NOTIONES

iNteracting netwOrk of inTelligence and securIty practitiOners with iNdustry and acadEmia actorS

D5.4

Monitoring of EU Research and Horizon Scanning -v3

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101021853.

Project Details

Acronym: **NOTIONES**

Title: **iNteracting netwOrk of inTelligence and securIty practitiOners with iNdustry and acadEmia actorS**

Coordinator: **FUNDACIÓN TECNALIA RESEARCH & INNOVATION** (SPAIN)

Reference: 101021853

Type: Coordination and support action

Program: HORIZON 2020

Theme: Pan-European networks of practitioners and other actors in the field of security

Topic-ID: SU-GM01-2020

Start: 01.09.2021 – 31.08.2026

Duration: 60 months

Consortium:

Id	Participant Name	Short name	Country
1	FUNDACIÓN TECNALIA RESEARCH & INNOVATION	TECNA	Spain
2	ZANASI ALESSANDRO SRL	Z&P	Italy
3	LAUREA UNIVERSITY OF APPLIED SCIENCES LTD	LAU	Finland
4	BULGARIAN DEFENCE INSTITUTE	BDI	Bulgaria
5	DEFENCE RESEARCH INSTITUTE	DRI	France
6	FONDAZIONE ICSA – INTELLIGENCE CULTURE AND STRATEGIC ANALYSIS	ICSA	Italy
7	BAR ILAN UNIVERSITY EUROPE INSTITUTE	BIU	Israel
8	AGENCY FOR THE PROMOTION OF EUROPEAN RESEARCH	APRE	Italy
9	TEKNOLOGIAN TUTKIMUSKESKUS VTT OY	VTT	Finland
10	Expert.AI SPA	EXP.AI	Italy
11	SAHER EUROPE	SAHER	Estonia
12	MARKETSCAPE A/S	MS	Denmark
13	TECOMS SRL	TECOMS	Italy
14	SYNYO GmbH	SYNYO	Austria
15	REGIONAL POLICE HEADQUARTERS IN RADOM	KWPR	Poland
16	BULGARIAN STATE AGENCY FOR NATIONAL SECURITY	DANS	Bulgaria
17	CARABINIERI LT.GENERAL LEONARDO LESO	LESO	Italy
18	FINANCIAL INTELLIGENCE UNIT OF LATVIA	FIU	Latvia

20	ISEM-INTERNATIONAL SECURITY AND EMERGENCY MANAGEMENT INSTITUTE, n.p.o.	ISEMI	Slovakia
21	KHARKIV NATIONAL UNIVERSITY OF INTERNAL AFFAIRS	KhNUIA	Ukraine
22	POLITSEI.JA PIIRIVALVEAMET	EPBG	Estonia
23	MINISTRY OF INTERIOR OF GEORGIA	MIA	Georgia
24	POLICE SERVICE OF NORTHERN IRELAND	PSNI	UK
25	SWEDISH POLICE AUTHORITY	SPA	Sweden
26	POLICIA JUDICIARIA PORTUGUESE	PJ	Portugal
27	MILITARY ACADEMY "GENERL MIHAILO APOSTOLSKI" – SKOPJE	MAGMA	North Macedonia
28	HOCHSCHULE FÜR DEN ÖFFENTLICHEN DIENST IN BAYERN	HFOED	Germany
29	GOBIERNO VASCO - DEPARTAMENTO SEGURIDAD	ERTZ	Spain
30	BEYOND THE HORIZON	BTH	Belgium

Deliverable Details

Number:	**D5.4**
Title:	**Monitoring of EU Research and Horizon Scanning -v3**
Lead beneficiary:	APRE
Work package:	WP5
Dissemination level:	PU (Public)
Nature:	Report (RE)
Due date:	30th April 2023
Submission date:	27th April 2023

Authors: **Livia Di Bernardini, Claudio Testani**, APRE;
Giulia Venturi, Maria Ustenko, Z&P

Contributors: **Andrew Staniforth, David Fortune**, SAHER;
Ciro Caterino, Vincenzo Masucci, Exp.AI;
Tuomas Tammilehto, LAUREA

Reviewers: **Olatz Ibañez**, TECNA;
Edoardo Sponzilli, ICSA; **Alessandro Zanasi**, Security Advisory Board

Version History:

Date	Version No.	Author	Notes
10/03/2023	0.1	Z&P, APRE	ToC
20/03/2022	0.11	Z&P	Section 2.2 (disinformation)
22/03/2022	0.2	SAHER	Sections 4.2 (PROTECTOR project) and 4.3 (iCognative technology)
27/03/2023	0.25	Exp.AI	Section 4.1 (NLP against disinformation)
27/03/2023	0.3	APRE	Section 3
28/03/2023	0.31	Z&P	Internal review
03/04/2023	0.33	LAUREA	Contribution by LAUREA added
07/04/2023	0.34	Z&P	Additional contribution added
11/04/2023	0.4	Z&P	Sections 2.1 (cryptocurrencies) and D2.3 (misinformation), Executive summary, section 5 (Main findings) and section 6 (Conclusions) – version ready for first review
20/04/2023	0.5	ICSA	Review by ICSA
24/04/2023	0.6	TECNA	Reviewed by TECNA
26/04/2023	1.0	SAB, Z&P	Reviewed by SAB; Final Version
26/04/2023	1.1	Z&P	Final Version

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101021853

Disclaimer: The content of this report reflects only the authors' view. The European Commission or the Agency are not responsible for the content and any use that may be made of the information.

Table of Content

List of Figures

List of Tables

Acronyms

AI	Artificial Intelligence
ANZPAA	Australia New Zealand Policing Advisory Agency
AutoML	Automated Machine Learning
BERT	Bidirectional Encoder Representations from Transformers
BF	Brain Fingerprinting
EC	European Commission
ERP	Event-Related Potential
DL	Deep Learning
GANs	Generative Adversarial Networks
GPT-2	Generative Pre-trained Transformer 2
HINTS	Human Interaction News Trustworthiness System
IDM	IDentity Management
IoT	Internet of Things
IRS-CI	Internal Revenue Service's Criminal Investigation Division
LEA	Law Enforcement Agency
ML	Machine Learning
NER	Named Entity Recognition
NLP	Natural Language Processing
PSM	Pathogenic Social Media
RCCC	Red-Clear-Clear-Clear
RGB	Red-Green-Blue
WP	Work Package

Executive Summary

This document represents the product of tasks T5.2 “*Research monitoring on EU projects*” and T5.3 “*Research monitoring through Horizon Scanning*” of NOTIONES Work Package 5, dedicated to innovation monitoring.
The work was carried out by adopting the methodology outlined in NOTIONES deliverable D5.1 “*Methodology for Innovation Monitoring*”.
The research activities and the findings are those obtained in months M19 and M20 since the beginning of the project (third run of the tasks).

Section 1 introduces the document by describing the work frame of tasks T5.2 and T5.3, and of the overall Work Package 5 of NOTIONES.

Section 2 reports on the research activities carried out in task T5.3 “*Research monitoring through horizon scanning*”. The main data source used was TheLens, but also open web and CORDIS were exploited. Datasets were explored searching for publications of relevance for the current NOTIONES *focus areas*.

Section 3 reports on the research activities carried out in task T5.2 “*Research monitoring on EU projects”* through cascade refinement stages of research, to identify the most interesting projects in terms of relevance for the NOTIONES *focus areas*.

Section 4 presents additional findings on selected relevant research topics: disinformation detection technology through NLP, the PROTECTOR project, and the iCognative technology.

Section 4.3 contains a summary of the most relevant findings of both tasks T5.2 and T5.3 through the common layout for the summarisation of the information proposed in deliverable D5.1. The following research projects and technologies are presented and their possible exploitation in NOTIONES is proposed: TRACE project, FERMI project, Chainalysis solutions, and PROTECTOR/PRECRISIS surveillance monitoring system.

Section 6 contains conclusive considerations and next steps.

1. Introduction

NOTIONES (iNteracting netwOrk of inTelligence and securIty practitiOners with iNdustry and acadEmia actorS) is a CSA (Coordination and Support Action) project, funded by the European Commission (EC), and aims to facilitate the supply side - academia, SMEs, and research centres - and demand side - security and intelligence practitioners - of Security innovation meet. The project results are expected to strengthen the European integration in the fields of Security and Intelligence, identifying the needs of Intelligence and Security practitioners.

NOTIONES Work Package WP5 aims at identifying new technologic opportunities and terrorist threats to support the European Security Research and Innovation by providing fresh inputs to reshape its research and development activities in order to directly address the practitioners' needs.

Tasks **T5.2 *"Research monitoring on EU projects"*** and **T5.3 *"Research monitoring through Horizon Scanning"*** of WP5 are dedicated to **innovation monitoring**, defined as the activity aimed at gaining understanding of important technological trends, along with their Intelligence and Security implications, by finding and interpreting the available information in order to provide a concrete benefit to the NOTIONES network of stakeholders.

This document represents the product of the third run of tasks T5.2 and T5.3 of WP5, which performed the research monitoring activities during M19 and M20, as depicted in Figure 1.

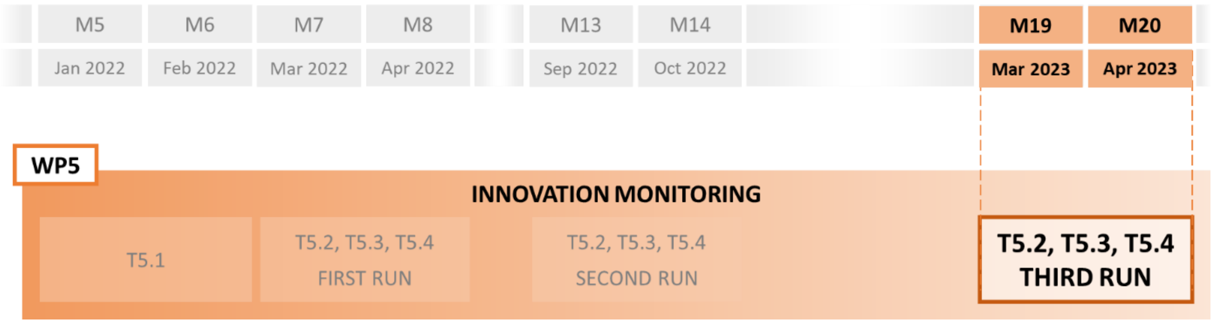


Figure 1 - Time diagram of the third run of innovation monitoring in WP5

For the reader's convenience, the tasks' descriptions are recalled below:

- T5.2 Research monitoring on EU projects: this task will perform a deep survey of the emerging technologies that are the most promising in the field of intelligence and security by exploiting the great variety and volume of knowledge produced by EU research projects. To this purpose, the project will rationalize and categorize knowledge exploiting the CORDIS database as a primary source for information. In addition to this, the expertise of all NOTIONES partners will be exploited.
- T5.3 Research monitoring through Horizon Scanning: this task will perform a deep survey of the emerging technologies that are the most promising in the field of intelligence and security by exploiting the great variety and volume of knowledge openly available. To this purpose, the project will rationalize and categorize knowledge exploiting open databanks of publications and patents. The gathering of information will be performed by a targeted search based on the keywords, by means of technology horizon scanning. "Horizon scanning" is intended as

systematic research of relevant technological developments with the purpose of highlighting opportunity and threats that may influence the capability of organizations and bodies providing intelligence and security services to achieve their objectives ad goals. Such analysis should also consider the maturity level of technologies, so to identify whether it is at research phase, development, prototyping or production.

It is worth reminding that task T5.4 "*Monitoring of emerging terrorist threats*" reports the findings in a separate deliverable, namely D5.13 "*Monitoring of Emerging terrorist threats -v3*".

1.1 Structure of the document

Section 1 introduced the document by describing the work frame of tasks T5.2 and T5.3, and of the overall Work Package 5 of NOTIONES.

Section 2 reports on the research activities carried out in task T5.3 "*Research monitoring through horizon scanning*".

Section 3 reports on the research activities carried out in task T5.2 "*Research monitoring on EU projects".*

Section 4 presents additional findings on selected relevant research topics.

Section 4.3 contains a summary of the most relevant findings of both tasks T5.2 and T5.3.

Section 6 contains conclusive considerations and next steps.

2. Research monitoring through Horizon Scanning

An essential part of the research monitoring activity is represented by Horizon Scanning, intended as a systematic research of technology trends with the purpose of highlighting opportunity and threats that may influence an organisation's capability to achieve its objectives ad goals – i.e., in NOTIONES, the security and intelligence practitioners' capability to operate.

Horizon Scanning aims at detecting new technologies, rapidly evolving and increasingly being adopted by industries, but also, in regard to already existing technologies, new combinations of such, transfer of technologies to other domains and/or new applications of existing technologies.

For the third run of task T5.3, Horizon Scanning was performed by researching technologies through the analysis of free online scholar and patent databanks.

The methodology adopted for task T5.3 originates from the methodology delivered in D5.1 "*Methodology for innovation monitoring*". The main data source used was TheLens [1], using the integrated search engine on scholarly works and patents and its export functionality, which allows to export up to 50.000 results in .csv, .ris, .json or BibTeX format. Apart from TheLens, open web and CORDIS [2] were also exploited. The datasets were primarily explored with the online statistical analysis tool of TheLens.

The research was performed by Mrs. Giulia Venturi (Orcid ID: 0000-0003-0445-2613) and Ms. Maria Ustenko (Orcid ID: 0000-0002-6506-7607) of Consortium partner Z&P.

Mrs. Venturi holds a Master's Degree in Physics in the University of Bologna (Italy) with Internship at the University of Cambridge (UK). She is expert in technology horizon scanning and in methodologies for strategic technology foresight.

Ms. Ustenko holds a Bachelor in Chemistry and Master in Nanotechnology. She graduated from PFUR, Engineering Academy led by Russian Space Association. She has both academic and industrial working experiences. Currently she is working as a technical researcher in the field of Artificial Intelligence.

With regard to the issues encountered and search features, the explanations provided in the first version of this deliverable (D5.2) remain valid.

In the next subsections, the results of the horizon scanning activities performed in the third run of WP5 are presented.

The focus areas tackled in this run of the horizon scanning task are the current focus areas tackled by the currently active NOTIONES working groups, namely:

- Tools for tracing cryptocurrencies used in criminal finances;
- AI for Disinformation validation in hybrid influencing.

2.1 Cryptocurrencies

The first horizon scanning performed in the third run of task T5.2 was made on *TheLens* using the keyword "*cryptocurrency*", revealing more than 2000 on scholar works and 140 patents.

These results were refined based on the indications coming from the NOTIONES Working Group WG3 "*Tools for tracing cryptocurrencies used in criminal finances*", which was active for quite some time before starting the current run of the task T5.2. This gave the researches the opportunity to count on the WG members' opinion and adapt the research according to the emerging requests coming from the practitioners participating in the working groups. They are interested in tracing new types of cryptocurrencies, in particularly crypto-currency based on the CryptoNote protocol, focused on increased transaction confidentiality - Monero. The interest of the practitioners was based on the fact that in Monero it is not possible to disclose information about transactions using specialized tools, although the theoretical possibility of tracking transactions does exist. However, it seems that it will not be the case for long time, as in 2020, the US Internal Revenue Service's Criminal Investigation Division (IRS-CI) put out reward for stockholders to develop tools to track Monero. The contract was awarded to blockchain analysis company Chainalysis and Integra FEC. Currently the tool is under development and many other blockchain analysis groups invest funds into designing to the similar tools, including European based companies as Cellebrite and others.

Following this input, another horizon scanning was performed similarly using the keywords "*monero* AND *tracing*". Without setting a date range limit, this search gave a result of 23 scholar works, four out of which were works cited by patents. The topic of tracing Monero and other privacy-enhanced cryptocurrencies, as well as Monero itself is a fairly new topic. Therefore, it was expected that the result of the research turned out to be few. However, it is possible to see a sharp increase in the interest in these areas, and as a result, a rebound in the trend is expected in the future.

The diagram of the research is shown in Figure 2, to assist the reader as he/she progresses in reading the report.

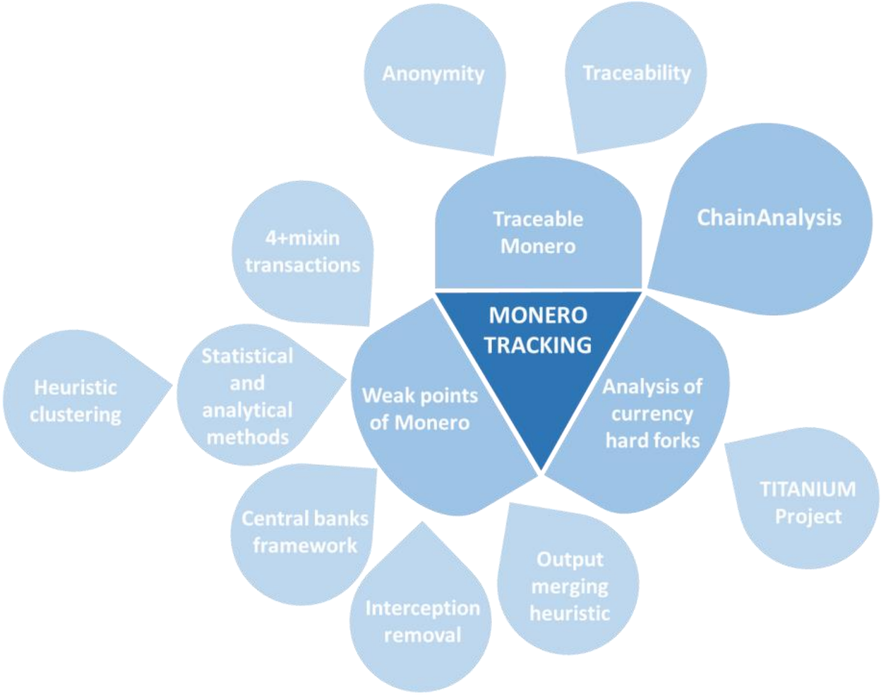


Figure 2 - Diagram of the horizon scanning research on cryptocurrency tracking

One of the scholar works found cited by patents was published in 2021 by Li et al. [3]. The work recognises that cryptocurrency can be the subject and means of committing crimes. Monero is considered an untraceable cryptocurrency and therefore many criminal acts can be committed with anonymity protection in cryptocurrency transactions. To assist law enforcement in investigating a crime, the authors propose to introduce a new cryptocurrency called Traceable Monero. The proposed traceable Monero scheme works in such a way that it makes it possible to track both (see Figure 3) – (i) disposable cash flow addresses and (ii) long-term addresses. Moreover, even though the system relies on a tracing authority, it is only activated in case of certain transactions that need to be investigated. Therefore, it gives a positive response to conflict resolution between the two fundamental divisions of security requirements in transactions with cryptocurrency, balancing such properties as anonymity and traceability. The security of Traceable Monero has been tested and demonstrating valuable results. Furthermore, the system prototype demonstrates that Traceable Monero has little overhead in creating and validating a transaction, unlike usual Monero transactions.

