combating disinformation and fake news, as well as the crimes caused by them under your responsibilities?

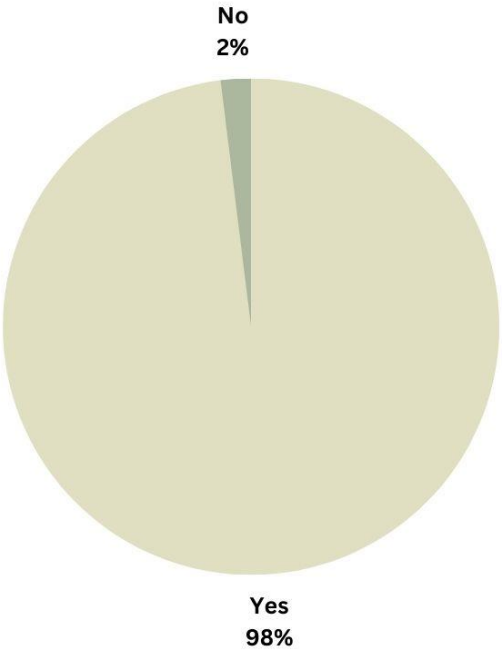
3. How often do you think the spread of disinformation and fake news leads to online or offline crimes?



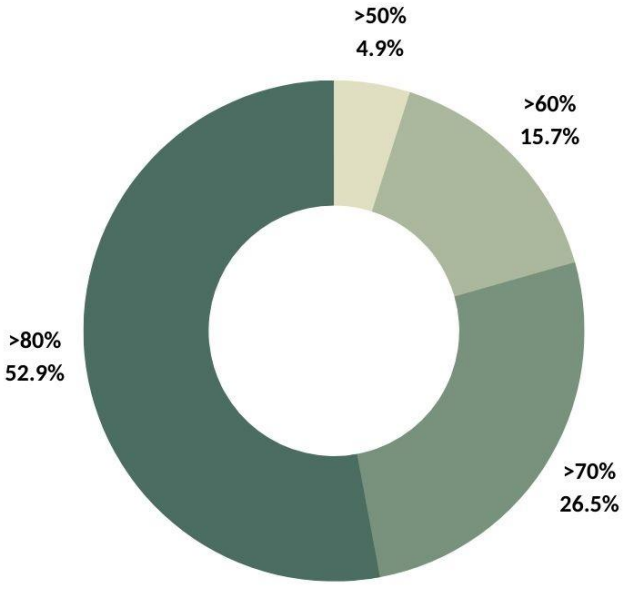
4. According to your knowledge, does your agency have specific tools to examine if an account that is spreading fake news is a physical actor or a bot?



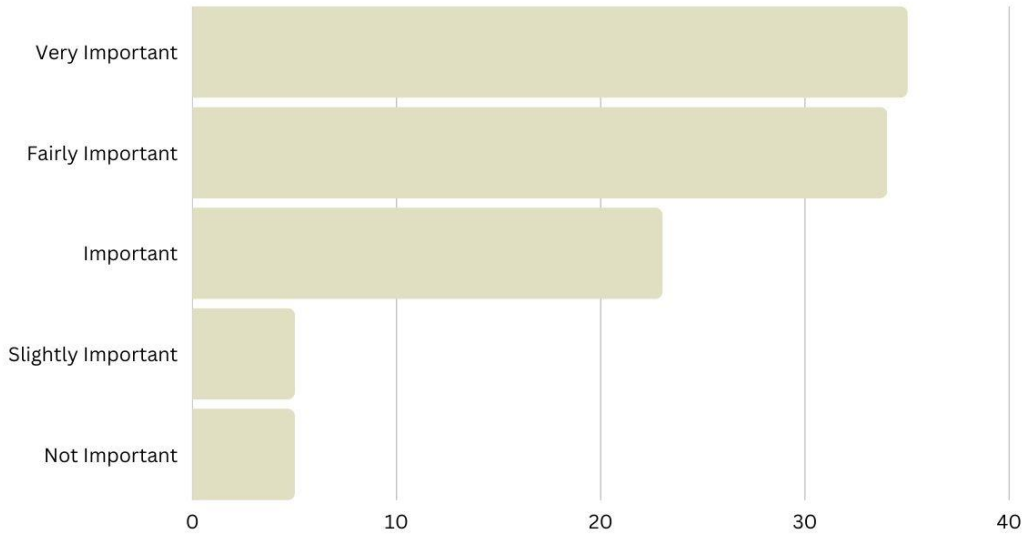
5. Do you believe that an AI-based predictive tool capable of forecasting most likely offline and online crimes induced by identified D&FN can lead to better allocation of law enforcement recourses?