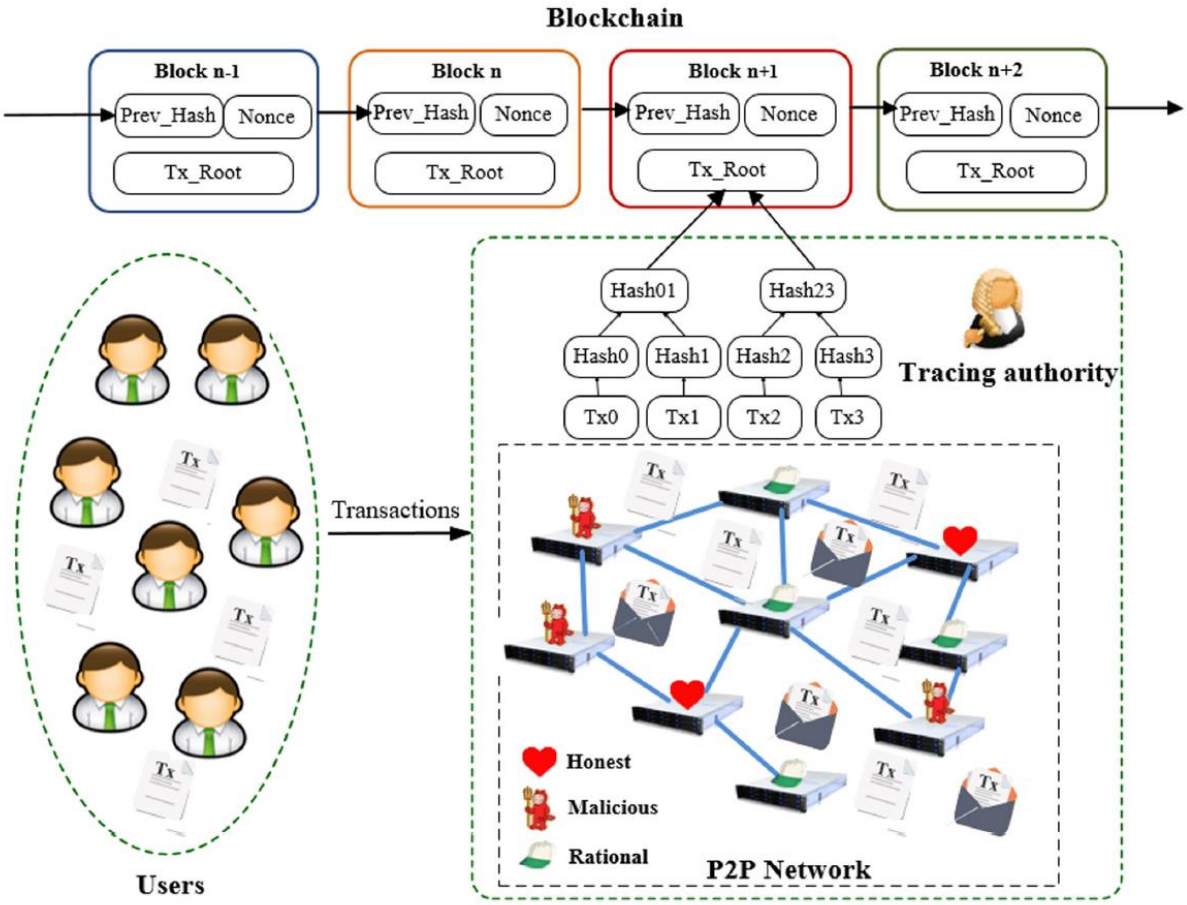


Figure 3 - System model of traceable Monero [3]

Another significant work was performed by Hinteregger and Haslhofer in 2019 [4]. The authors introduce a brand-new method for tracing Monero transactions, by analysing currency hard forks (Figure 4). They tested the efficiency of the method by performing passive analysis for traceability of data from monero, monerov, and monero original blockchains. Even recognizing the efficiency of their method, the authors do admit the complexity of the design of studied cryptocurrency, which most probably will remain resistant to the conventional tracing methods applied for any other cryptocurrency, making Monero invulnerable to the vector attack, at the same time throttling the developers in Monero community. This work was carried out within the European Commission funded project TITANIUM (Project ID: 740558). The project was identified while horizon scanning and reported in the subsection 3.1 and in previous runs of the NOTIONES research.

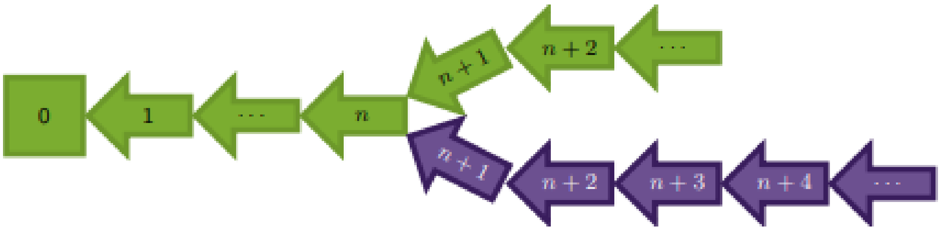


Figure 4 - Illustration of a currency hard fork [4]

Another interesting paper was published by Möser *et al.* [5]. This paper is particularly interesting as the authors do not offer a complete tool or a method to trace Monero transactions, but they rather perform the analysis of the weak points of the Monero cryptocurrency, which can be later used for designing the tracing tool. The authors demonstrate at least three shortcoming points where the non-traceability could be improved. For example, Monero introduced 4+mixin transactions which however are easy to be deanonymized. Also, the authors proved that about 91% of transactions that were run by non-RingCT protocol can be traced by using GuessNewest heuristic approach. Cao et al. in 2020 [6] also studied the weakness points of Monero, focusing on the potential vulnerabilities od peer-to-peer network, which Monero is built on. The paper suggests that by isolating the set of nodes this type of network may fail an attack on the blockchain safety.

Another method for tracing that is based on transaction analysis was studied by Barcelo [7] and by Koshy et al. [8]. The scientists proved that over 62% of transactions may be deducted by using a combination of statistical and analytical methods. Barcelo proved that by using with some external information, one can find true identities. Reid et al. [9] grouped wallets by using heuristic clustering and eventually deanonymize transactions in Ripple network.

On the other hand, it is possible not only trace but also control at present time the transactions basing on a central party. Möser et al. proposed a protocol that holds several banks in its core and manages to control all the transactions on behalf of users, while Danezis et al. [10] introduced a framework to manage cryptocurrencies that demands the central banks to command over the policies.

More scientist researched the traceability of Monero. They distinguish the methods of tracing in many categories, one of which is heuristic group of methods, which aims those few weaknesses that the cryptocurrency has. Möser et al. proposed to use the intersection removal to trace the transactions, while Kumar et al. [11] counted on output merging heuristic, as it is suggested to be the real outputs that may be followed. Both the papers suggest monitoring the repeated applications called zero mixin removal, as they may result in a chain reaction. It was proved that each ring (which is used in every transaction and mixes the output with decoy outputs) has one member and can be traced as any other cryptocurrency, i. e Bitcoin.

It is worth mentioning that almost every scholar work mentions the approaches and the future works of the technology provider Chainalysis [12]. Chainalysis is an American blockchain analysis group active since 2014. For many years they have been cooperating with law enforcement agencies, helping them to conduct finance crime investigations, covering dark web marketplaces as well. Chainalysis group has already been identified by the researchers of NOTIONES during the very first run of horizon scanning and was later voted with the high esteem by all the working groups active at that time. The researchers attempted to contact the group with no success at that time. It is planned, however, to reach the company out again to establish contact with the company and practitioners from NOTIONES.

2.2 Disinformation

The second horizon scanning performed in the third run of task T5.2 took as a starting point the NOTIONES focus area about *disinformation* and proceeded in a pure exploratory way, through wide-ranging research.

The diagram of the research is shown in Figure 5, to assist the reader as he/she progresses in reading the report.

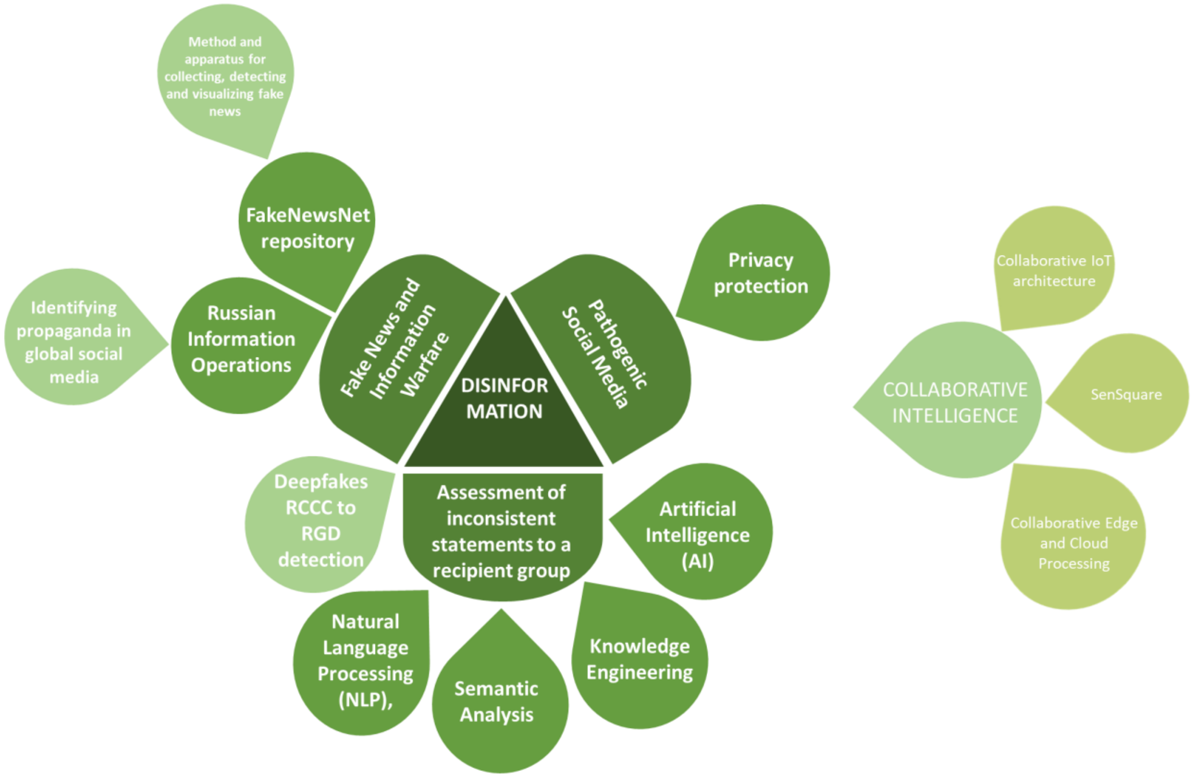

Method and apparatus for collecting, detecting and visualizing fake news
FakeNewsNet repository
Identifying propaganda in global social media
Russian Information Operations
Fake News and Information Warfare
DISINFOR MATION
Pathogenic Social Media
Privacy protection
Deepfakes RCCC to RGD detection
Assessment of inconsistent statements to a recipient group
Artificial Intelligence (AI)
Natural Language Processing (NLP),
Semantic Analysis
Knowledge Engineering
Collaborative IoT architecture
COLLABORATIVE INTELLIGENCE
SenSquare
Collaborative Edge and Cloud Processing

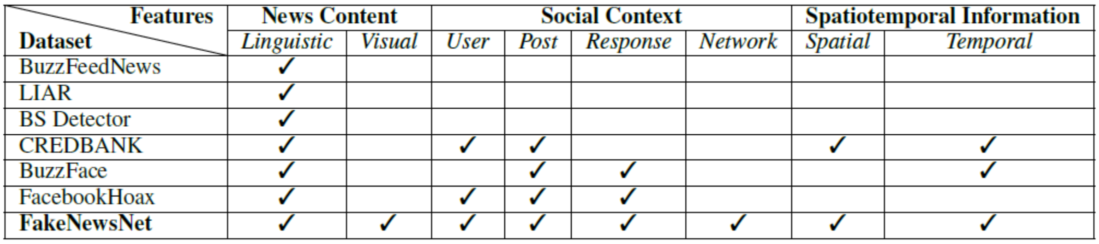
Figure 5 - Diagram of the horizon scanning research on disinformation

The search on the online scholar database *TheLens* with keyword "*disinformation*" in all search fields, with no date range limitation, led to 9077 results, of which 39 results had citing patents.

The research focused on scholarly publications with citing patents, considering the enhanced technological content at pre-development stage.

An interesting work published in 2018 by Alvari et al. from the Arizona State University [13] was found, related to the identification of Pathogenic Social Media (PSM) accounts such as terrorist supporters who exploit large communities of supporters for conducting attacks on social media. Early detection of these accounts is crucial as they are high likely to be key users in making a harmful message *viral*. The authors use causal inference to identify PSMs within a short time frame around their activity. The proposed algorithm is applied to groups of accounts sharing similar causality features and is followed by a classification algorithm to classify accounts as PSM or not. The authors claim that this approach applied to a real-world dataset from Twitter demonstrates effectiveness and efficiency, with precision of 0.84 for detecting PSMs only based on their first 10 days of activity. This work by Alvari et al. is cited by the patent "*Privacy protection systems and methods*" [14] published on September 2022 in the US jurisdiction, which does not seem particularly relevant for NOTIONES.

A second work, published in 2020 by Shu et al. from the Arizona State University [15] and cited by hundreds of publications since, presents a fake news data repository called *FakeNewsNet*, which contains two comprehensive datasets with diverse features in news content, social context, and spatiotemporal information. The repositor is built by utilizing fact-checking websites to obtain news contents for fake news and true news such as PolitiFact [16] and GossipCop [17].



Features / Dataset	News Content		Social Context				Spatiotemporal Information	
	Linguistic	*Visual*	*User*	*Post*	*Response*	*Network*	*Spatial*	*Temporal*
BuzzFeedNews	✓							
LIAR	✓							
BS Detector	✓							
CREDBANK	✓		✓	✓			✓	✓
BuzzFace	✓			✓	✓			✓
FacebookHoax	✓		✓	✓	✓			
FakeNewsNet	✓	✓	✓	✓	✓	✓	✓	✓

Figure 6 - Comparison of FakeNewsNet with existing news detection datasets [15]

This work by Shu at al. is the main source cited by the patent "*Method and apparatus for collecting, detecting and visualizing fake news*" [18], published on November 2022 in the US jurisdiction by the Arizona State University. The authors propose to create a hierarchical macro-level propagation network of the fake and real online news articles (comprising news nodes, social media post nodes, and social media repost nodes) and a hierarchical micro-level propagation network of the fake and real online news articles (comprising reply nodes). Then, an analysis is performed on the structural and temporal features of the hierarchical macro-level propagation network, and on the structural, temporal, and linguistic features of the hierarchical micro-level propagation network. Fake news is identified among the online news articles based on the analysis of such features.

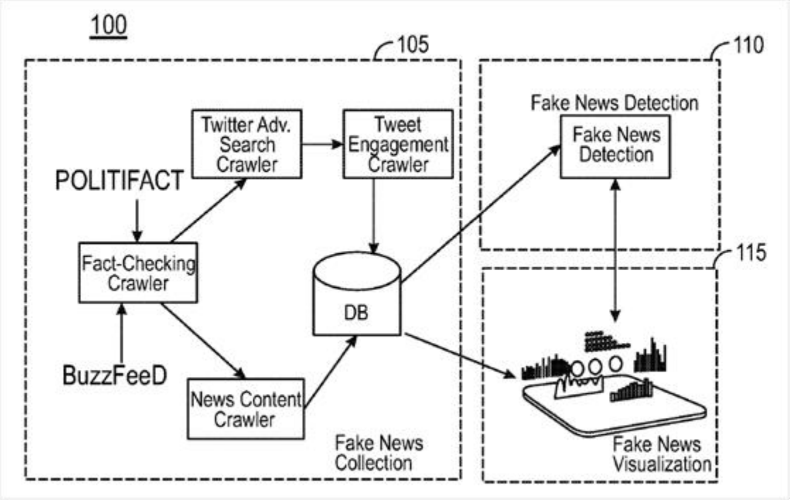
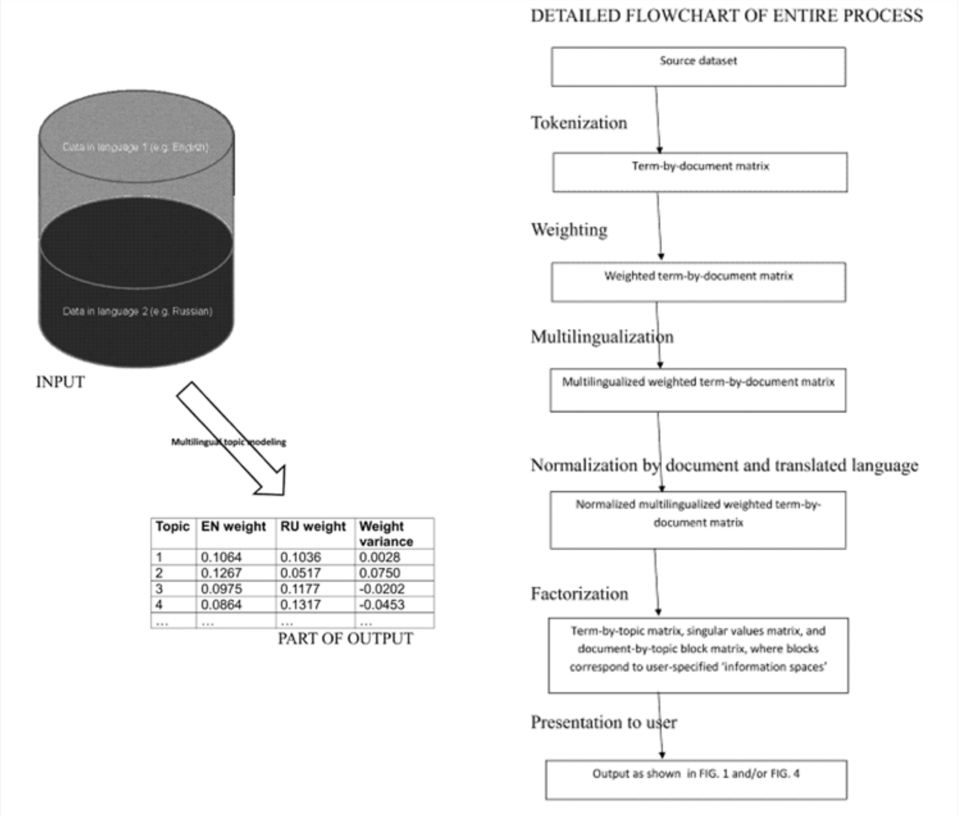


Figure 7 - Schematic description of patent US 11494446 B2

In general, the Arizona State University seems to be very active in the area od Fake News detection.

Another work, published in 2017 by Chew at al. [19], reports about Russian Information Operations researched by means of Unsupervised Multilingual Topic Modeling. Topic modelling is a technique widely used to detect hidden topicality of text corpora, including those from social media [20]. The authors look at differences between what is discussed in Russian language versus what is discussed in English language, using unsupervised machine learning to detect large-scale trends, anomalies, similarities and differences. Applying this approach to different Twitter datasets, the authors claim to be able to "*draw out several interesting and non-obvious insights about Russian cyberspace and how it differs from its English counterpart*". They "*show how these insights reveal aspects of how master* *narratives are instantiated, and how sentiment plays out on a large scale, in Russian discourse relating to NATO".*

This work by Chew at al. is cited by the patent "*Identifying propaganda in global social media*" [21], published on November 2018 in the US jurisdiction by GALISTEO Consulting Group Inc., a small consulting firm offering services in the critical areas of national security, policy analysis, business planning, and corporate culture change based in New Mexico (US).


Data in language 1 (e.g. English)
Data in language 2 (e.g. Russian)
INPUT
Multilingual topic modeling
PART OF OUTPUT
DETAILED FLOWCHART OF ENTIRE PROCESS
Source dataset
Tokenization
Term-by-document matrix
Weighting
Weighted term-by-document matrix
Multilingualization
Multilingualized weighted term-by-document matrix
Normalization by document and translated language
Normalized multilingualized weighted term-by-document matrix
Factorization
Term-by-topic matrix, singular values matrix, and document-by-topic block matrix, where blocks correspond to user-specified 'information spaces'
Presentation to user
Output as shown in FIG. 1 and/or FIG. 4

Topic	EN weight	RU weight	Weight variance
1	0.1064	0.1036	0.0028
2	0.1267	0.0517	0.0750
3	0.0975	0.1177	-0.0202
4	0.0864	0.1317	-0.0453
...	...	...	...

Figure 8 - Schematic description of patent US 10140289 B2

The patent regards a "*scalable method for automatically deriving the topics discussed most prevalently in unstructured, multilingual text, and simultaneously revealing which topics are more biased towards one or another information space*"[1]. This method allows to determine which topics are more discussed, for example, in one language than another. An analyst's attention can then be "*focused on the most important differences between national discourses, and insight more quickly gained into the areas (both topics and geographic regions) in which propaganda of the sort envisaged in Russian strategic doctrine may be taking hold*".

Further work published in 2019 by McCloskey [22] reported on Image forensics addressed to online disinformation campaigns, with focus on Generative Adversarial Networks (GANs), e.g. deepfakes, that can be trained to generate synthetic imagery which is in some ways indistinguishable from real imagery. The authors analyzed the structure of the generating network of a popular GAN implementation [23], and showed that "*the network's treatment of exposure is markedly different from a real camera"* and that "*this cue can be used to distinguish GAN-generated imagery from camera imagery, including effective discrimination between GAN imagery and real camera images used to train the GAN*".

[1] Here, the concept of the *information space* is derived from Russian strategic doctrine on information warfare; an example of an *information space* would be the portion of social media in which the Russian language is used.

This work by McCoskey is cited by the patent "RCCC to RGB domain translation with deep neural networks" [24], published on July 2021 in the US, Chinese and German jurisdictions by Ford Global Technologies Llc. The patent discloses a system and a method for translating, e.g., mapping, a Red-Clear-Clear-Clear (RCCC) image to a Red-Green-Blue (RGB) image. In an example implementation, the system and the method can receive, at a deep neural network, an image having a Red-Clear-Clear-Clear (RCCC) image pattern, wherein the deep neural network includes a generator and a discriminator; generate, at the generator, a Red-Green-Blue (RGB) image based on the image having the RCCC image pattern; determine, at the discriminator, whether the RGB image is machine-generated or is sourced from the real data distribution; and update at least one weight of the generator when the discriminator determines the RGB image is machine-generated.

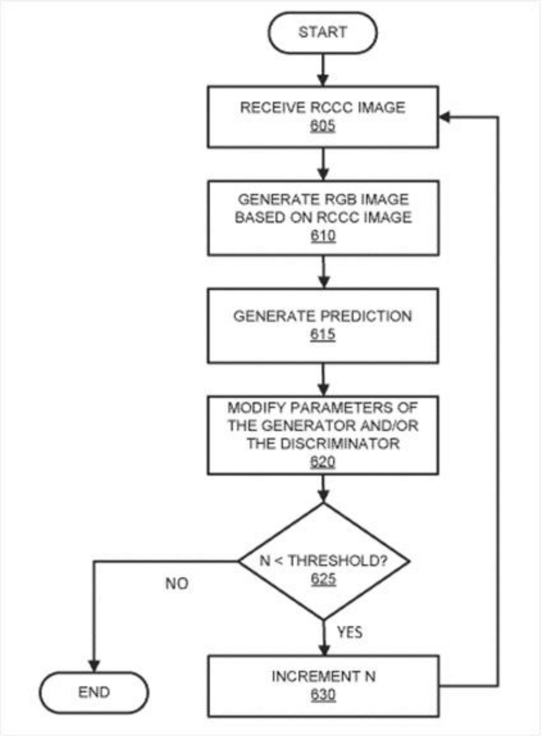

START
RECEIVE RCCC IMAGE
605
GENERATE RGB IMAGE BASED ON RCCC IMAGE
610
GENERATE PREDICTION
615
MODIFY PARAMETERS OF THE GENERATOR AND/OR THE DISCRIMINATOR
620
N < THRESHOLD?
625
NO
YES
END
INCREMENT N
630

Figure 9 - Schematic description of patent US 11068749 B1

In addition to this, Lakshmanan et al. published in 2019 a work on data mining methods to detect fake news online [25]. The paper disseminates the efforts of different communities on combating fake news by providing "*a panoramic view of the state-of-the-art of research on various aspects including detection, propagation, mitigation, and intervention of fake news*". Then, it provides a concise and intuitive summary of prior research on "*data integration, truth discovery and fusion, probabilistic databases, knowledge graphs and crowdsourcing from the lens of fake news*". This work is particularly interesting because it is cited by the patent "Method for detecting false messages in social media" [26], published on December 2020 in the Chinese jurisdictions by the Northwestern Polytechnical University in Xi'an.

Among the patents investigated, another one appears highly interesting for NOTIONES. The International Business Machines Corporation (IBM) published in September 2022 a patent titled "Assessment of inconsistent statements to a recipient group" [27], which discloses embodiments that relate to a machine-implemented method for "*assessing a statement by an entity to a recipient group, particularly in a social media environment, for an inconsistency with a previous (historical) statement to the same recipient group, and notifying the entity of the inconsistency. In certain embodiments, an* *alternative statement is provided*". The patent mentions Natural Language Processing (NLP), Semantic Analysis, Knowledge Engineering and Artificial Intelligence (AI) methods.

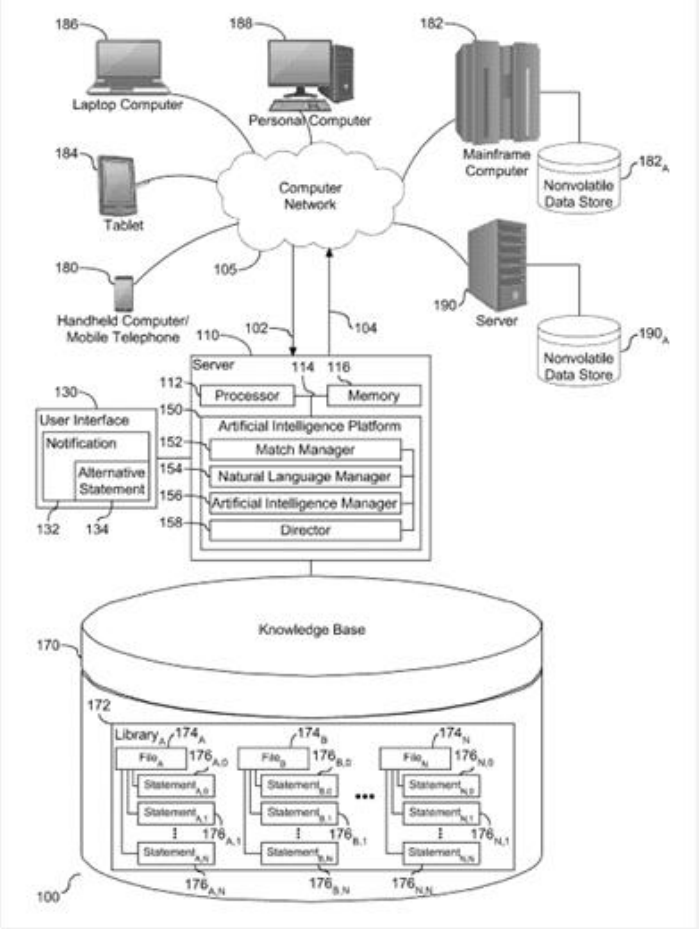


Figure 10 - Schematic description of patent US 11443208 B2

The search about "Disinformation" on the patent database of *TheLens* led also to other interesting results. For example, Microsoft Technology Licensing Llc published the patent "Data privacy pipeline providing collaborative intelligence and constraint computing" [28] on September 2022. The patent discloses techniques for deriving collaborative intelligence based on constraint computing or constraint querying: "*At a high level, a data trustee can operate a trustee environment that derives collaborative intelligence subject to configurable constraints, without sharing raw data. The trustee environment can include a data privacy pipeline through which data can be ingested, fused, derived, and sanitized to generate collaborative data without compromising data privacy. The collaborative data can be stored and queried to provide collaborative intelligence subject to the configurable constraints. In some embodiments, the data privacy pipeline is provided as a cloud service implemented in the trustee environment and can be spun up and spun down as needed.*"

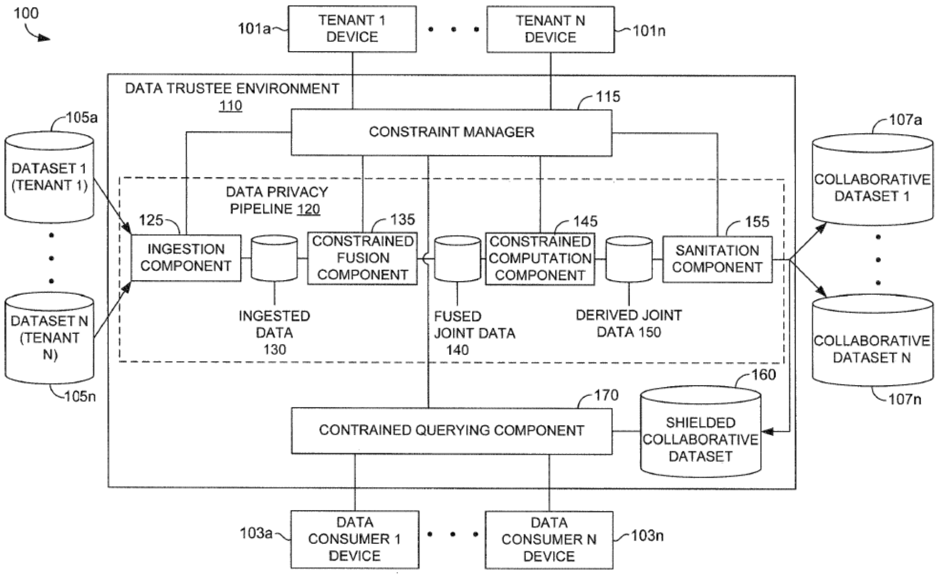


Figure 11 - Schematic description of patent US 11455410 B2

This patent cites the paper “A Collaborative Internet of Things Architecture for Smart Cities and Environmental Monitoring” by Montori et al. [29] about the concept of crowdsensing in order to produce the amount of data that Internet of Things (IoT) scenarios need in order to be pervasive. The authors introduce an architecture, namely *SenSquare* [30], able to handle both the heterogeneous data sources coming from open IoT platform and crowdsensing campaigns and display a unified access to users. The platform deals with heterogeneous data classification, mobile crowdsensing management for environmental data, information representation, and unification, IoT service composition and deployment.

The patent also cites the paper “Live Data Analytics With Collaborative Edge and Cloud Processing in Wireless IoT Networks” by Sharma et al. [31] about big data analytics applied to IoT. Here, the key challenges are how to “*effectively extract useful features from the massive amount of heterogeneous data generated by resource-constrained IoT devices in order to provide real-time information and feedback to the end-users, and how to utilize this data-aware intelligence in enhancing the performance of wireless IoT networks*”. As the authors explain: “*Although there are parallel advances in cloud computing and edge computing for addressing some issues in data analytics, they have their own benefits and limitations. The convergence of these two computing paradigms, i.e., massive virtually shared pool of computing and storage resources from the cloud and real-time data processing by edge computing, could effectively enable live data analytics in wireless IoT networks*”. In this regard, the authors propose a novel framework for coordinated processing between edge and cloud computing/processing that can exploit the network-wide knowledge and historical information available at the cloud center to guide edge computing units towards satisfying various performance requirements of heterogeneous wireless IoT networks. They also identify and describe the potential key enablers for the proposed edge-cloud collaborative framework, the associated key challenges and some interesting future research directions.

2.3 Misinformation

The third horizon scanning performed in the third run of task T5.2 focused on the area of *misinformation*. The scanning of the scholar works and patents using the keyword "*misinformation*" revealed more than one and a half thousands of related works, including patents. Below the most interesting findings are reported.

In the first patent "Mitigating misinformation in encrypted messaging networks" granted in 2021 to the IBM corp. [32] , a method for mitigating misinformation in encrypted messaging environments is proposed. This method involves receiving content from an originating user, encrypting it, and storing the content on a messaging server. The method is very complete as includes:

(i) *"receiving content from an originating user,*
(ii) *encrypting the content into an originating message using a first encrypting key, appending an originating message identifier to the originating message,*
(iii) *storing the originating message identifier on a messaging server in conjunction with transmitting the originating message to a first device corresponding to a first recipient,*
(iv) *decrypting the originating message using a first decrypting key,*
(v) *storing the content on the first device to produce locally stored content*
inserting the originating message identifier within metadata for the locally stored content"

The second patent "Method and system for detection of misinformation" by Accenture [33] proposes a system for automatically detecting misinformation using a cross-stitch based semi-supervised end-to-end neural attention model that can generalize and identify emerging misinformation. The author disclosed a new method for detecting misinformation. It provides the output via an attention mechanism, which means that the network pays more attention to smaller however more significant pieces of information. The scientists implemented the proposed misinformation detection system using a semi-supervised cross-stitch-based end-to-end neural attention model that is tuned to use the large amount of raw data available. The cross-stich model can detect and colligate emerging misinformation content since it trains using massive relevant knowledge, coming from heterogeneous sources, as social medias posts user details, users' activities, etc.

In the (pending) patent "System and method for detecting misinformation and fake news via network analysis" by prof. Elan Pavlov from MIT [34], a method for detecting misinformation without analyzing articles is proposed, using a graph containing users and articles with weights for user and article nodes. This significant work on detecting misinformation was performed by the same scientist in 2022. Prof. Pavlov patented a new system for detecting misinformation (HINTS - the Human Interaction News Trustworthiness System). The idea behind the system is to predict a possible post or a news article that may be liked by the user, based on his previous behaviour.

The system uses a mixed graph that includes both user and article nodes and edges between them. It also uses seed nodes that are manually labelled as trustworthy or untrustworthy to help train the algorithm to detect fake news. The algorithm then runs for a predetermined number of rounds to update the weights of both the user and articles and converge the graph. The output is a set of highest weights for users and/or articles that may be flagged for possible remedial action (Figure 12).

Although the proposal is worth to be mentioned in the study as the algorithm might bring significant results during the investigations, it is fair to say that there were no official records detected of the system to be tested by security practitioners or any type of law enforcement agencies while running this research.

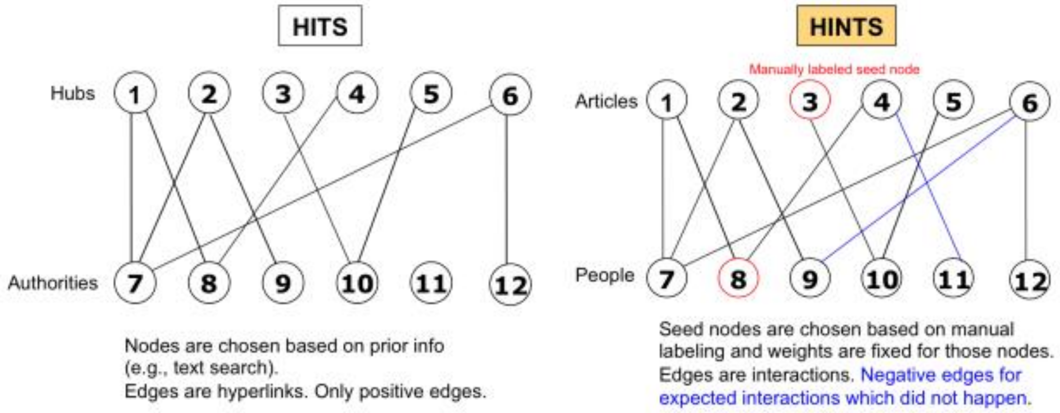

HITS
Hubs
Authorities
1
2
3
4
5
6
7
8
9
10
11
12
Nodes are chosen based on prior info (e.g., text search).
Edges are hyperlinks. Only positive edges.
HINTS
Manually labeled seed node
Articles
People
1
2
3
4
5
6
7
8
9
10
11
12
Seed nodes are chosen based on manual labeling and weights are fixed for those nodes.
Edges are interactions. Negative edges for expected interactions which did not happen.

Figure 12 - HINTS algorithm method. [35]

The (pending) patent [35] proposes a method for presenting thought-provoking information about a media product to reduce misinformation consumption.

Another (pending) patent [36] by Ericsson Telefon Ab proposes a method for generating machine learning models that selectively aggregate models trained locally using data stored in client devices, while avoiding misinformation.

3. Research monitoring on EU projects

A search on the European Community CORDIS (Community Research and Development Information Service) Platform[2] was performed, with regard to the most promising research projects in the field of intelligence and security. To this purpose, the search performed during the first and second runs of Task T5.2 was repeated and enlarged, to include newly funded actions.

Following the indication of the deliverable D5.1 "*Methodology for Innovation Monitoring*", the activities of the task T5.2 have been focused on a further survey of the most promising emerging technologies in the field of intelligence and security by highlighting the available results from EU research projects. To this purpose, the actual report is a tentative to rationalise and categorise knowledge exploited from the CORDIS database.

The keyword-based retrieval of data from CORDIS and the desk research (research, evaluation and possible re-elaboration of information already collected by others, typically in textual format) were adopted as analysis techniques.

The dataset retrieval was obtained by searching on the CORDIS database the research projects mentioning keywords relevant to the currently active NOTIONES working groups:

- Cryptocurrency tracking and investigation;
- Disinformation/Misinformation/Propaganda tracking, detection, countering, policing.

The research was performed by Ms. Livia di Bernardini and by Mr. Claudio Testani (Orcid ID: 0000-0002-5312-6016, Hi=13) of Consortium partner APRE, with contributors from Consortium partner LAUREA.

Ms. Di Bernardini holds a Master of Arts in International Relations (Università degli Studi Roma Tre) with focus on the Common Security and Defence Policy (CSDP) of the EU and a second level Master's degree in Cybersecurity, Public policy, regulation and management (Luiss Guido Carli University). She is currently employed at APRE where she carries out several projects in the framework of the Horizon Europe cluster 4 "Digital, Industry and Space".

Mr. Testani holds a Master's degree in Aerospace Structural Engineering (Univ. La Sapienza, Roma, Italy) and a PhD in Material Science (Univ. Tor Vergata, Roma, Italy). Moreover, he holds the Italian ASN (qualification for Associate Professor) and he is member of the teaching board of the TorVergata University PhD School. He is member of the European Enterprise Network sector group for Aeronautic, Defence and Aerospace and is member of the APRE - Cluster 4 (Industry, Digital and Space) Expert Team for Horizon Europe.

In the next subsections, the results of the research project monitoring activities performed in the third run of WP5 are presented.

3.1 Cryptocurrency tracking and investigation

Five (5) projects were found to have interesting results in terms of relevance with the NOTIONES focus area *"Cryptocurrency tracking and investigation"*. The projects are listed in Table 1.

[2] https://cordis.europa.eu

Acronym	Title	Start - end year	Mapped at
CryptoVolatility	New, realistic and robust models for cryptocurrency volatility	2022 - 2024	3rd round
TRACE	Tracking illicit money flows	2021 - 2024	1st round 3rd round
ANITA	Advanced tools for fighting oNline Illegal TrAfficking	2018 - 2021	1st round 2nd round 3rd round
TITANIUM	Tools for the Investigation of Transactions in Underground Markets	2017 - 2020	1st round 2nd round 3rd round
RAMSES	Internet Forensic platform for tracking the money flow of financially-motivated malware	2016 - 2019	3rd round

Table 1 Selected projects for "Cryptocurrency tracking and investigation"

Below is reported a synthesis of main information about the projects.