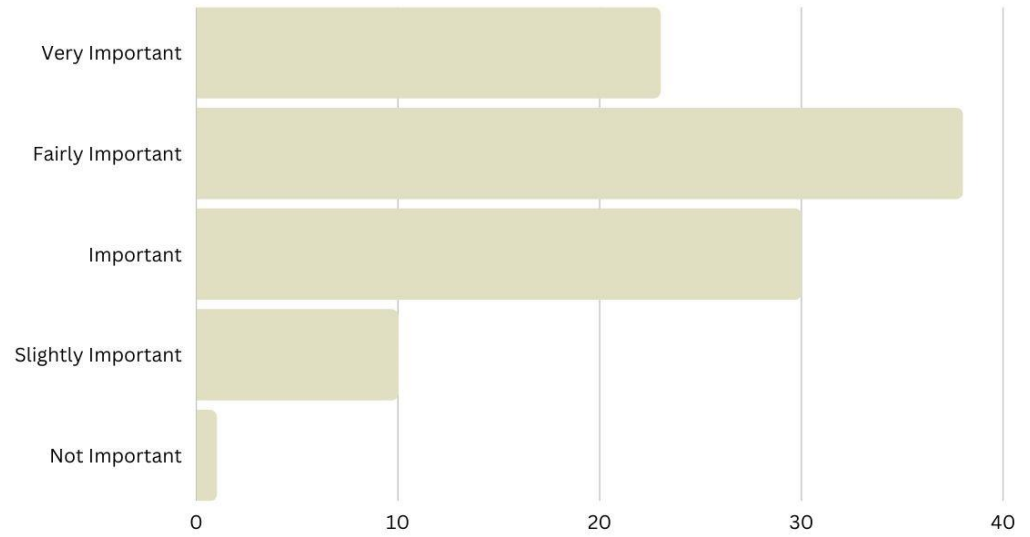
6. What is the minimum accuracy level that is acceptable on the assessments of the origin of D&FN?



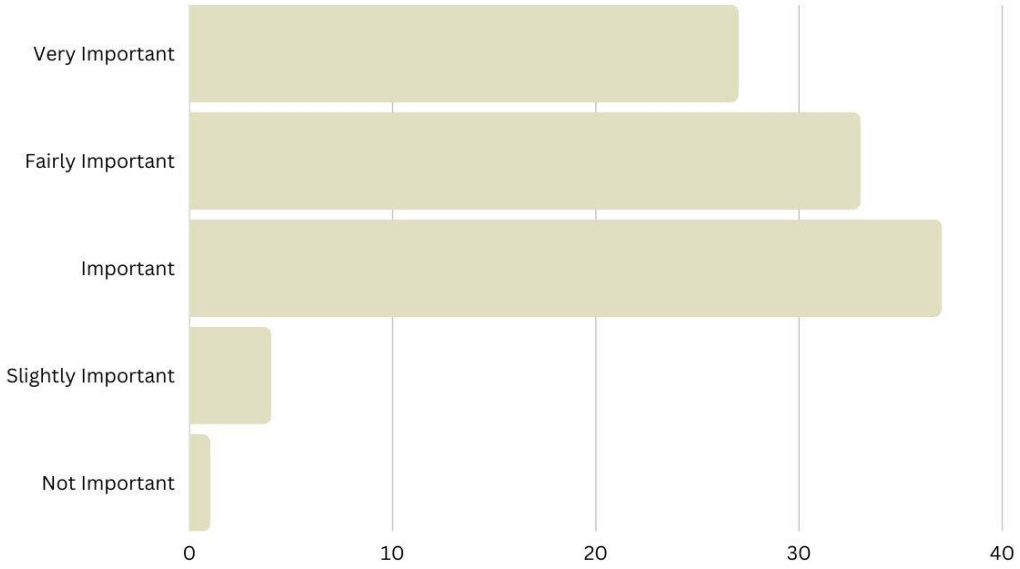
7. According to your knowledge, how important is it to examine if an account that is spreading fake news is a physical actor or a bot?



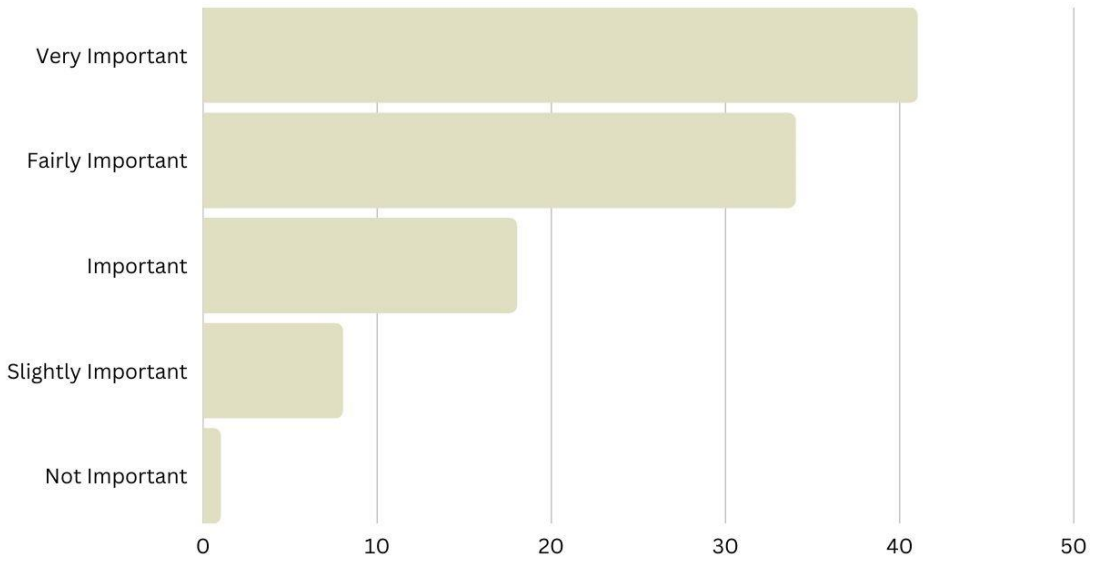
8. According to your knowledge, how important is it to predict which kind of crimes the D&FN will eventually lead to?



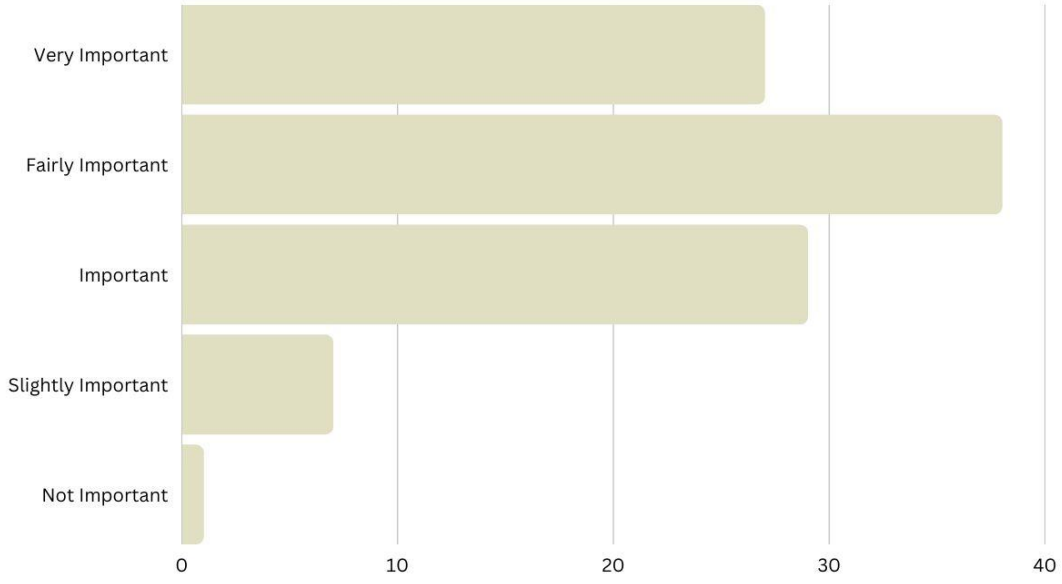
9. According to your knowledge, how important is it to constantly update one's insights into disinformation and fake news by processing textual content from websites and social media channels through machine learning?



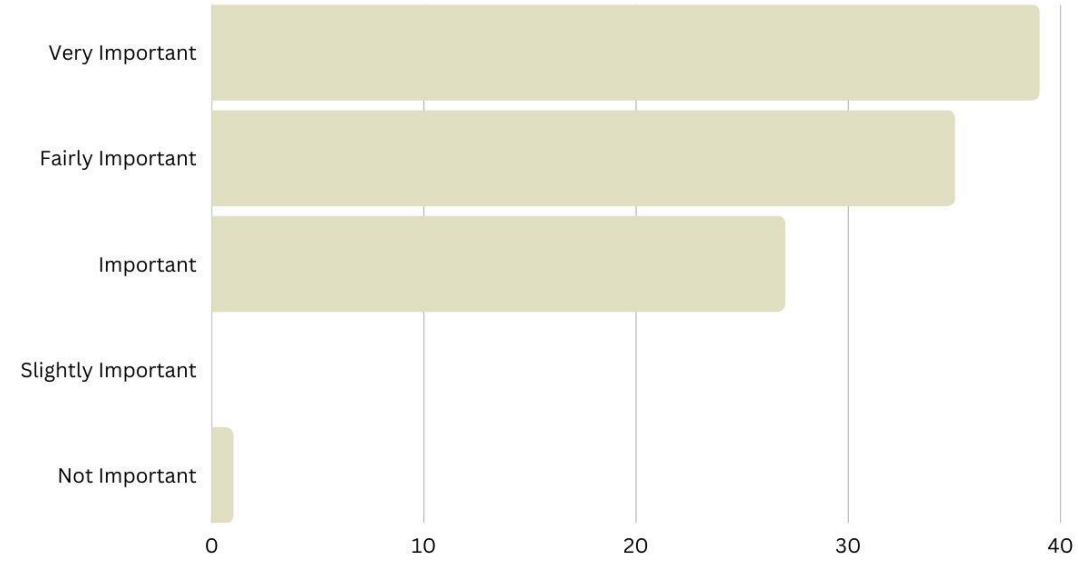
10. According to your knowledge, how important is it to predict who are the potential victims and targets of crimes related to D&FN?