CryptoVolatility [ongoing]

Project	CryptoVolatility
Full Title	New, realistic and robust models for cryptocurrency volatility
GRANT AGREEMENT ID:	101022759
Source of information	CORDIS
EU contribution	€ 190 680,96
Coordinator	TAMPEREEN KORKEAKOULUSAATIO SR, Finland
Website:	-
Coordinator Contact:	https://www.tuni.fi/en/about-us/contact-us
Funding Scheme	EXCELLENT SCIENCE - Marie Skłodowska-Curie Actions
Start Date	1 September 2022
End Date	31 August 2024

Forecasting cryptocurrency volatility has become an increasingly important topic in quantitative finance. Many studies have shown that cryptocurrency prices are driven to a large extent by sentiments, more so than equity prices. However, existing econometric studies have been limited by their focus on using conditional volatility models that were developed for equity or commodity price volatility to fit cryptocurrency data, despite the lack of robustness and effectiveness of these models in this context.

To address these shortcomings, this project proposes the development of new and more realistic conditional volatility models that are specifically designed for cryptocurrency data. The project will also provide cross-disciplinary estimation techniques that are more reliable in forecasting cryptocurrency price volatility.

One innovative aspect of this proposed framework is the use of artificial neural networks to measure sentiments, combined with empirical observations of cryptocurrency prices. The aim is to build a machine that can produce discrete sentiment phases each day by analyzing news articles and internet search data. This will allow for the identification of the number of phases and the determination of which phase an observation at a given time belongs to.

Once the sentiment phases have been identified, the project will use particle filtering techniques to estimate model parameters and filter out the continuous conditional volatility process in real-time. This approach is expected to yield more accurate and reliable results compared to existing models.

The project is not only relevant for academics, but also for regulators and investors. Regulators can use the sentiment labels from the neural network to design policies that can help prevent and mitigate financial crises in the future. Investors can benefit from a better understanding of how cryptocurrency volatility behaves, allowing them to make more informed investment decisions.

Overall, the project aims to propose new and more realistic models for forecasting cryptocurrency volatility, accompanied by reliable estimation techniques. The use of artificial neural networks to measure sentiments is a novel and innovative approach that is expected to improve the accuracy and effectiveness of cryptocurrency volatility forecasting.

TRACE [ongoing]

Project	TRACE
Full Title	Tracking illicit money flows
GRANT AGREEMENT ID:	101022004
Source of information	https://cordis.europa.eu/project/id/101022004
Call for Proposal	H2020-SU-SEC-2018-2019-2020
EU contribution	€ 6 980 082,50
Coordinator	COVENTRY UNIVERSITY
Website:	https://trace-illicit-money-flows.eu/
Coordinator Contact:	https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/contact-form/PROJECT/999612161/101022004
Funding Scheme	H2020
Start Date	01/07/2021
End Date	30/06/2024

TRACE was already listed among the relevant projects for NOTIONES in deliverable D5.2. No updates were found for this project. NOTIONES contacted the Project coordinator and an interaction was successfully established. TRACE project will attend NOTIONES Second Conference, that will be held on the 12th of May 2023 in Paris and deliver a presentation of their main findings.

ANITA [ended in 2021]

Project	ANITA
Full Title	Advanced tools for fighting oNline Illegal TrAfficking
GRANT AGREEMENT ID:	787061
Source of information	CORDIS

EU contribution	€ 4 999 580
Coordinator	ENGINEERING - INGEGNERIA INFORMATICA SPA
Website:	https://www.anita-project.eu/
Coordinator Contact:	https://www.eng.it/find-us/offices-contacts
Funding Scheme	Secure societies - Protecting freedom and security of Europe and its citizens
Start Date	1 May 2018
End Date	31 October 2021

The main goal of ANITA project was to enhance the investigation capabilities of Law Enforcement Agencies (LEAs) by providing them with a range of tools and techniques to combat online trafficking of counterfeit/falsified medicines, NPS, drugs, and weapons. The goal was reached through various methods such as knowledge modeling and reasoning services, monitoring of online marketplaces, identification of criminal identities on the web, unmasking of fake information, discovery of criminal groups involved in trafficking, analysis of trends and behavioral patterns, tracking of crypto-currency transactions, and interoperability with existing LEA systems. These tools and techniques now allow LEAs to conduct more effective and efficient investigations using information obtained through lawful warrant.

In order to counter illegal trafficking of drugs, counterfeit medicines, NPS a firearm, ANITA designed and developed a user-centred investigation system analysing heterogeneous online and offline content. ANITA combined advanced technologies such as data source analysis of crypto-currency network, transactions and blockchain tools, Big Data analytics, and sophisticated knowledge modeling methodologies. The consortium of ANITA also developed an adaptive, cognitive user modeling framework to incorporate human perception and cognition principles in the system and enable domain-related and user-oriented intelligence applications for identifying patterns and correlations among illegal trafficking events.

More information on ANITA results and approach were already described in deliverable D5.3 “*Monitoring of EU Research and Horizon Scanning -v2*”.

Results

1	Cybercrime threat intelligence: A systematic multi-vocal literature review	Report
2	LEAs cooperation policy	Report

TITANIUM [ended in 2020]

Project	TITANIUM
Full Title	Tools for the Investigation of Transactions in Underground Markets
GRANT AGREEMENT ID:	740558
Source of information	https://cordis.europa.eu/project/id/740558
Call for Proposal	H2020-SEC-2016-2017 Secure societies - Protecting freedom and security of Europe and its citizens
EU contribution	€ 4 991 600
Coordinator	AIT AUSTRIAN INSTITUTE OF TECHNOLOGY GMBH, Austria
Website:	Startseite - AIT Austrian Institute Of Technology
Coordinator Contact:	Contact the Organisation
Funding Scheme	H2020 - Secure societies - Protecting freedom and security of Europe and its citizens
Start Date	01/05/2017
End Date	30/04/2020

TITANIUM was already listed among the relevant projects for NOTIONES in deliverables D5.2 and D5.3. No updates were found for this project.

Results

	Relevant Publications		via OpenAIRE :
Datasets	Spams meet Cryptocurrencies: Sextortion in the Bitcoin Ecosystem Author(s): Paquet-Clouston Masarah; Romiti Matteo; Haslhofer Bernhard; Charvat Tomas Published in: Zenodo		OpenAIRE
	Ransomware Payments in the Bitcoin Ecosystem Author(s): Paquet-Clouston, Masarah; Haslhofer, Bernhard; Dupont, Benoit Published in: Zenodo		
Software	Source Code for An Empirical Analysis of Anonymity in Zcash	Author(s): Kappos, Georgios; Haaroon Yousaf; Maller, Mary; Meiklejohn, Sarah DOI: 10.5281/zenodo.1443274; 10.5281/zenodo.1443273 Publisher: Zenodo	OpenAIRE

RAMSES [ended in 2019]

Project	RAMSES
Full Title	Internet Forensic platform for tracking the money flow of financially-motivated malware
GRANT AGREEMENT ID:	700326
Source of information	https://cordis.europa.eu/project/id/700326
Call for Proposal	H2020-FCT-2014-2015
EU contribution	€ 3 532 000
Coordinator	Politecnico di Milano
Website:	https://ramses2020.eu/
Coordinator Contact:	https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/contact-form/PROJECT/999879881/700326
Funding Scheme	H2020
Start Date	01/03/2016
End Date	30/11/2019

The RAMSES project ended on 2019. The project designed tools for deployment and integration into the RAMSES platform of banking Trojan analyzer and Bitcoin tracker. The Banking Trojan analyzer is based on a memory forensics framework for banking Trojan and ransomware analysis and detection. While the Bitcoin Tracker is based on a modular framework for extracting intelligence from the Bitcoin network to analyze the malicious use of this cryptocurrency.

RAMSES was an innovative project that provided Law Enforcement Agencies (LEAs) with a comprehensive, intelligent, and scalable platform to aid digital forensic investigations. This platform can extract, analyze, link, and interpret information obtained from the Internet regarding financially motivated malware.

To achieve its objectives, the project involved in the research customers, developers, and malware victims to gain a better understanding of how and where malware could be spread. The obtained information helped to pinpoint the source of the threat. RAMSES relied on disruptive Big Data technologies to extract, store, and analyze enormous amounts of structured and unstructured data. The project focused on two case studies: ransomware and banking Trojans.

RAMSES utilized the latest technologies to develop an intelligent software platform that includes public and deep web scraping, manipulation and steganalysis for images and videos, tracking malware payments, and analysis of malware samples. The platform also incorporates Big Data analysis and visualization tools.

Validation pilots for RAMSES were conducted in three different EU countries - Portugal, Belgium, and Spain. The first pilot was a mono-LEA pilot at each site, while the second was a collaborative investigation pilot involving several LEAs.

RAMSES is a crucial project that provided LEAs with the tools they need to combat cybercriminals effectively. By leveraging Big Data technologies and the latest software, RAMSES truly is a valuable asset in the fight against financially motivated malware.

Results

1	Overview of existing approaches and best practices for digital surveillance by Law Enforcement Agencies	Report
2	Report on the use and prevalence of image and video steganography over Social Media	Report
3	Stakeholders identification and liaison activities	Document
4	Training material and plan of training sessions for law enforcement agents	Report
5	Social and ethical implications of digital surveillance	Briefing paper
6	Final report on other cryptocurrencies	Report
7	Ethical and privacy monitoring and evaluation of RAMSES platform	Report
8	Design of the analysis system and specifications	Report

3.2 Disinformation/Misinformation/Propaganda tracking, detection, countering, policing

Fifteen (15) projects were found to have interesting results in terms of relevance with the NOTIONES focus area *"Disinformation/Misinformation/Propaganda tracking, detection, countering, policing"*. The projects are listed in Table 2 Selected projects for "Disinformation/Misinformation/Propaganda".

Table 2 Selected projects for "Disinformation/Misinformation/Propaganda"

Acronym	Title	Start-end year	Mapped at
AI4TRUST	AI-based-technologies for trustworthy solutions against disinformation	2023 - 2026	3rd round
TITAN	AI for Citizen Intelligent Coaching against Disinformation	2022 – 2025	3rd round
vera.ai	VERification Assisted by Artificial Intelligence	2022 – 2025	3rd round
FERMI	Fake nEws Risk MItigator	2022 – 2025	3rd round
VIGILANT	Vital IntelliGence to Investigate ILlegAl DisiNformaTion	2022 – 2025	3rd round
DisAI	Improving scientific excellence and creativity in combating disinformation with artificial intelligence and language technologies	2022 – 2025	3rd round
FARE_AUDIT	Fake News Recommendations - an Auditing System of Differential Tracking and Search Engine Results	2022 – 2024	3rd round
RADICALISATION	We're not neo-Nazis anymore': Radicalisation strategies in online far-right propaganda and disinformation campaigns	2020 – 2023	3rd round
MISTRUST	Correcting misinformation: The role of source (un)trustworthiness on the effects of repetition and contradiction in judgments of information's truth-value.	2020 – 2022	3rd round
PROVENANCE	Providing Verification Assistance for New Content	2018 – 2022	3rd round
FANDANGO	FAke News discovery and propagation from big Data ANalysis and artificial intelliGence Operations	2018 – 2021	3rd round
SOMA	Social Observatory for Disinformation and Social Media Analysis	2018 – 2021	3rd round

WeVerify	Wider and enhanced verification for you	2018 – 2021	3rd round
Co-Inform	Co-Creating Misinformation-Resilient Societies	2018 – 2021	3rd round
SocialTruth	Open Distributed Digital Content Verification for Hyper-connected Sociality	2018 – 2021	3rd round

AI4TRUST [ongoing]

Project	AI4TRUST
Full Title	AI-based-technologies for trustworthy solutions against disinformation
GRANT AGREEMENT ID:	101070190
Source of information	CORDIS
EU contribution	€ 5 950 682,50
Coordinator	Fondazione Bruno Kessler, Italy
Website:	Not yet available
Coordinator Contact:	https://www.fbk.eu/it/contatti/
Funding Scheme	RIA - Research and Innovation action
Start Date	1 January 2023
End Date	28 February 2026

The AI4TRUST system will use advanced AI solutions to monitor multiple social media platforms and filter out social noise. It will analyze multimodal content (text, audio, visual) in multiple languages with novel AI algorithms, providing nearly real-time updates. Additionally, the system will cooperate with an international network of human fact-checkers who will periodically provide validated data to update the algorithms.

The resulting quantitative indicators, including infodemic risk, will be inspected through the lens of social and computational social sciences. This will help creating customizable and reliable data reports, providing media professionals with trustworthy information. The AI4TRUST system is based on a human-centered approach to technology development, aligned with European social and ethical values, and is expected to become a standard tool for data analysts working on disinformation.

TITAN [ongoing]

Project	TITAN
Full Title	AI for Citizen Intelligent Coaching against Disinformation
GRANT AGREEMENT ID:	101070658
Source of information	CORDIS
EU contribution	€ 5 734 395
Coordinator	Engineering - Ingegneria Informatica Spa, Italy
Website:	Not yet available
Coordinator Contact:	https://www.eng.it/find-us/offices-contacts
Funding Scheme	RIA - Research and Innovation action
Start Date	1 September 2022
End Date	31 August 2025

The TITAN project aims to empower citizens to question, investigate, and understand whether a statement is true. To achieve this, TITAN will develop an engaging ecosystem that is open, distributed and citizen-centric.

Through intelligent coaching, TITAN will introduce AI-driven, intuitive, and personalised 'question-and-response' interaction systems that will enable citizens to effectively and efficiently investigate statements to determine their veracity. The interaction will focus the attention of the investigating citizen on the logical interpretation and critical assessment of the implied reasoning and arguments in the statement, while guiding the citizen to appropriate fact-checking and media literacy tools and services.

The TITAN ecosystem will empower citizens to conduct investigations on their own or in collaboration with other concerned citizens. The intelligent coaching conversational schemes will be personalised based on the investigating citizen's profile, digital skills, media literacy skills, and possible difficulties in critical thinking, as well as the linguistic characteristics of the statement under investigation. At the end of each interaction cycle, the citizen will have improved their critical thinking and media literacy skills, making them better equipped to detect disinformation in the future.

The TITAN project will be developed using a human-centred approach, which will involve diverse groups of citizens in co-creation sessions at all phases of implementation. By engaging citizens in the development of the ecosystem, TITAN hopes to create a more effective and accessible tool for combatting disinformation. Through this citizen-centric approach, the TITAN project aims to create a more informed and critical-thinking society, better equipped to address the challenges posed by disinformation.

vera.ai [ongoing]

Project	vera.ai
Full Title	VERification Assisted by Artificial Intelligence
GRANT AGREEMENT ID:	101070093
Source of information	CORDIS
EU contribution	€ 5 691 875
Coordinator	ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS, Greece
Website:	https://www.veraai.eu/home
Coordinator Contact:	https://www.certh.gr/8E7D7A5A.en.aspx
Funding Scheme	HORIZON.2.4 - Digital, Industry and Space
Start Date	15 September 2022
End Date	14 September 2025

The "vera.ai" project aim to create trustworthy AI solutions that will include a fact-checker-in-the-loop approach and AI models that constantly check updated sources and multimodal data, verified in the InVID-WeVerify plugin and the Truly Media/EDMO platform.

"vera.ai" aims to create trustworthy AI solutions that can counter advanced disinformation techniques. These innovative solutions will be co-created with media professionals and researchers, and will lay the groundwork for future research in AI-based disinformation detection.

The AI models developed by "vera.ai" will have several novel characteristics, including fairness, transparency, and the ability to handle multimodal and multilingual content. They will also be continuously updated to adapt to evolving disinformation campaigns, using a fact-checker-in-the-loop approach to ensure accuracy and reliability. Additionally, "vera.ai" will develop tools for deepfake detection to counter the perils of AI-generated content.

"vera.ai" will employ a multidisciplinary co-creation approach, combining open-source algorithms with continuously collected fact-checking data from real-life content verified in the InVID-WeVerify plugin and the Truly Media/EDMO platform. Social media and web content will be analyzed and contextualized to expose disinformation campaigns and measure their impact.

The results of "vera.ai" will be validated by professional journalists and fact checkers from project partners, external participants, the community of InVID-WeVerify verification plugin users, and media literacy, human rights, and emergency response organizations. By developing trustworthy AI solutions, "vera.ai" aims to combat the spread of disinformation and protect the integrity of online information.

FERMI [ongoing]

Project	FERMI
Full Title	Fake nEws Risk MItigator
GRANT AGREEMENT ID:	101073980
Source of information	CORDIS
EU contribution	€ 3 999 815
Coordinator	Hochschule Fur Den Offentlichen Dienst In Bayern, Germany
Website:	https://fighting-fake-news.eu/
Coordinator Contact:	https://www.fhvr.bayern.de/de/aktuelles/service/kontakt.html
Funding Scheme	IA - Innovation action (HORIZON.2.3 - Civil Security for Society)
Start Date	1 October 2022
End Date	30 September 2025

The EU-funded FERMI project aims to address the issue of disinformation by applying a comprehensive and cross-disciplinary methodology to analyse disinformation and fake news, as well as their sources.

The FERMI project will leverage a holistic approach to investigate the spread of disinformation and fake news, taking into account all the socioeconomic factors that may affect their propagation and impact on multiple dimensions of society. The project will provide a set of innovative technological developments that enable the detection and monitoring of the spread of disinformation and fake news, as well as the implementation of relevant security countermeasures. Additionally, FERMI will produce tailor-made training material that will help European Police Authorities, relevant stakeholders, and EU citizens combat the spread and limit the impact of disinformation and fake news, while increasing digital trust.

The impact of disinformation and fake news can have severe consequences on society. Their diffusion online may cause uncertainty and fear, intensify crisis situations, and weaken European societies, thus aggravating their divisions. The spread of disinformation and fake news can polarize public debates and lead to physical violence offline and other hate crimes. Furthermore, the use of fake accounts, AI-generated fake content, and bots that can spread disinformation and fake news at scale pose additional problems.

Therefore, FERMI's holistic and cross-disciplinary methodology is crucial for providing a comprehensive understanding of disinformation and fake news and their sources, and for developing innovative technological developments and tailor-made training materials to mitigate their negative impact on European societies.

VIGILANT [ongoing]

Project	VIGILANT
Full Title	Vital IntelliGence to Investigate ILlegAl DisiNformaTion
GRANT AGREEMENT ID:	101073921
Source of information	CORDIS
EU contribution	€ 3 376 604,50
Coordinator	The Provost, Fellows, Foundation Scholars & The Other Members Of Board, Of The College Of The Holy & Undivided Trinity Of Queen Elizabeth Near Dublin – Ireland
Website:	https://www.vigilantproject.eu/
Coordinator Contact:	tcd.ie/contacts/
Funding Scheme	HORIZON-AG - HORIZON Action Grant Budget-Based
Start Date	1 November 2022
End Date	31 October 2025

The spread of disinformation and hateful content online has become a major concern for law enforcement agencies, but the tools and technologies available to them are often inadequate. The EU-funded VIGILANT project aims to address this issue by developing a platform that can track and analyze disinformation to help police crackdown on internet hate crime. This new platform will feature advanced disinformation identification and analysis tools and technologies, including state-of-the-art artificial intelligence methods tailored to police needs. It can be used on all major social media platforms and websites, and for all types of content, including text, image, and video, and in multiple languages.