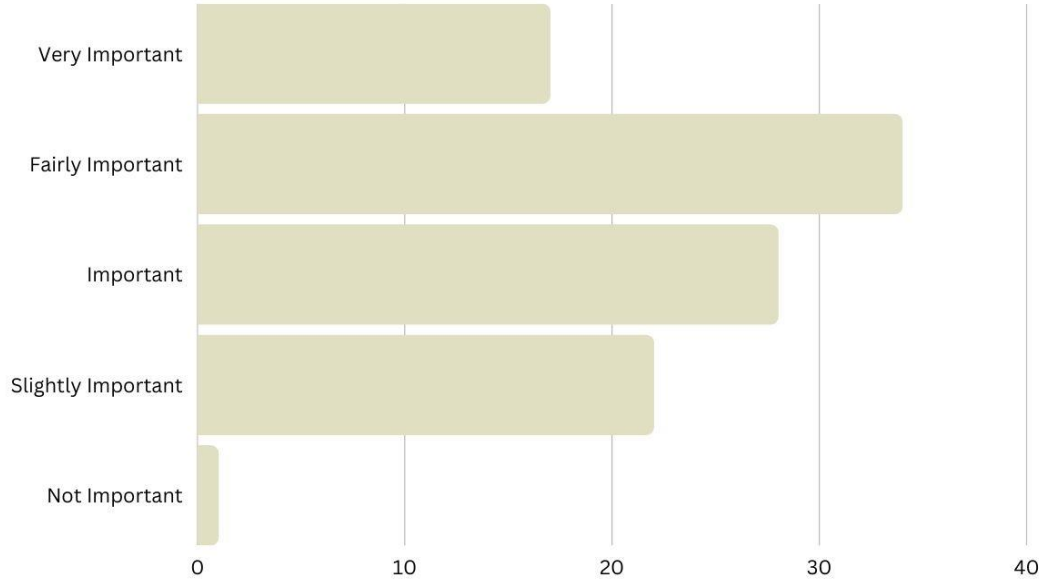
11. According to your knowledge, how important is it to predict the environment and context in which the criminal event may occur due to the D&FN?



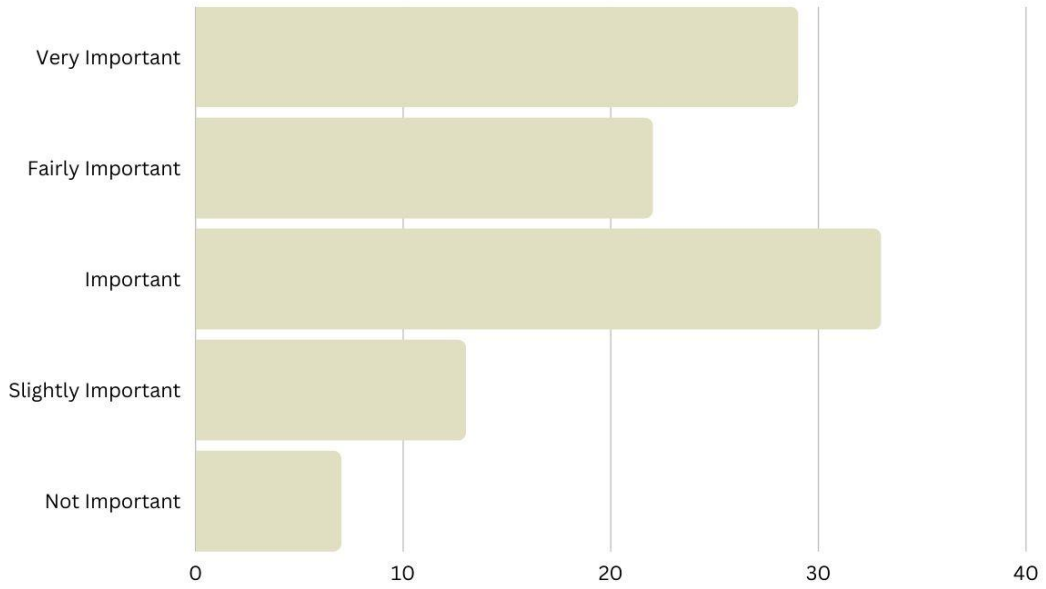
12. According to your knowledge, how important is it to estimate the level of threat and risk posed by disinformation and fake news?



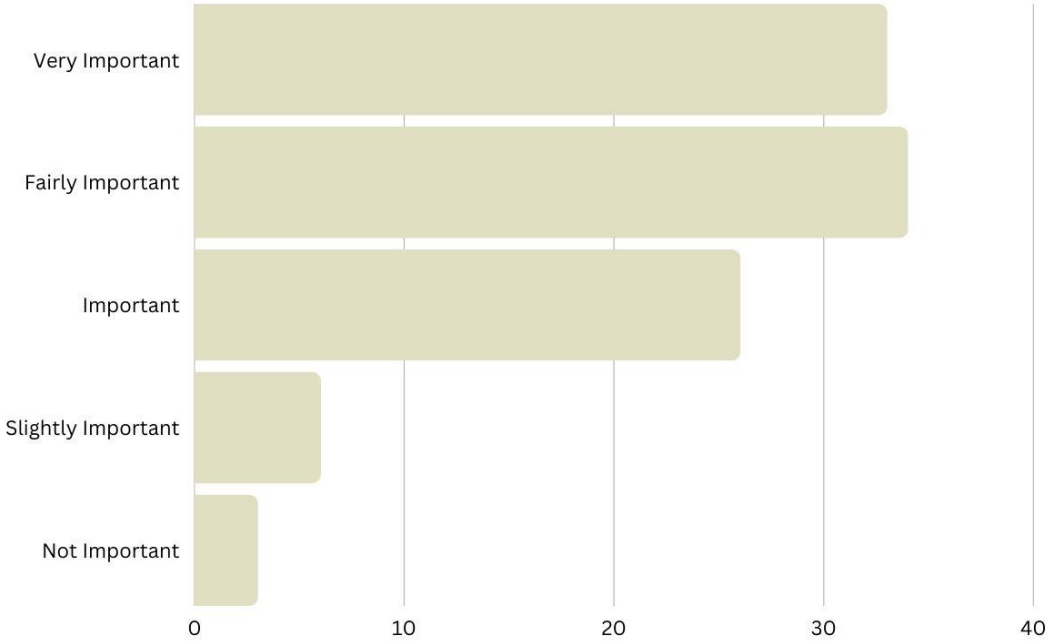
15. According to your knowledge, how important is it to design and deliver mechanisms that determine any economic factors that could influence the discourse outcomes and lead to potential polarization cases?



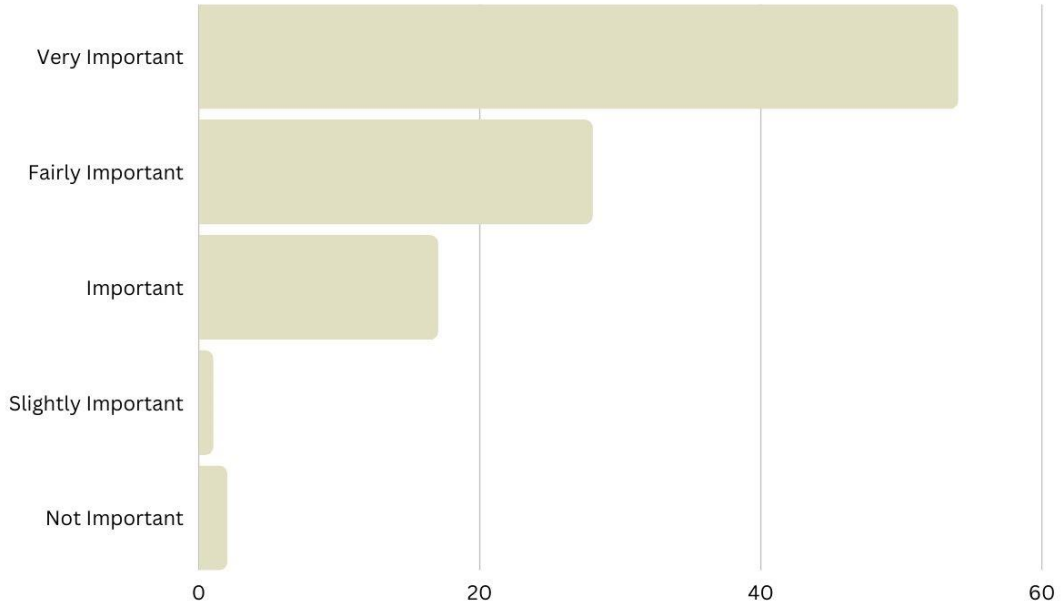
16. According to your knowledge, how important is it to quantify the costs of violent extremism caused by disinformation and fake news?