Many police authorities do not have access to specialized tools or technologies to combat disinformation, and some of the better-equipped agencies are using off-the-shelf products designed for commercial use, which are not capable of dealing with the complexities of disinformation. The VIGILANT project solves this problem by providing an integrated platform of advanced disinformation identification and analysis tools and technologies that were developed as part of highly successful FP7 and H2020 projects. These tools will be tailored to meet the specific needs of law enforcement agencies, with an ethical-by-design and user-centric approach. The VIGILANT project also covers disinformation from all major sources, including major social media platforms and fake news websites, and it is suitable for investigating hate speech, violent nationalist or separatist movements, radicalization, extremist groups, incels, lone wolves, and other counter-terrorism threats.

The VIGILANT project also includes innovative solutions to leverage the knowledge and experience of other stakeholder organizations and social media companies. Additionally, the project includes training for law enforcement agencies in the use of VIGILANT and in conducting disinformation investigations as part of a long-term sustainable training network. The interdisciplinary consortium responsible for the project includes expertise from social sciences and humanities, ethics, computer science, and commercial entities, as well as five European police authorities. Overall, the VIGILANT project is a promising development that could help law enforcement agencies combat the spread of disinformation and hateful content online.

DisAI [ongoing]

Project	DisAI
Full Title	Improving scientific excellence and creativity in combating disinformation with artificial intelligence and language technologies
GRANT AGREEMENT ID:	101079164
Source of information	CORDIS
EU contribution	€ 1 499 750
Coordinator	Kempelenov Institut Inteligentnychtechnologii - Slovakia
Website:	-
Coordinator Contact:	https://kinit.sk/contact/
Funding Scheme	CSA - Coordination and support action
Start Date	1 December 2022
End Date	30 November 2025

The DisAI project aims to tackle the growing threat of disinformation on social media by using computer detection through natural language processing. The project focuses on ensuring that such AI processing is available for low-resource languages like Slovak, and the Kempelen Institute of Intelligent Technologies is working with the project to achieve this goal. The primary objective of the project is to enhance scientific excellence in trustworthy AI and multimodal natural language processing and multilingual language technologies to combat disinformation, with the ultimate aim of protecting European democratic values.

The DisAI project is of great importance, given the increasing amount of disinformation being spread, and the crucial role of AI and language technologies in detecting it. To meet the European goals, it is necessary to develop tools and methods for combating disinformation in low resource languages as well. KInIT, the project's Slovakian partner, is playing a leadership role in shaping the research and innovation landscape of Slovakia, by boosting cross-sectoral, interdisciplinary, and international collaboration.

The project will increase the research capacity and excellence of scientists at different career levels in multilingual and multimodal language technologies and trustworthy AI. The project partners, including the German Research Center for Artificial Intelligence, Centre for Research and Technology Hellas, and Copenhagen University, will collaborate on a joint research project on claim matching, as well as networking and mobility activities. The transfer of knowledge on innovation and creativity for facilitating industry-academia collaboration and boosting the skills of research managers and administrative staff will be part of the capacity-building program. The project will provide new opportunities for collaboration and exchanges, and contribute to growing awareness of online disinformation.

FARE_AUDIT [ongoing]

Project	FARE_AUDIT
Full Title	Fake News Recommendations - an Auditing System of Differential Tracking and Search Engine Results
GRANT AGREEMENT ID:	101100653

Source of information	CORDIS
EU contribution	€ 150 000
Coordinator	Laboratorio De Instrumentacao E Fisica Experimental De Particulas Lip, Portugal
Website:	-
Coordinator Contact:	https://ciencias.ulisboa.pt/pt/lip
Funding Scheme	ERC-POC - Proof of Concept Grant
Start Date	1 December 2022
End Date	31 May 2024

While efforts have been made to identify and remove "fake news" websites and minimize disinformation spread on social media, little attention has been given to the role of search engines. The FARE_AUDIT project aims to address this gap by offering an innovative tool to audit search engines that can be widely used.

The FARE_AUDIT tool will help to better understand how browsing history influences search engine results, particularly in directing users to disinformation. It will also create a system that democracy-promoting institutions and concerned citizens can use to identify new disinformation in near-real-time. Additionally, the tool aims to breach information bubbles by simulating how search results would differ if users had a different online profile. By relying on web-crawlers, the tool is privacy-protecting and does not require any real user data. Furthermore, the proposed system anticipates the shift from cookie-tracking to fingerprinting and takes advantage of the expected small time overlap between both systems to broaden its scope.

The FARE_AUDIT project is expected to have a meaningful social impact by increasing public awareness of the role of search engines in disinformation spread and equipping organizations with a tool to detect and monitor disinformation, especially in political contexts. Overall, this novel tool could represent a significant step towards curbing the spread of disinformation and promoting the transparency and accountability of search engines.

RADICALISATION [ongoing]

Project	RADICALISATION
Full Title	We’re not neo-Nazis anymore’: Radicalisation strategies in online far-right propaganda and disinformation campaigns
GRANT AGREEMENT ID:	845643
Source of information	CORDIS
EU contribution	€ 235 191,36
Coordinator	CEU GMBH – Austria
Website:	https://www.ceu.edu/project/radicalisation
Coordinator Contact:	https://www.ceu.edu/contact
Funding Scheme	MSCA-IF-GF - Global Fellowships
Start Date	15 April 2020
End Date	14 April 2023

The EU-funded RADICALISATION project intends to identify the radicalisation strategies employed by far-right groups in their online propaganda. By gaining a deep understanding of the linguistic, semiotic, and visual resources used by these groups to recruit new members, the project seeks to contribute to the development of effective prevention and deradicalisation programs.

The urgency of this project is evident from the recent rise in far-right violence, which is often underreported by the media and law enforcement agencies. The internet has facilitated the rapid dissemination of far-right ideology and the mobilisation of individuals from white supremacist movements. This poses an enhanced security threat to the EU and it is considered an existential threat to European values. The project's primary goal is to identify the radicalisation strategies employed by far-right groups and to gain a deep understanding of their linguistic and semiotic repertoire.

The project's interdisciplinary approach is designed to generate new knowledge relevant to a range of disciplines. By investigating extremist online propaganda, the project aims to contribute to the development of joint programs on countering violent extremism and radicalisation, as set out in the European Commission's 'A Global Strategy for the European Union's Foreign and Security Policy'. Ultimately, the project seeks to lay the foundations for the creation of educational and deradicalisation programs that can effectively counter the spread of far-right extremist ideology and protect fundamental European values.

MISTRUST [ended in 2022]

Project	MISTRUST
Full Title	Correcting misinformation: The role of source (un)trustworthiness on the effects of repetition and contradiction in judgments of information's truth-value.
GRANT AGREEMENT ID:	844296
Source of information	CORDIS
EU contribution	€ 147 815,04
Coordinator	Iscte - Instituto Universitário de Lisboa - Portugal
Website:	-
Coordinator Contact:	https://www.iscte-iul.pt/contents/1380/contact-us
Funding Scheme	MSCA-IF-EF-ST - Standard EF
Start Date	1 December 2020
End Date	30 November 2022

The EU-funded MISTRUST project examined the correction mechanism for fake news and proposed methods to convince people to question false information.

The project's focus was on the untrustworthiness of the source of the false claims. The researchers hypothesized that providing information about the source's lack of credibility would prompt individuals to scrutinize and analyze information more deeply, countering the effects of repetition and contradiction. The results of the project contributed to the development of effective misinformation-correction actions as well as informed future policies to deal with the increasing amount of fake news and misinformation spread to the public.

PROVENANCE [ended in 2022]

Project	PROVENANCE
Full Title	Providing Verification Assistance for New Content
GRANT AGREEMENT ID:	825227
Source of information	CORDIS
EU contribution	€ 2 438 810
Coordinator	DUBLIN CITY UNIVERSITY, Ireland
Website:	http://www.provenanceh2020.eu/
Coordinator Contact:	https://www.dcu.ie/contact-us
Funding Scheme	IA - Innovation action
Start Date	1 December 2018
End Date	31 May 2022

The project PROVENANCE was designed to help people navigate through online content and develop digital literacy skills by creating a personalized virtual companion that would evaluate the quality of online content and provide contextual information to users.

The PROVENANCE Social Network Monitor used advanced tools to verify multimedia content and contextualize it with relevant information such as the quality of writing and visual manipulation. The project developed an intermediary-free solution for digital content verification using blockchain technology and to empower users with greater control over social sharing based on values of trust, openness, and fair participation.

The project was co-created with representatives of civil society across four distinct use-cases in the social media domain and aimed to significantly advance the state of the art in content verification, understanding of information cascades, openness of algorithms, and user control over personal data. Ultimately, the project laid the foundation for a new federated social network grounded in trust, openness, and fair participation and support the development of an observatory on information veracity and social media best practice.

Results:

1	Disinformation and Manipulation in Digital Media - Information Pathologies	Monographic books
2	Report on Use Cases and User Requirements Formalisation	Report
3	Technical Evaluation Support Tool	Report
4	Innovative Tools for Citizen Empowerment in the Fight Against Misinformation	Book chapter

FANDANGO [ended in 2021]

Project	FANDANGO
Full Title	FAke News discovery and propagation from big Data ANalysis and artificial intelliGence Operations
GRANT AGREEMENT ID:	780355
Source of information	Cordis

EU contribution	€ 2 879 250
Coordinator	ENGINEERING - INGEGNERIA INFORMATICA SPA, Italy
Website:	https://fandango-project.eu/
Coordinator Contact:	https://www.eng.it/find-us/offices-contacts
Funding Scheme	IA - Innovation action
Start Date	1 January 2018
End Date	31 March 2021

The FANDANGO project used cross-sector big data management and analytics, along with an effective interoperability scheme, to generate new business and societal impacts involving media companies, governmental institutions, the overall industrial ecosystem, and society as a whole.

The project aggregated and verified different typologies of news data, media sources, social media, and open data in order to detect fake news and provide an efficient and verified communication for all European citizens. The European tradition in democracy, journalism, and transparency should play a worldwide example in the fast-changing society, where all citizens are overwhelmed by new technologies and social challenges.

The idea of FANDANGO was to break data interoperability barriers by providing unified techniques and an integrated big data platform to support traditional media industries to face the new "data" news economy with better transparency to the citizens under a Responsible, Research and Innovation prism. The project was validated and tested in three specific domains, namely climate, immigration, and the European context, which are typical scenarios where fake news can influence perception with respect to social and business actions, and where news can be verified and validated by trustable information based on facts and data.

Results:

1	Dataset related to article "Volume-of-Interest Aware Deep Neural Networks for Rapid Chest CT-Based COVID-19 Patient Risk Assessment" **Author(s):** Chatzitofis, Anargyros; Cancian, Pierandrea; Gkitsas, Vasileios; Carlucci, Alessandro; Stalidis, Panagiotis; Albanis, Georgios; Karakottas, Antonis; Semertzidis, Theodoros; Daras, Petros; Giannitto, Caterina; Casiraghi, Elena; Sposta, Federica Mrakic; Vatteroni, Giulia; Ammirabile, Angela; Lofino, Ludovica; Ragucci, Pasquala; Laino, Maria Elena; Voza, Antonio; Desai, Antonio; Cecconi, Maurizio; Balzarini, Luca; Chiti, Arturo; Zarpalas, Dimitrios; Savevski, Victor **Published in:** Zenodo	Dataset
2	Second iteration piloting and validation report	Report
3	User Requirements	
4	Data Interoperability and data model design	
5	FANDANGO Reference Architecture description	
6	Impact Report	
7	First iteration piloting and validation report	
8	Data model and components	
9	Data lake integration plan	
10	FANDANGO platform setup defining process	
11	Pilots execution and evaluation plans	
12	Final Exploitation plan and technology uptake	
13	Report replicability of the solution	
14	Market Analysis and preliminary business requirement	

16	Application areas business requirements and preliminary exploitation plan	

SOMA [ended in 2021]

Project	SOMA
Full Title	Social Observatory for Disinformation and Social Media Analysis
GRANT AGREEMENT ID:	825469
Source of information	https://cordis.europa.eu/project/id/825469
EU contribution	€ 987 437,50
Coordinator	Athens Technology Center Anonymi Viomichaniki Emporiki Kai Techniki Etaireia Efarmogon Ypsilis Technologias - Greece
Website:	https://www.disinfobservatory.org/
Coordinator Contact:	https://www.atc.gr/contact-us/
Funding Scheme	Horizon2020 CSA - Coordination and support action
Start Date	1 November 2018
End Date	30 April 2021

The Horizon 2020 funded SOMA project (Social Observatory for Disinformation and Social Media Analysis) was launched to demonstrate how disinformation spreads on social media, and to identify new tools and methods for detecting, analysing, and countering it.

The project involved collaboration between researchers from various disciplines across Europe, and included partners from academic institutions, research centres, and media organisations. Some of the specific research areas that SOMA focused on included:

- the use of bots and automated accounts to spread disinformation,
- the role of social media algorithms in promoting disinformation, and
- the impact of disinformation on democratic processes.

Disinformation can take many forms, such as fake news stories, manipulated images or videos, rumours, conspiracy theories, and propaganda. The spread of disinformation can be deliberate and organised, or it can occur unintentionally because of misinformation or mistakes. EC has defined disinformation as verifiably false or misleading information that is created, presented, and disseminated for economic gain or to intentionally deceive the public, and may cause public harm. Public harm comprises threats to democratic political and policy-making processes as well as public goods such as the protection of EU citizens' health, the environment or security [37].

SOMA launched a multitude of services to provide support to a European community that will jointly fight disinformation, especially aiming at restoring trust to social media.

A crucial starting point was to increase awareness among the key stakeholders about social media trends and topics of concern, while providing tools to tackle disinformation. Thus, to start the work, SOMA formed a community of more than 100 organisations from over 20 countries enabling effective collaboration between those fighting disinformation. The community is called “European Observatory against Disinformation” and it has its own webpage to promote the community and its members’ work (https://www.disinfobservatory.org/the-observatory/).

The collaborative verification platform – based on Truly Media, developed by ATC and Deutsche Welle – also links to other verification tools such as TruthNest [38]. Users can easily approach sources such as the European Parliament, European Commission and Eurostat to request official positions on specific issues.

SOMA also researched existing methodologies to assess news sources, identifying 12 indicators to create the SOMA Transparency Index [39]. These indicators were grouped under six dimensions related to the transparency and trustworthiness of news outlets: headline, author, sources, contents, wording, and advertisement.

The impact on technological tools of SOMA were related to producing three new technological tools and one algorithm (see the table below). The tools have been directly integrated into the Truly Media platform [40], while the algorithm has been published in a scientific journal and in Bitbucket, an Open-Source repository.

Results:

#	Title of the Deliverable	Short Description
1	Evaluating Distributed scalable information cascade analysis. 4th section.	Section of report that details the models used for (dis)information cascade modelling at scale and evaluate the potential to scale from mono platform to multi- and interoperable platform and media information flow analysis.
2	Evaluating safe space solutions including data management and processing setups 1st section	Section of report that details the models used for (dis)information cascade modelling at scale and evaluate the potential to scale from mono platform to multi- and interoperable platform and media information flow analysis.
3	Final Sustainability Plan	Report that details the final strategy be adopted to guarantee the Observatory sustainability and its feasibility.
4	Report on the centers for excellence studies on online disinformation.Set up and activities.	Report that provides information about the two centres of excellence, about their set up and operational activities
5	Impact assessment results	Report that presents the results from the impact assessment and shows the major impacts produced by the Observatory at European level
6	Research Data exchange (and transparency) solution with platforms. 2nd section	Report that compiles the findings and recommended solutions and actions needed in order to construct a sustainable data exchange model for stakeholders, focusing on an differentiated perspective, one for journalists and the broader community, and one for university-based academic researchers.
7	Social media Observatory Guide	Guideline to use the Observatory
8	Media Literacy Public Activities	Report that presents the 2 public events organized in order to promote Media literacy, following the findings of the Workshops.

9	Data Intelligence toolkit description	Report containing the description of the architecture, the concept, the models and the design for the realization of the Data Intelligence toolkit
10	Media Literacy Workshop Series & Reports	Report that presents the 10 stakeholder meetings between experts at the European level.
11	Algorithms of Data Intelligence, Complex Network Analysis, Artificial Intelligence for the Observatory AI Driven	Report containing the analyses, state of the art and study related to Complex Network, Data & Geo Intelligence, Artificial Intelligence algorithms.
12	Outlier (disinformation) detection solution. 3rd section	Report that details the tools and work done on outlier detection and evaluate the transferability to disinformation and the potential actions needed in order to carry out effective detection.
13	SOMA impact assessment methodology	Report that described the methodology for the SOMA impact assessment. The methodology description contains 10 Source Transparency Indicators and variables that can be used to measure the impacts of SOMA during the project lifetime and which form the Source Transparency Index (STI).
14	The measure of online disinformation	Whitepaper reporting main evidences from the analysis performed in WP5 in order to provide measures and data on the effects of the disinformation on three topics to policy makers trying to influence decision making through the use of data collected during the project.
1	Beyond Fact-Checking: Network Analysis Tools for Monitoring Disinformation in Social Media Author(s): Guarino, S., Trino, N., Chessa, A., & Riotta, G. Published in: 2019 Publisher: Springer Cham	Publication
2	Characterizing Networks of Propaganda on Twitter: a Case Study Author(s): Guarino, S., Trino, N., Celestini, N., Chessa, A., & Riotta, G. Published in: Submitted to Applied Network Science. Springer, Cham., 2020 Publisher: Springer Cham	Publication
3	SOMA PROJECT: ESTABLISHING THE EUROPEAN OBSERVATORY AGAINST DISINFORMATION Author(s): Marina Klitsi, Simona De Rosa, Luca Tacchetti, Silvia	Publication

	Cavasola, Lynge Asbjørn Møller and Nikos Sarris Published in: 2020 Publisher: icme	

WeVerify [ended in 2021]

Project	WeVerify
Full Title	Wider And Enhanced Verification For You
GRANT AGREEMENT ID:	825297
Source of information	CORDIS
EU contribution	€ 2 499 450
Coordinator	SIRMA AI EAD - Bulgaria
Website:	https://weverify.eu/
Coordinator Contact:	https://sirma.com/
Funding Scheme	IA - Innovation action
Start Date	1 December 2018
End Date	30 November 2021

The EU-funded WeVerify project aimed at tackling the challenges of content verification, which is difficult even for experienced journalists, human rights activists, and media literacy scholars. WeVerify applied a participatory verification approach, open-source algorithms, low-overhead human-in-the-loop machine learning, and intuitive visualisation to analyze and contextualize web content within the broader online ecosystem. This exposed fabricated content and provided cross-modal content verification, social network analysis, micro-targeted debunking, and a blockchain-based public database of known fakes.

A key outcome of WeVerify was a platform for collaborative, decentralized content verification, tracking, and debunking that is open source to engage communities and citizen journalists alongside newsroom and freelance journalists. Additionally, a premium version of the platform was offered to support more advanced newsroom needs. The platform was further supplemented by a digital companion to assist with verification tasks.

The results of the WeVerify project were validated by professional journalists and debunking specialists from project partners, external participants from the First Draft News network, the community of more than 2,700 users of the InVID verification plugin, and by media literacy, human rights, and emergency response organizations.