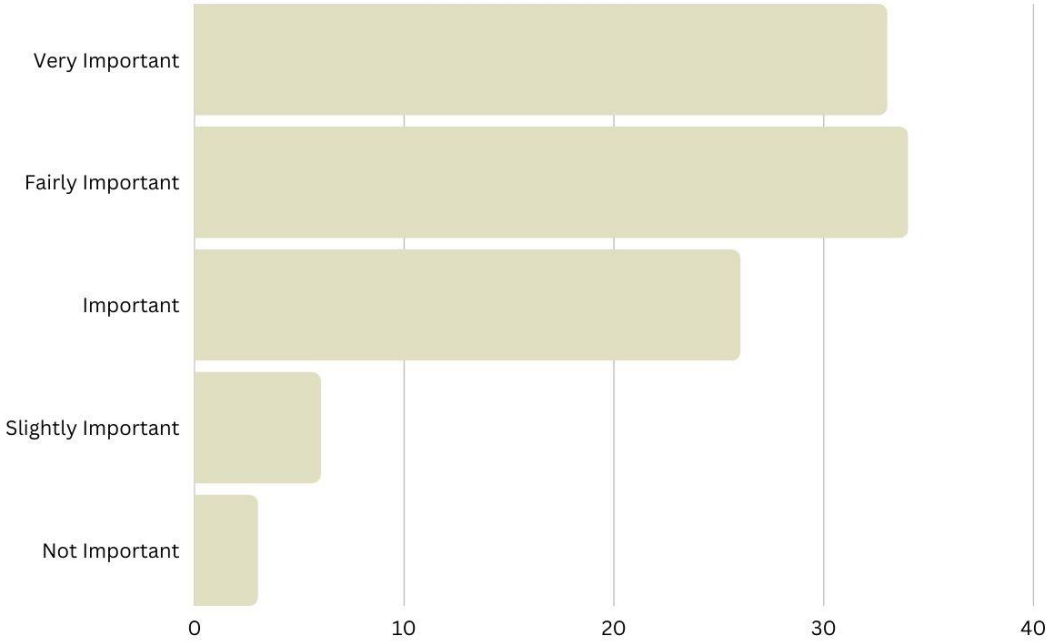
13. According to your knowledge, how important is it to: increase the knowledge of law enforcement and security officers about cultural aspects and the perception of disinformation among citizens?



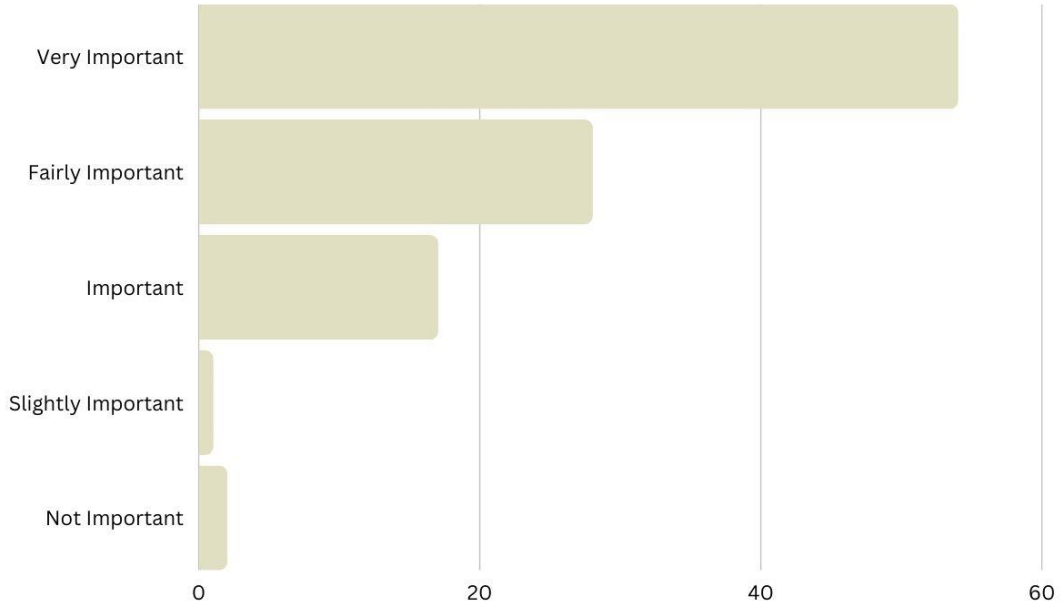
14. According to your knowledge, how important is it to create technological tools to enhance the capabilities of security officers to analyze in near-real-time large volumes of data to combat D&FN?



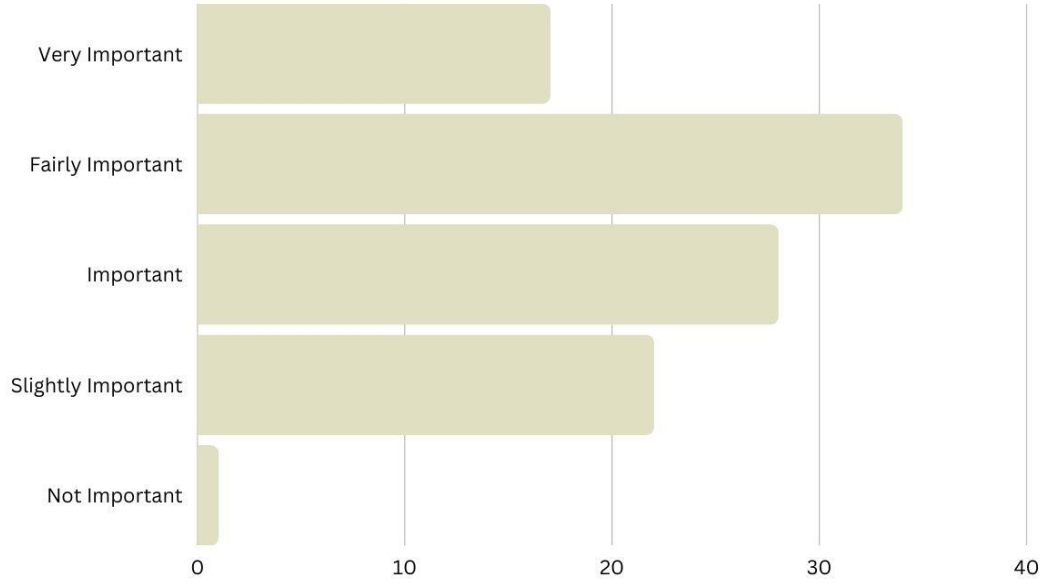
13. According to your knowledge, how important is it to: increase the knowledge of law enforcement and security officers about cultural aspects and the perception of disinformation among citizens?



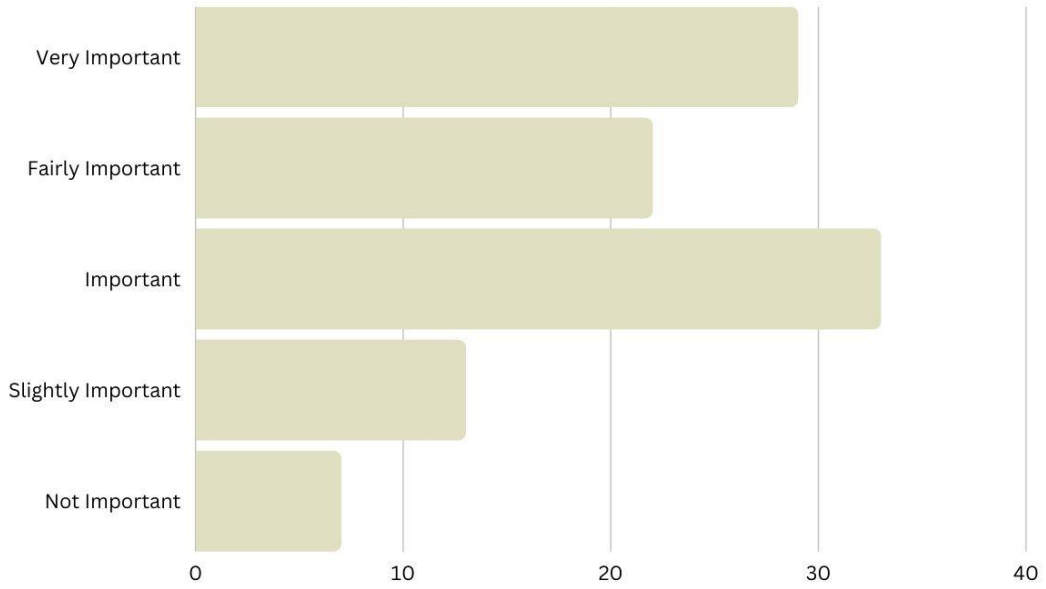
14. According to your knowledge, how important is it to create technological tools to enhance the capabilities of security officers to analyze in near-real-time large volumes of data to combat D&FN?