Results:

1	Ukraine-related Disinformation Dataset **Author(s):** Iknoor Singh **Published in:** Zenodo	Datasets
2	InVID FIVR-200K **Author(s):** Kordopatis-Zilos, Giorgos; Papadopoulos, Symeon; Patras, Ioannis; Kompatsiaris, Yiannis **Published in:** Zenodo	
3	Ukraine-related Disinformation Dataset	

	Author(s): Iknoor Singh **Published in:** Zenodo	
4	WeVerify Architecture Definition	Reports
5	WeVerify Use Case Analysis Report	
6	"Blockchain Database of ""Known Fakes"" v2.0"	
7	Cross-modal verification tools - final version	Demonstrators, pilots, prototypes
8	Disinformation Flow Analysis: Tools and Methodology - Final Version	
9	User-Facing WeVerify Tools - Final Version	
10	WeVerify Open Platform and Benchmark Evaluation v3.0	
11	User-Facing WeVerify Tools v2	
12	"Blockchain Database of ""Known Fakes"" v1.0"	
13	Disinformation Flow Analysis: Tools and Methodology v.2	
14	Cross-modal verification tools - v1	

Co-Inform [ended in 2021]

Project	Co-Inform
Full Title	Co-Creating Misinformation-Resilient Societies
GRANT AGREEMENT ID:	770302
Source of information	CORDIS
EU contribution	€ 4 110 758,75
Coordinator	STOCKHOLMS UNIVERSITET - Sweden
Website:	https://coinform.eu/
Coordinator Contact:	https://www.su.se/om-universitetet/kontakt
Funding Scheme	RIA - Research and Innovation action
Start Date	1 April 2018
End Date	31 July 2021

The Co-Inform project focused on creating socio-technical solutions that empower citizens, journalists, and policymakers to increase resilience to misinformation and make more informed decisions.

Co-Inform's goal was to co-create these solutions in collaboration with citizens, journalists, and policymakers. These solutions included detecting and combating misinforming posts and articles on social media, supporting misinformation-resilient behavior, bridging communication between the public, external fact-checking journalists, and policymakers, understanding and predicting which misinforming news and content are likely to spread across which parts of the network and demographic sectors, infiltrating echo-chambers on social media to expose confirmation-biased networks to different perceptions and corrective information, and providing policymakers with advanced misinformation analysis to support their policy-making process and validation.

To achieve these objectives, Co-Inform brought together a team of multidisciplinary scientists and practitioners to develop co-creational methodologies and practices for engaging stakeholders in combating misinformation posts and news articles. The team used advanced intelligent methods for detecting misinformation, predicting misinformation flow, and real-time processing and measurement of crowds' acceptance or refusal of misinformation. The Co-Inform tools and platform were made freely available and open-sourced to maximize their benefit and reuse.

The Co-Inform project directly engaged three main stakeholder groups throughout this process: citizens, journalists, and policymakers. By empowering these groups with better tools and information, Co-Inform aimed to increase resilience to misinformation and generate more informed behaviors and policies.

SocialTruth [ended in 2021]

Project	SocialTruth
Full Title	Open Distributed Digital Content Verification for Hyper-connected Sociality
GRANT AGREEMENT ID:	825477
Source of information	CORDIS
EU contribution	€ 2 505 027
Coordinator	ICCS Institute of Communications and Computer Systems (Greece)
Website:	http://www.socialtruth.eu/
Coordinator Contact:	cdemest@cn.ntua.gr
Funding Scheme	IA – Innovation Action
Start Date	1 December 2018
End Date	30 November 2021

The goal of the project SocialTruth was to create an open, democratic, pluralistic, and distributed ecosystem that would allow easy access to various verification services (both internal and third-party), ensuring scalability and establishing trust in a completely decentralised environment.

Social networks, media, and web platforms are the standard way in how our societies operate for the purposes of communication, information exchange, conducting business, co-creation, learning and knowledge acquisition. However, this extreme growth and adoption of Social Media, in combination with their poor governance and the lack of quality control over the digital content being published and shared, has led information veracity to a continuous deterioration. SocialTruth tackled this challenge and provided an innovative and distributed way to achieve both content and author credibility verification and detection of fake news. In this way, it contributed to increase trust in and reliability of Social Media. SocialTruth solution can be used to detect fake news by both professionals (i.e. journalists) and individuals (daily social media users), allowing for improved governance and information veracity in Social Media.

In particular, SocialTruth solution presents the following features:

- Open ecosystem with standard interfaces and no vendor lock-in, favouring pluralism and the deployment of reusable, interoperable and interchangeable verification services;
- Powered by blockchain technology, for distributed reputation and trust, enhanced security and auditability, with no intermediaries and central authorities;
- Expert meta-verification engines with open design that can intelligently fuse multiple verification results;
- Lifelong Learning Machines that constantly accumulate experience and learn new paradigms of fake news;
- Digital Companion ensuring convenient access to both professional and individual users.

Results:

#	**Title of the Deliverable**	**Short Description**

1	Requirements and use cases	Report on the use cases and requirements and humanrelated aspects that should be covered by the SocialTruth solution
2	SocialTruth Semantic Analyzer	Report on the initial design of SocialTruth’s semantic analyzer system.
3	SocialTruth Blockchain	Report on the design and implementation of the blockchain architecture
4	Distributed System Architecture, Data Modelling and Interfaces	Report on the detailed, functional and non-functional, specifications for the distributed SocialTruth Architecture and the associated interfaces.
5	Refined Distributed System Architecture	Report on the detailed technical specifications for the revised distributed SocialTruth Architecture based on feedback collected
6		
7	SocialTruth LifeLong Learning Expert System	Report on the design and implementation of the SocialTruth expert system for ranking and classification together with relevant intelligent learningbased functionalities
8	SocialTruth Deep Learning Multimedia Verification	Report on the initial design of SocialTruth’s deep learning multimedia verification system.
9	SocialTruth Open Verification Ecosystem - Release 2	Report on the design and implementation of the open ecosystem building blocks including the standard SocialTruth APIs final release
10	SocialTruth Digital Companion	Report on the design and implementation of the Digital Companion
11	Implementation of the BERT-derived architectures to tackle disinformation challenges	Sebastian Kula; Rafał Kozik; Michał Choraś, Neural Computing & Applications, 2021
12	A predictive model for estimating citizens' beliefs regarding the risk perception of dissemination and dispersal of fake content	Lazar, Iuliana Mihaela, Paun; Ana Catalina, Cognition, Brain, Behavior, 2020
13	Advanced Machine Learning techniques for fake news (online disinformation) detection: A systematic mapping study	Choraś, M.; Demestichas, K.; Giełczyk, A.; Herrero, Á.; Ksieniewicz, P.; Remoundou, K.; Urda, D.; Woźniak, M., Applied Soft Computing, 2021
14	Food for Thought: Fighting Fake News and Online Disinformation	Demestichas, K.; Remoundou, K.; Adamopoulou, E., IT Professional, 2020
15	New explainability method for BERT-based model in fake news detection	Szczepanski M., Pawlicki M., Kozik R., Choras M., Scientific Reports, 2021

Many Conference proceedings are also available as results of the SocialTruth project, as well as two book sections (see CORDIS page).

4. Additional findings

This section describes the text analysis techniques that can be used to characterize and identify cases of disinformation/misinformation.

4.1 Disinformation detection technology through NLP

4.1.1 Background

Distinguishing true news from false news is a very complex task, and automated support has been researched in the past and currently, with specific algorithms designed for this purpose [41]. In its simplest form, a disinformation detector can be thought of as working on text classification problems: an AI model is trained under supervision to classify text, i.e. assign the probability of occurrence of predefined categories in each text it processes.

In fact, most existing research concerns supervised methods (AI models trained on data manually labelled by human annotators); semi-supervised and unsupervised methods are less widely used [42] [43]. Before diving into specific use cases, it is essential to emphasize the usefulness of representing text, or any other type of data, numerically as feature vectors (embedding). These numerical representations, obtained at the word, phrase or even word, phrase or document level, represent spatial (semantic) relationships with other words/sentences/documents, which an artificial intelligence model can use to group, classify or even calculate semantic similarities numerically, instead of having to search for and compare specific keywords.

Models such as Word2Vec (group of related models that are used to produce word embeddings, published in 2013) and GloVe (Global Vectors model for distributed word representation, published in 2014) can be used to transform words into embedding vectors. For sentence-level embeddings artificial intelligence models such as Universal Sentence Encoder [44] or another model customised for a specific language and use case can be used.

In the context of the detection and analysis of disinformation activities, both classical Machine Learning (ML) and Deep Learning (DL) models are widely used. However, over the past four years, neural network models based on attention in natural language processing, called Transformers [45], have demonstrated high levels of accuracy [46]. Studies show that accuracy has improved due to the development of more efficient ways of using meta-data, data and information.

More efficient ways of using metadata, such as speaker credibility and information about new model - called Grover [47] - for generating fake texts using the Generative Pre-trained Transformer 2 (GPT-2, [48]) architecture showed that the overall reliability score of misinformation increases when it is rewritten by a neural network. Disinformation increases when it is rewritten by the Grover text generator. The authors also found that Grover's neural network can effectively detect computer-generated articles. They argue that to combat AI-generated fake news, access to generators is crucial.

However, OpenAI [49] challenged the supremacy of GPT-2-generated texts by demonstrating that the tuning of a RoBERTa-based detector [50] achieved consistently higher accuracy than the tuning of a GPT-2-based detector with equivalent capabilities.

Jwa et al. [51] proposed a model for misinformation detection based on the architecture of the BERT (Bidirectional Encoder Representations from Transformers) transformer and fine-tuned on data from CNN and the Daily Mail. The relationship between the headline and the body of the news text was analysed.

The research by Marcellino et al. [52] introduced an improved model to efficiently detect conspiracy theory topics, based on a hybrid ML model that combined BERT word embeddings (numerical feature vectors) with linguistic position markers obtained from an ML analysis tool able to combine qualitative content analysis with the identification of patterns, tone and appreciation of feeling in the use of words.

Fagni et al. [53] collected a dataset of deepfake tweets - TweepFake - to evaluate 13 computer-generated (deepfake) texts. The results showed that automatically distinguishing between human-composed and computer-generated tweets is a challenge due to continuous improvements in generator performance and the limited length of the tweet. Disinformation detection exploits the stylistic distortions present in a text.

AI text generators often introduce artefacts into their texts, which can be learned and recognized by discriminators [41]. However, the datasets used to train the models are likely to be biased and this can cause errors in detection [54]. Furthermore, it is not sufficient to address disinformation as a simple matter of as a simple matter of machine-generated text identification. The most inherent characteristic of disinformation is that the 'truth' is made ambiguous by introducing false and/or misleading facts, not whether a text is human- or machine-generated.

Another approach to detecting fake news is to use a knowledge base of verified facts or articles. Ghosh et al. [55] introduce a fake news detector consisting of two sub-modules: a sub-module for detecting veracity based on information retrieval models and a sub-module based on style. The truthfulness check consists of two steps: the most relevant documents are retrieved from a carefully prepared knowledge base and, given the true (i.e. factual) information contained in these documents, the truthfulness of a statement is deduced.

Shaar et al. [56] propose a model that learns to classify relevant documents to identify previously verified statements. The BERT transformer neural network is used as a sentence encoder to obtain a numerical representation of an input text. The cosine similarity is calculated to classify the numerical representation of the input statements and the verified statements in the dataset. An interesting product of this type is called FactSparrow [57] introduced by Repustar, which uses a Twitter bot as a fact request and delivery mechanism; anyone can mention the FactSparrow bot in social media conversations and retrieve the relevant facts of the discussed topic.

Identification using a knowledge base can also benefit greatly from Named Entity Recognition (NER), a specific activity in which the artificial intelligence model extracts useful information (proper names, organisations, locations, medical codes, time expressions, quantity monetary values, percentages, etc.) from raw, unstructured textual data.

Sentiment analysis at entity level [58] (assigning numerical values to sentiment expressions and aggregating them for analysis, so that a document can be assigned a positive or negative score and a high or low sentiment value) makes it possible to analyse and compare texts at a more granular level.

The use of artificial intelligence to further extract relationships between entities makes it possible to construct advanced knowledge graphs [59] for efficient information extraction and visualisation.

Thanks to the increasing shift towards AutoML solutions [60] [61] [62], less technologically savvy AI practitioners can benefit from state-of-the-art AI models and shift the focus away from the fine-tuning of AI model parameters and the development of specific data sets.

4.1.2 Stylometric analysis

This section reports on an innovative technique used to identify cases of misinformation based on evidence in the literature [63] [64] of stylistic changes in the language of those who lie or deceive.

Starting from this premise, and thanks to the use of techniques based on automatic text comprehension and in particular computational stylometry, the section looks at concrete cases of application [65].

Stylometric Analysis aims at analyzing the key traits of a text in terms of vocabulary use and writing style which can help to identify the author, or some information related to their profile, such as gender and age. These traits are called linguistic features, whose collection and gathering conducted by statistical and computational methods can outline the *Authorial DNA*, namely the unique writing style of an author (as unique as the fingerprint) defined by the set of the characteristics of their language, which originate from psychological and sociological properties and peculiarities. Indeed, variation in psychological and sociological aspects of the author define a stylistic variation. These psychological factors include personality, mental health, and being a native speaker or not; sociological factors include age, gender, education level, and region of language acquisition [66]. Some linguistic features relevant for stylometric the analysis are the following [67]:

i. Frequency of *n*-grams (sequences of *n* letters/syllables/words),
ii. Frequency of punctuation markers, frequency of errors,
iii. Average length of sentences,
iv. Frequency of repetitions,
v. Richness of vocabulary,
vi. Terminology choices,
vii. The use of punctuation marks and so on.

Figure 13 presents an example of some stylemes extracted analyzing text of Cristoforetti, Moore and Chomsky using NLU technique.

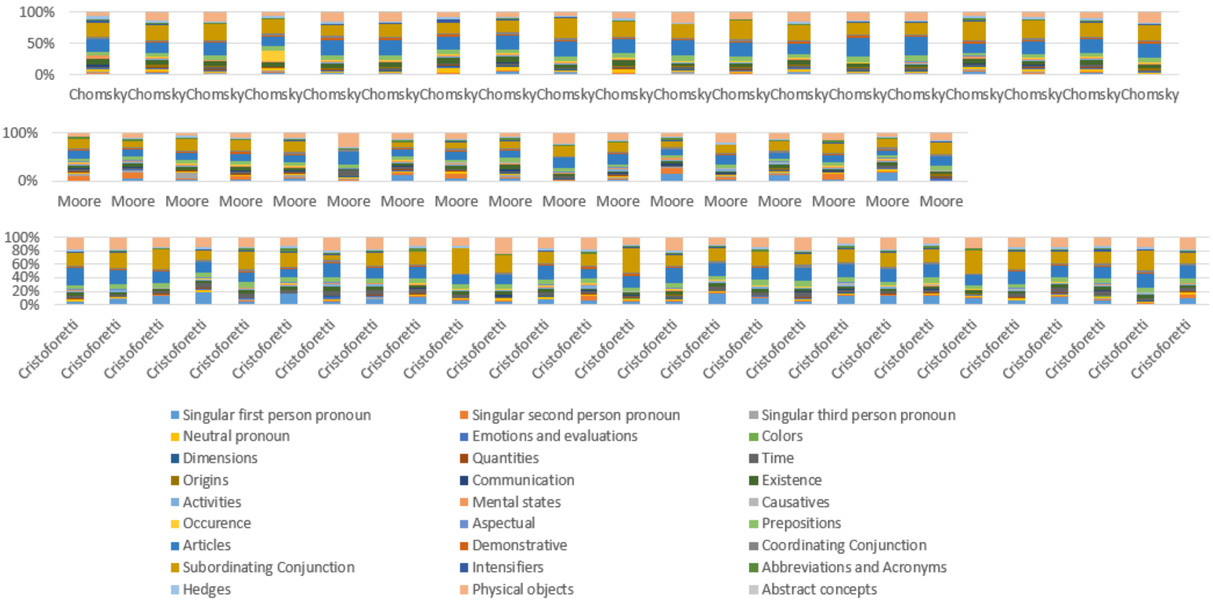


Figure 13 - Stylemes extracted analyzing text of Cristoforetti, Moore and Chomsky

In Figure 14, a different representation of these values shows how these stylistic indices change from one person to another.

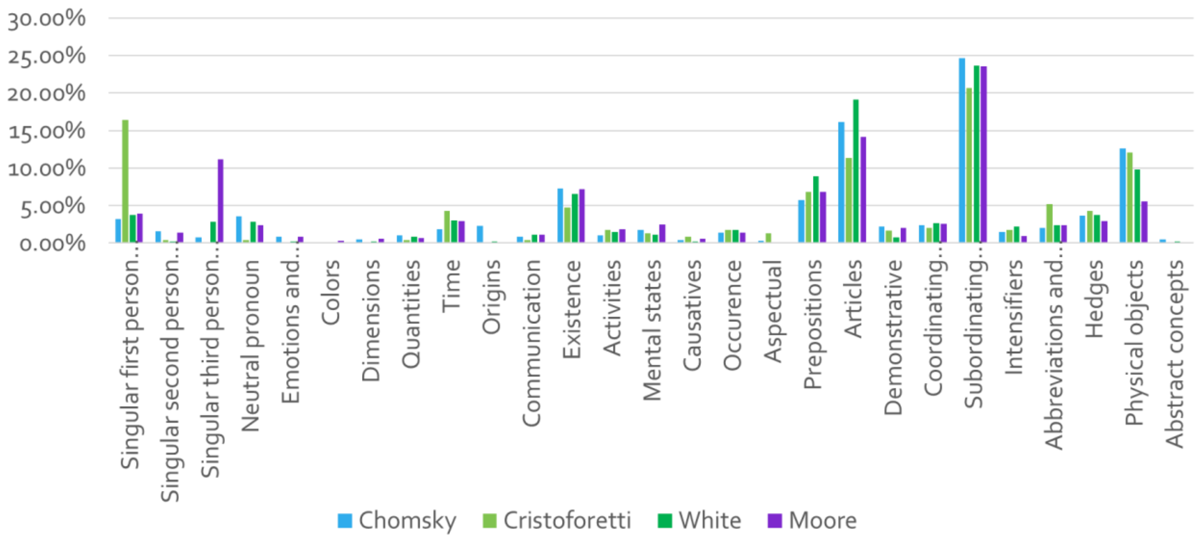


Figure 14 - Stylistic indices for Chomsky, Cristoforetti, White and Moore

4.1.3 DL/ML approach on labeled fake news dataset

In order to apply machine learning algorithms, extensive and tagged corpora are required for the training phase of the algorithms themselves. To date, corpora like these are not available on the internet. Therefore, the choice fell on a big repository of fake news on Kaggle [68], containing more than twenty-one thousand documents tagged as reliable or unreliable. They were organized in a *.csv* file, where each document had an ID, a title, an author, a text, and a label (“1” for “reliable” and “0” for “unrealiable”).
Starting from this .csv file, which was converted into an excel table, a work of pre-processing was made: only documents in English were kept, and documents with missing cells were discarded. More than twenty thousand documents (20369 documents) remained.

The number of stylistic features extracted and used for the ML and DL modules is 400.

Then, an ad-hoc parser was built in order to extract all the texts from the excel file. For each one of these, a corresponding .json and .txt file have been generated. During the parsing phase, texts have been enriched with metadata as document source, document type and language, and finally they have all been indexed in Elastic Search.

The best results are obtained with the Logistic, SimpleLogistic and RandomForest algorithms, with a brilliant 85% F-score.

Exploiting twelve thousand files of the same corpus described previously, and the same stylometric features, a further experimentation has been led using Deep Learning instruments. A designed neural network contained: 4 hidden layers have been used, for a total of 27146 trainable parameters. The neural network was trained with 10200 documents, and after 100 Epochs the result on the training set has been the following:

```
accuracy: 0.9137.
```

The next test was made on the remaining 1800 documents. The final result on the test set is: `accuracy: 0.91333333563804626.`

In this contribution, Table 3 that shows the performance of a stylometry-based machine learning model on different datasets is introduced. The table includes columns that represent the test set accuracy, epochs, training time, and additional metrics for three datasets (WelfakeNews [69], FN44k [70], and FN20k [71]), as well as the merged dataset that combined all three values.

It is expected to be demonstrated how the higher accuracy of the pooled model across different datasets suggests that pooling datasets with different characteristics and biases can make the model more resilient to different types of fake news and improve its ability to generalize to new data.

The following table shows the test set accuracy, epochs, training time, and additional metrics for three different datasets (WelfakeNews, FN44k, and FN20k), as well as the merged dataset that combines all three.

Table 3. The results of the accuracy tests.

	Neural network + stylometry (800 fea)					
	Test set accuracy	**Epoch**	**Time adds**	**fn44k acc**	**welfake acc**	**fn20k acc**
WelfakeNews	0.92	50	50 sec (gpu)	0.04		0.05
FN44k	0.99	50	50 sec (gpu)		0.44	0.6
FN20k	0.92	50	50 sec (gpu)	0.62	0.31	
merged	0.72	50	147 sec (gpu)	0.9	0.79	0.81

Each column represents the following:

- Test set accuracy: The accuracy of the model on a held-out test set of data.
- Epoch: The number of times the model has been trained on the entire dataset.
- Tempo add: The time it took to train the model.
- fn44k acc: The accuracy of the model on a specific external dataset of the data called FN44k.
- welfake acc: The accuracy of the model on a specific external dataset called WelfakeNews.
- fn20k acc: The accuracy of the model on a specific external dataset called FN20k.

These external datasets were used for each model to test the performance of the model on certain types of fake news and evaluate its ability to generalize to different datasets. However, it is important to note that these external datasets may not reflect the entire dataset or the actual distribution of fake news. The last row "merged" shows the performance of a model that was taught on a training set consisting of 90% of each of the three datasets (WelfakeNews, FN44k, and FN20k), with the remaining 10% of each dataset used as a validation set. The performance of the model was evaluated on the validation set and the other three datasets separately.

This approach is commonly used in machine learning to evaluate the generalization performance of a model on unseen data. By evaluating the model's performance on the validation set during training, it helps preventing overfitting, which occurs when a model performs well on the training data but poorly on the validation or test set.

The contrast between the lower accuracy of the merged dataset on its own test set and the higher overall accuracy on all three of the other datasets suggests that the merged model may be more robust to different types of fake news and can generalize better to new data. The possible explanation to this may be the fact that the model was trained on a diverse range of data, which can help it to learn more generalizable patterns.

Overall, the trade-off between model accuracy on individual datasets and generalization performance to new data is an important consideration in machine learning. In practice, it is often necessary to strike a balance between these two goals when creating training and validation sets and evaluating model performance.

4.2 PROTECTOR project

This section provides an overview of the PROTECTOR System, an AI facial recognition and behavioural analysis surveillance monitoring system developed by FBK in the PROTECTOR project. The system has been piloted and tested in Antwerp, Trento and Sofia by LEAs and local public authorities. The PROTECTOR project is developing new tools to protect places of worship from hate crime and terrorism and a follow-on project called PRECRISIS commences in May 2023.

Table 4 PROTECTOR project information

Project	PROTECTOR
Full Title	PROTECTing places of wORship
GRANT AGREEMENT ID	ISFP-2020-AG-PROTECT-101034216
Source of information	Project Coordinator
EU contribution	€1.860.130,80
Coordinator	Saher (Europe) OÜ
Website	https://www.protector-project.eu/
Coordinator Contact:	Andrew Staniforth andy@saher-eu.com; Dave Fortune dave@saher-eu.com;
Funding Scheme	EU's Internal Security Fund (ISFP)
Start Date	April 2021
End Date	April 2023

4.2.1 Background

Supported by the G20 Inter Faith Forum and being conducted by 5 Law Enforcement Agencies (LEAs) from Belgium, Germany, Ireland, Italy, and Sweden, the PROTECTOR (*PROTECTing places of wORship*) project is designing, developing, and delivering the next generation of integrated measures to enhance the protection of all places of worship from hate crime and terrorism.

Receiving funding from the EU's Internal Security Fund (ISFP-2020-AG-PROTECT-101034216) and being coordinated by SAHER (Europe), the PROTECTOR project recognizes that attacks on places of worship is a serious security concern across the world, as extremists and terrorists promulgating hate and violent ideologues continue to commit crimes against places of worship with alarming regularity. Recent attacks on places of cult have revealed the exploitation of their intrinsic vulnerabilities that result from their open nature and public character. Threat assessments by Europol and the EU Intelligence and Situation Centre (INTCEN) confirm the focus of places of worship in Europe as a terrorist target, which is also openly incited in terrorist publications on the internet.

To achieve the general objective of PROTECTOR, a new surveillance monitoring system with technology tools including advancements in artificial intelligence (AI) and facial recognition has been developed and is currently being tested to enhance the situational awareness, operational decision-making and investigative capability of Law Enforcement Agencies when preventing and responding to hate crimes and terrorist incidents at places of worship as part of the wider protection of public spaces.

4.2.2 Technology tools development and integration

Under *Phase 3: Technology tools development* of the project led by PROTECTOR Technical Coordinators at Fondazione Bruno Kessler (FBK) in Italy, the development of technology tools for the surveillance monitoring system have been developed, via the completion of the following key tasks:

- **Visual data acquisition and analysis:** In this task, visual data gathered from static surveillance cameras and mobile sensors was collected and analyzed to detect and track the motion of people and relevant objects (e.g., vehicles) in the scene. Specifically, streams from multiple visual sensors were automatically analyzed to determine people and objects positions, their trajectories, alongside the collection of simple information about their interactions. State-of-the-art single-stage object detectors and tracking approaches based on deep architectures were considered and extended to ensure accuracy in a wide range of environmental conditions providing 3D object pose information.
- **Video analysis for activity recognition and anomaly event detection:** This task implemented technologies for high-level analysis of visual data. Images and videos collected from static and mobile cameras were processed to automatically individuate relevant activities in the scene and detect anomalous events. Activity recognition was achieved considering state-of-the-art two-streams CNNs (i.e., operating on RGB and optical flow images), empowered with transfer learning models to improve robustness to different environmental conditions (e.g., camera view, illumination changes). The anomaly detection module considered both supervised techniques, e.g., exploiting state-of-the-art CNNs, and unsupervised deep models to enable the identification of abnormal events that are not pre-specified by human operators.
- **Identification, acquisition, and pre-processing of online data:** This task collected data from news sites and social media to identify terrorist and violent extremist content. In particular, the automatic collection of large pools of examples about online terrorist propaganda, the radicalization process and recruitment communication strategies on the Internet and social media and their evolution in time. Data collection utilized web and social media crawlers, also using freely available social network APIs (for example. Twitter and YouTube). Specific crawling techniques to collect relevant data to ensure accuracy as well as legal and ethical integrity of PROTECTOR were considered. This entailed content extraction of named entities and metadata, focused on the locations of places of worship. A monitoring workflow was implemented to cover the different languages of the use cases (i.e., English, Italian, and Bulgarian) and retrieval analyses and information online across different countries, which were then stored and ready to be reused to detect anomalous events.
- **Online data analytics for situational awareness:** This task involved the collection of information from the web and social media to identify content related to potential threats related to violent extremism and terrorism. Raw web and social media data were cleaned and normalized to allow Natural Language Processing, Social Network Analysis, and indicators for linking to specific scenarios. The relationship between social media and news content was analysed, with the

specific goal to identify which online players mainly contribute to spreading hateful content and disinformation online. Additional comparisons between current social network posts, online news, and historical data about past/attempted attacks to places of worship (and further soft targets) provided by LEAs was carried out.

- **Operational and situational awareness tool:** This task integrated previous task results into intelligent HMIs for a multi-dimensional approach, to improve the visualization of large amounts of information from heterogeneous sources. Information fusion was implemented to combine knowledge derived from different data sources and discover relevant patterns, e.g., information resulting from processing images and online data, was exploited to detect anomalous events. Moreover, in this task advanced visualization techniques were applied to enhance the perception and cognition of security officers to improve situation understanding and decision-making. Using the new visualization capabilities for exploiting the data processing techniques of the PROTECTOR project, LEAs are capable of interacting with huge amounts of information decomposed into several dimensions about elements of current ongoing operations that facilitate their work. This approach will help security officers implement a collective response to potential threats.
- **Identity Management:** This task carried out the development of different components and services necessary for the realization and fruition of the products, while considering the requirements elicited from the use cases and those identified earlier in the project. Participants already employ differing Identity Management (IDM) solutions, with which the components now integrate; these, together with the product deployment model, have determined the adoption or development of specific IDM solutions – for instance the federation with existing Identity Providers and authentication methods vs. the enrolment of new identities and their authenticators.
- **Privacy Management:** This task was devoted to developing the necessary mechanism to provide the data subjects with a solution to provide stakeholders with capabilities to comply with applicable regulations and requirements. These included defining and enforcing access controls to data for well-defined and limited purposes, safeguarding data with encryption at rest and in transit, and maintaining tamper-evident auditable logs.

4.2.3 Testing and evaluation

To test and evaluate the PROTECTOR surveillance monitoring system a series of pilots have been designed. As an integral part in the planning of the pilot programme, the pilot testing criteria and evaluation matrix have been designed. The pilot tests were planned to take place in three separate locations and scheduled to ensure sufficient time between each pilot to evaluate and review operational effectiveness. The first pilot was be conducted in Trento, Italy and was coordinated by the local municipality police during September 2022. The second pilot was conducted in Antwerp, Belgium, and was coordinated by the police in Antwerp, being aligned to a major counterterrorism and hostage siege training exercise at a synagogue in Antwerp during December 2022. The third and final pilot will be conducted in Sofia, Bulgaria, and coordinated by the European Institute with support from local police on 30th March 2023.

4.2.4 Next steps

The finalised PROTECTOR surveillance monitoring system will be presented at a workshop during the PROTECTOR conference to take place in Nicosia, Cyprus on 25th April 2023, jointly hosted by sister project PROSECUW (*PROtection and SECUrity for places of Worship*).

Table 5 PROSECUW project information

Project	PROSECUW
Full Title	PROtection and SECUrity for places of Worship
GRANT AGREEMENT ID	ISFP-2020-AG-PROTECT-101034232
Source of information	Project Coordinator
EU contribution	€1.121.895
Coordinator	Center for Social Innovation (CSI)
Website	https://prosecuwproject.eu/
Coordinator Contact:	Panayiota Constanti panayiota.constanti@csicy.com
Funding Scheme	EU's Internal Security Fund (ISFP)
Start Date	May 2021
End Date	April 2023

To further develop the PROTECTOR system, a new EU funded project has received support from the Internal Security Fund (ISFP-2022-TFI-AG-PROTECT-02-101100539). The PRECRISIS (*PRotECting public spaces thRough Integrated Smarter Innovative Security)* project includes primary consortium partners from the PROTECTOR project and will commence on 1st May 2023. This 2-year initiative is coordinated by SAHER (Europe), and will further develop the PRTOTECTOR system, integrating new and advanced AI and surveillance technologies for smart-city adoption and will be further tested for full operational deployment at the close of the PRECRISIS project in April 2025.

Table 6 PRECRISIS project information

Project	PRECRISIS
Full Title	PRotECting public spaces thRough Integrated Smarter Innovative Security
GRANT AGREEMENT ID	ISFP-2022-TFI-AG-PROTECT-02-101100539
Source of information	Coordinator
EU contribution	€1.996.347,15
Coordinator	Saher (Europe) OÜ
Website	-
Coordinator Contact	Andrew Staniforth andy@saher-eu.com; Dave Fortune dave@saher-eu.com;
Funding Scheme	EU's Internal Security Fund (ISFP)
Start Date	May 2023
End Date	April 2025

4.3 iCognative technology

This section provides an update on iCognative, the brain fingerprinting technology which has now been operationally deployed.

The iCognative deception detection technology has been developed by Brainwave Science [72], a company based in the USA founded in 2012. According to Brainwave Science, iCognative technology has applications in national security, counterterrorism, border security, human and drug trafficking and immigration control for defence, intelligence, and law enforcement agencies worldwide.

4.3.1 Background

iCognative deception detection technology uses the P300 brain wave response mechanism which identifies whether information is stored in the brain by precisely measuring brainwaves. Unlike a conventional polygraph, which detects an emotional stress response on the theory that people are more stressed when lying, iCognative only detects whether the information exists within the brain or not. The system is fully automated using a laptop and specialized wireless headset and is claimed by Brainwave science to have an accuracy rate of over 99%.

While the neurological origins of the electrical surge remain unclear, neuroscientific research has used the discovery of P300 to advance the field of brain fingerprinting (BF) which detects concealed information stored in the brain by measuring brainwaves. As an example, P300 responds to words or pictures relevant to a crime scene or terrorist bomb-making knowledge, therefore detecting information by measuring cognitive information processing of high value to investigators.

Developments in BF over recent years has led to the design of several applications in brain-computer interfacing (BCI) for P300 which have now been realized, following recognition that it has several desirable qualities that aid in the implementation of computer systems. For example, the P300 response waveform is consistently detectable and is elicited in response to precise stimuli. The P300 response waveform can also be evoked in nearly all subjects with little variation in measurement techniques, which help simplify interface designs and permit greater usability, resulting in technology now being available with a monitoring headset linked to a computer dashboard visually displaying results in a graph in real time. Moreover, the P300 wave is recognized in neuroscience as part of the event-related potential (ERP) component elicited in the process of decision making. It is an endogenous potential, as its occurrence links not to the physical attributes of a stimulus, but to a person's reaction to it.

4.3.2 Operational application

To date, BF is considered a technique of proven accuracy for US Government tests, and it has been ruled as admissible in one US court as scientific evidence. The Brainwave Science device has been tested by several US federal government agencies and has now been purchased by several law enforcement agencies with promising results.

During 2021, Dubai Police homicide detectives reportedly solved a challenging murder case using iCognative Brainwave Science technology for the very first time [73]. In a remarkable example of embracing new innovative technologies and scientific developments, under the direction of Lieutenant

General Abdullah Khalifa Al Marri, Commander-in-Chief of Dubai Police, new neuroscience tools were used to measure the brain waves of suspects leading to a breakthrough in the murder investigation.

Described as the 'next-generation' deception detection technology, advancing traditional law enforcement lie detection techniques currently in operation, Major General Dr Ahmed Eid Al-Mansoori, Director of the Department of Criminology in Dubai, oversaw the unique approach alongside forensic psychology experts to apply brainwave science in their investigative tasks, conducting experiments for a year, resulting in the capture of information leading to the identification of the murder suspect. The ground-breaking use of neuroscience technology by Dubai Police has provided evidence of its operational potential for criminal investigations.

According to Muhammad Issa Al-Hammadi, Director of the Criminology Department in Dubai, the application of the brainwave technology was dependent upon the understanding of psychology, explaining that: "A person's memory stores life events and details and experiences, and when they appear before them again the brain stimulates and emits waves as a result of its retrieval of these events, and thus the extent of a person's knowledge of the events can be measured through electrical waves the brain emits after viewing the pictures.

Muhammad Issa Al-Hammadi went on to reveal that: "*Dubai Police measured these waves originating from the brain after a person sees images related to the location or tool of the crime, and then provides an accurate analytical reading about whether the person or suspect was present in the location of the crime, and whether they know the tool used or even identify the victim by the frequency of these waves*."

4.3.3 International awareness

As the deployment of iCognative reveals further evidence of its practical application, awareness of its use is being further examined and explored. As an example, at the 2022 Australia New Zealand Policing Advisory Agency (ANZPAA) conference, dedicated to Navigating the Next Generation of Policing, a series of future challenges and opportunities were presented and discussed by international experts gathered in Melbourne. A fascinating input was provided by Future Trends Analyst Michael McQueen, whose opening presentation on 'Preparing now for what's next', highlighted the work of Dubai Police in solving a murder case using BF technology.

This presentation at the very outset of the ANZPAA conference sparked significant interest in the iCognative technology. Moreover, at the recent Police World Summit held in Dubai during March 2023, Dubai Police were exhibiting and showcasing the iCognative Brainwave Science technology, providing another example of where this new approach to deception detection is gaining attention and uptake at seminal events promoting new and emerging technology, tools, and techniques to improve and develop law enforcement agency policy, practice and procedure.

4.3.4 Next steps

It is becoming increasingly evident that BF technologies, tools and techniques provides a fascinating opportunity to profoundly impact upon intelligence operations and law enforcement investigations. Although BF is not currently used or applied in the EU, there remains the potential for significant positive impact upon and within member states' criminal justice systems. The practical application and adoption of BF and related P300 techniques and technologies remains controversial, and although there is a growing body of evidence to support its positive use, scientific opinion remains divided as to its accuracy and efficacy. The use and application of iCognative remains in its infancy and further evidence through controlled scientific trials are required to convince many in authority that investments in such technologies are appropriate, proportionate, and necessary.

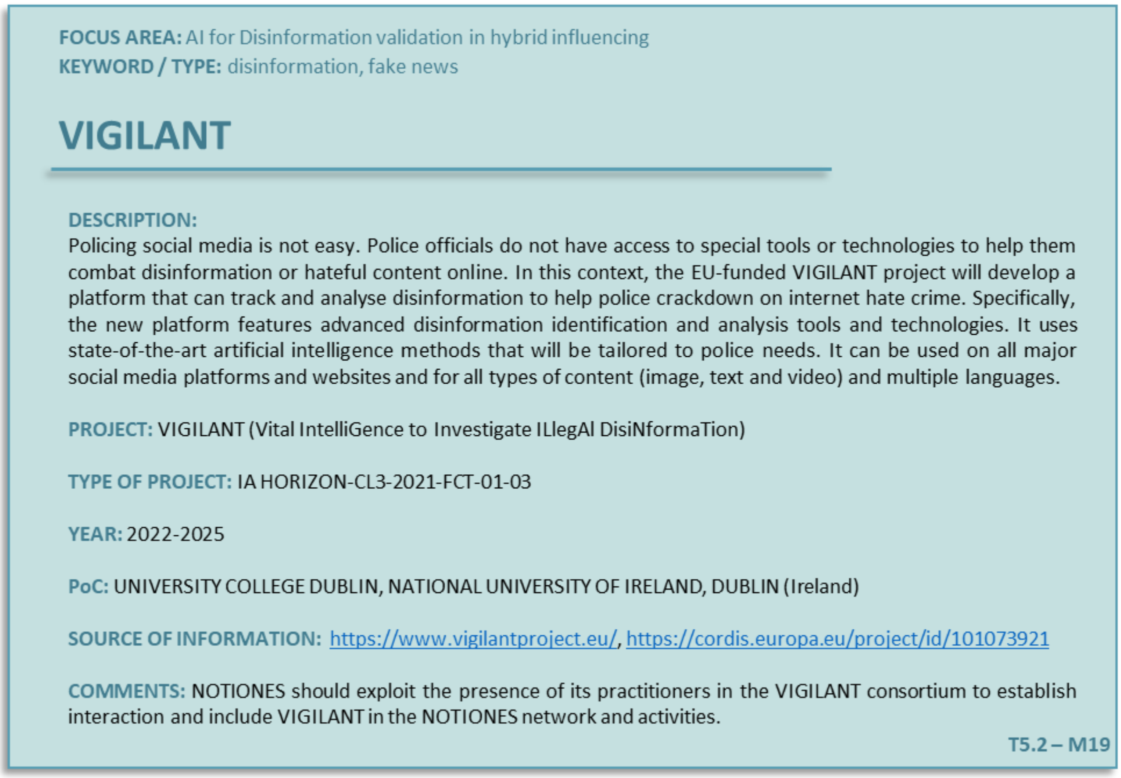
iCognative and related BF currently are at the cutting edge of deception detection techniques for intelligence and law enforcement agencies, and authorities are right to remain cautious given legal and ethical implications of its use, although this cautious approach should not prevent further examination and continued monitoring of iCognative and related technologies to assess and analyse their future potential.