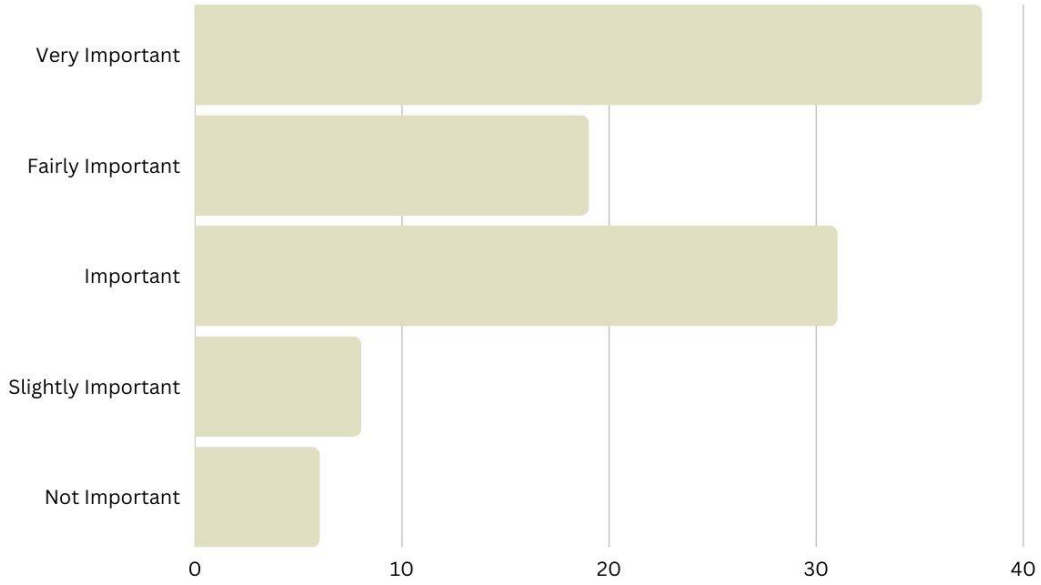
15. According to your knowledge, how important is it to design and deliver mechanisms that determine any economic factors that could influence the discourse outcomes and lead to potential polarization cases?



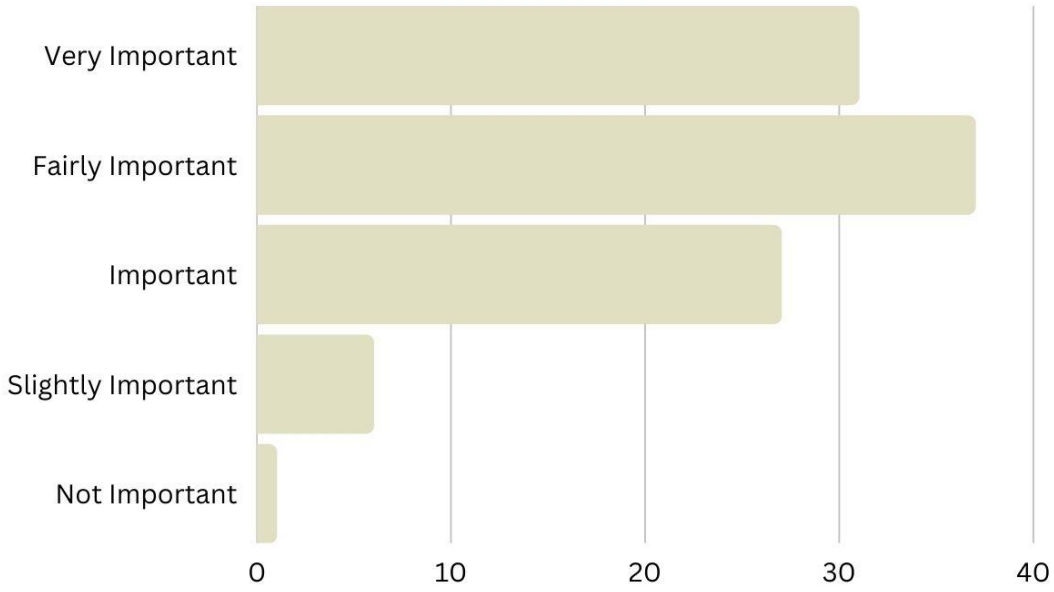
16. According to your knowledge, how important is it to quantify the costs of violent extremism caused by disinformation and fake news?



17. According to your knowledge, how important is it to grasp a community's resilience to disinformation and fake news?



18. According to your knowledge, how important is it to develop a behavioral profiler to identify who is driving a campaign and for what purpose (e.g., for short-term disruption, long-term influence, economic damage, etc.)?



Annex D Use Cases Template

Use Case Leader	
Use Case number	
Use Case Description	
Scenario	
Usability Evaluation	
Factors	
Actors	
Technologies currently available (Users)	
Technologies desired by the PROJECT platform	
To-be-examined End-User Requirements	
Goals and Objectives	