5. Main findings

This section presents the main findings of the research described in the previous sections of this document. The common layout for the summarization of the information proposed in deliverable D5.1 is used.

VIGILANT project

VIGILANT is a very promising EC-funded project that tackles disinformation with specific focus on Police officials' needs and will last until 2025.



FOCUS AREA: AI for Disinformation validation in hybrid influencing
KEYWORD / TYPE: disinformation, fake news

VIGILANT

DESCRIPTION:
Policing social media is not easy. Police officials do not have access to special tools or technologies to help them combat disinformation or hateful content online. In this context, the EU-funded VIGILANT project will develop a platform that can track and analyse disinformation to help police crackdown on internet hate crime. Specifically, the new platform features advanced disinformation identification and analysis tools and technologies. It uses state-of-the-art artificial intelligence methods that will be tailored to police needs. It can be used on all major social media platforms and websites and for all types of content (image, text and video) and multiple languages.

PROJECT: VIGILANT (Vital IntelliGence to Investigate ILlegAl DisiNformaTion)

TYPE OF PROJECT: IA HORIZON-CL3-2021-FCT-01-03

YEAR: 2022-2025

PoC: UNIVERSITY COLLEGE DUBLIN, NATIONAL UNIVERSITY OF IRELAND, DUBLIN (Ireland)

SOURCE OF INFORMATION: https://www.vigilantproject.eu/, https://cordis.europa.eu/project/id/101073921

COMMENTS: NOTIONES should exploit the presence of its practitioners in the VIGILANT consortium to establish interaction and include VIGILANT in the NOTIONES network and activities.

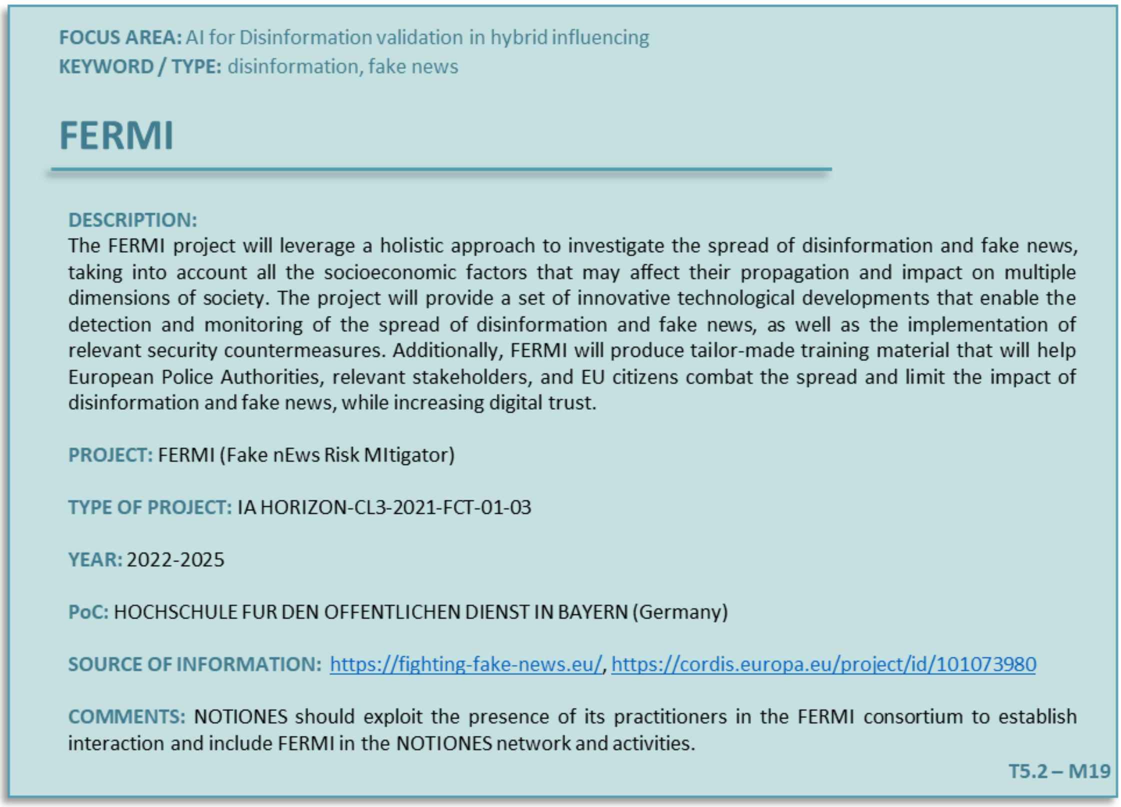
T5.2 – M19

Figure 15 - Main results: VIGILANT project

NOTIONES will contact the Project coordinator and propose to start interaction with the Working Group dedicated "AI for disinformation validation in Hybrid influencing".

FERMI project

FERMI is a very promising EC-funded project that tackles disinformation with a holistic approach and will last until 2025.



FOCUS AREA: AI for Disinformation validation in hybrid influencing
KEYWORD / TYPE: disinformation, fake news

FERMI

DESCRIPTION:
The FERMI project will leverage a holistic approach to investigate the spread of disinformation and fake news, taking into account all the socioeconomic factors that may affect their propagation and impact on multiple dimensions of society. The project will provide a set of innovative technological developments that enable the detection and monitoring of the spread of disinformation and fake news, as well as the implementation of relevant security countermeasures. Additionally, FERMI will produce tailor-made training material that will help European Police Authorities, relevant stakeholders, and EU citizens combat the spread and limit the impact of disinformation and fake news, while increasing digital trust.

PROJECT: FERMI (Fake nEws Risk MItigator)

TYPE OF PROJECT: IA HORIZON-CL3-2021-FCT-01-03

YEAR: 2022-2025

PoC: HOCHSCHULE FUR DEN OFFENTLICHEN DIENST IN BAYERN (Germany)

SOURCE OF INFORMATION: https://fighting-fake-news.eu/, https://cordis.europa.eu/project/id/101073980

COMMENTS: NOTIONES should exploit the presence of its practitioners in the FERMI consortium to establish interaction and include FERMI in the NOTIONES network and activities.

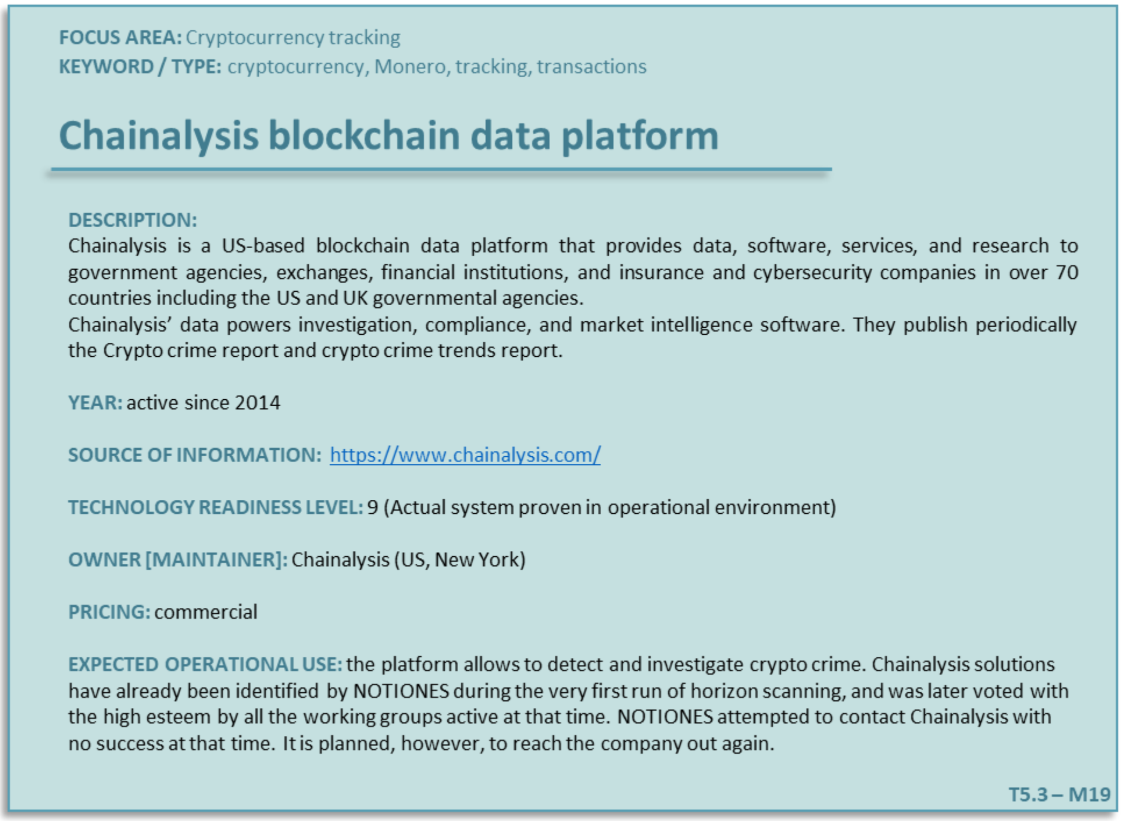
T5.2 – M19

Figure 16 - Main results: FERMI project

NOTIONES contacted the Project coordinator and an interaction was successfully established. FERMI will be present at the second NOTIONES Conference in Paris in May 2023 and will hopefully collaborate with the NOTIONES Working Group dedicated "AI for disinformation validation in Hybrid influencing".

Chainalysis

The Chainalysis platform allows to detect and investigate crypto crime.



FOCUS AREA: Cryptocurrency tracking

KEYWORD / TYPE: cryptocurrency, Monero, tracking, transactions

Chainalysis blockchain data platform

DESCRIPTION:

Chainalysis is a US-based blockchain data platform that provides data, software, services, and research to government agencies, exchanges, financial institutions, and insurance and cybersecurity companies in over 70 countries including the US and UK governmental agencies.

Chainalysis' data powers investigation, compliance, and market intelligence software. They publish periodically the Crypto crime report and crypto crime trends report.

YEAR: active since 2014

SOURCE OF INFORMATION: https://www.chainalysis.com/

TECHNOLOGY READINESS LEVEL: 9 (Actual system proven in operational environment)

OWNER [MAINTAINER]: Chainalysis (US, New York)

PRICING: commercial

EXPECTED OPERATIONAL USE: the platform allows to detect and investigate crypto crime. Chainalysis solutions have already been identified by NOTIONES during the very first run of horizon scanning, and was later voted with the high esteem by all the working groups active at that time. NOTIONES attempted to contact Chainalysis with no success at that time. It is planned, however, to reach the company out again.

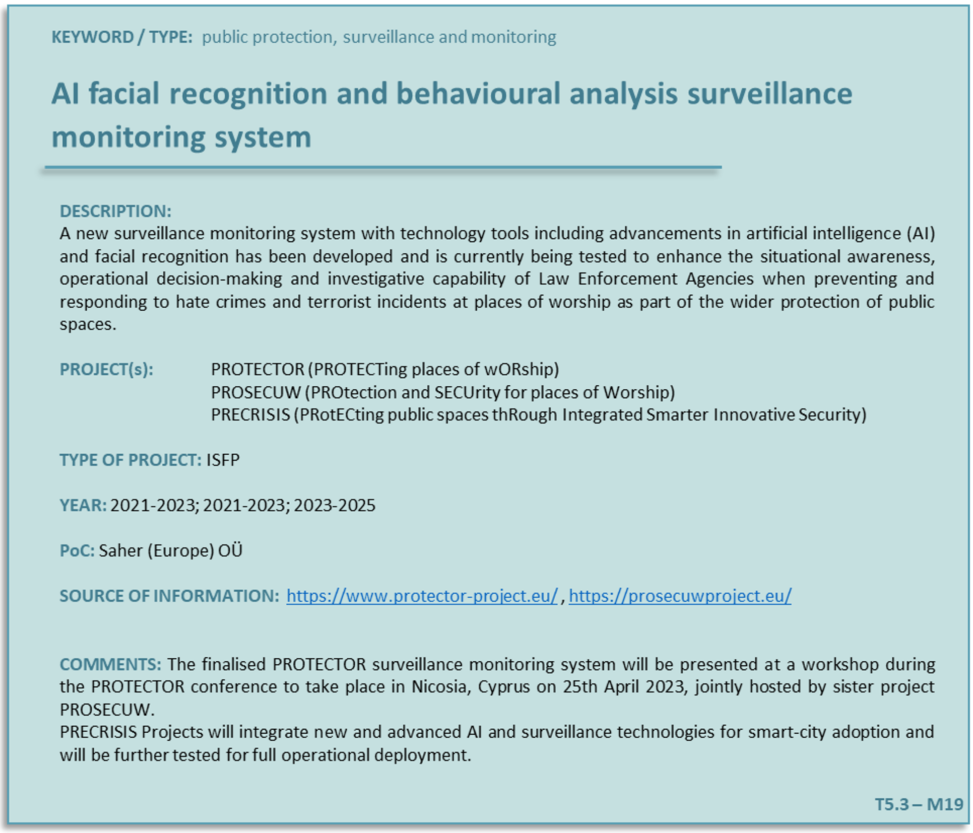
T5.3 – M19

Figure 17 - Main results: Chainalysis blockchain data platform

Chainalysis solutions have already been identified by NOTIONES during the very first run of horizon scanning, and was later voted with the high esteem by all the working groups active at that time. NOTIONES attempted to contact Chainalysis with no success at that time. It is planned, however, to reach the company out again.

PROTECTOR/PRECRISIS surveillance monitoring system

The PROTECTOR Project developed a new surveillance monitoring system with technology tools including advancements in artificial intelligence (AI) and facial recognition which is currently being tested to enhance the situational awareness, operational decision-making and investigative capability of Law Enforcement Agencies when preventing and responding to hate crimes and terrorist incidents at places of worship as part of the wider protection of public spaces.



KEYWORD / TYPE: public protection, surveillance and monitoring

AI facial recognition and behavioural analysis surveillance monitoring system

DESCRIPTION:
A new surveillance monitoring system with technology tools including advancements in artificial intelligence (AI) and facial recognition has been developed and is currently being tested to enhance the situational awareness, operational decision-making and investigative capability of Law Enforcement Agencies when preventing and responding to hate crimes and terrorist incidents at places of worship as part of the wider protection of public spaces.

PROJECT(s): PROTECTOR (PROTECTing places of wORship)
PROSECUW (PROtection and SECUrity for places of Worship)
PRECRISIS (PRotECting public spaces thRough Integrated Smarter Innovative Security)

TYPE OF PROJECT: ISFP

YEAR: 2021-2023; 2021-2023; 2023-2025

PoC: Saher (Europe) OÜ

SOURCE OF INFORMATION: https://www.protector-project.eu/ , https://prosecuwproject.eu/

COMMENTS: The finalised PROTECTOR surveillance monitoring system will be presented at a workshop during the PROTECTOR conference to take place in Nicosia, Cyprus on 25th April 2023, jointly hosted by sister project PROSECUW.
PRECRISIS Projects will integrate new and advanced AI and surveillance technologies for smart-city adoption and will be further tested for full operational deployment.

T5.3 – M19

Figure 18 - Main results: PROTECTOR/PRECRISIS surveillance monitoring system

NOTIONES has already interacted with PROTECTOR by participating in its Ethics & AI seminar in November 2022 and it is suggested to continue the interaction also in the future, with the follow-up Project PRECRISIS.

6. Conclusions and next steps

This document represents the product of the third run of tasks T5.2 and T5.3 of WP5, which performed the research monitoring activities during months M19 and M20 (March, April 2023).

Task T5.2 performed the horizon scanning activity through exploratory research on the online scholar and patent database The Lens and the open web. Results were found about: Traceable Monero, Weak points of Monero, Analyss of currency hard forks (TITANIUM Project), Chainalysis, methods and algorithms for collecting, detecting, and visualising fake news, propaganda, deep fakes, and inconsistent statements to a recipient group, as well as patents for misinformation detection.

Task T5.3 performed the monitoring of EU-funded research projects on the CORDIS database and found five projects with topic related to the focus area "Tools for tracing cryptocurrencies used in criminal finances" and fifteen projects with topic related to the focus area "AI for Disinformation validation in hybrid influencing".

WP5 also performed thematic deepening on selected topics: disinformation detection technology through NLP, the PROTECTOR project, and the iCognative technology.

In conclusion, the results of the third run of tasks T5.2 and T5.3 were documented in this report. The main findings are summarised in section 4.3, highlighting technologies and EU projects that are most promising for the purposes of NOTIONES:

- TRACE project for fighting disinformation;
- FERMI project for fighting disinformation;
- Chainalysis solutions for cryptocurrency tracking;
- PROTECTOR/PRECRISIS surveillance monitoring system.

The TRACE project will attend NOTIONES Second Conference in May 2023 and will deliver a presentation of their main findings.

It should be noted that the Horizon Europe Framework Programme envisages a specific call in 2024 about "Tracing of cryptocurrencies transactions related to criminal purposes" (Topic ID: HORIZON-CL3-2024-FCT-01-06 [74]). The next run of task T5.2 will monitor this call and will try to establish contacts with the winning proposal(s).

With regard to the next steps, two main actions are foreseen in the next runs of tasks T5.2 and T5.3 in months M25-M26:

- The research of T5.2 will be repeated on CORDIS with updated search parameters;
- New research topics will be targeted in the next run of the tasks, corresponding to the new focus areas tackled by upcoming working groups.

Updates will be included in the next release of the deliverable D5.5 "*Monitoring of EU Research and Horizon Scanning -v4*", due in M26.

References

[1] TheLens. [Online]. Available: https://www.lens.org/.

[2] Publications Office of the European Union, “COmmunity Research and Development Information Service,” [Online]. Available: https://cordis.europa.eu/en.

[3] Y. Li, G. Yang, W. Susilo, Y. Yu, M. Ho Au and D. Liu, “Traceable Monero: Anonymous Cryptocurrency with Enhanced Accountability,” *IEEE,* p. 13, 2021.

[4] A. Hinteregger and B. Haslhofer, “An Empirical Analysis of Monero Cross-Chain traceability,” p. 8, 2019.

[5] M. Malte, K. Soska, E. Heilman, K. Lee, H. Heffan, S. Srivastava, K. Hogan, J. Hennessey, A. Miller, A. Narayanan and N. Christin, “An Empirical Analysis of Traceability in the Monero blockchain,” *Proceedings on Privacy Enhancing Technologies,* no. 3, pp. 143-163, 2018.

[6] T. Cao, J. Yu, J. Decouchant, X. Luo and P. Verissimo, “Exploring the Monero Peer-to-Peer Network,” *Lecture Notes in Computer Science,* p. 578–594, 2020.

[7] J. Barcelo., “User privacy in the public bitcoin blockchain,” 2014.

[8] P. Koshy, D. Koshy and P. McDaniel, “An analysis of anonymity in bitcoin using p2p network traffic,” *Springer, international conference on Financial Cryptography and Data Security,* pp. 469-485, 2014.

[9] F. Reid and M. Harrigan, “An analysis of anonymity in the bitcoin system,” *Security and Privacy in Social Networks,* pp. 197-223, 2013.

[10] G. Danezis, Meiklejohn and S., “centrally banked cryptocurrencied,” 2015.

[11] A. Kumar, C. Fischer, S. Tople and P. Saxena, “A Traceability Analysis of Monero’s Blockchain,” *Springer International Publishing,* p. 153–173, 2017.

[12] Chainalysis, “The Blockchain Data Platform,” [Online]. Available: https://www.chainalysis.com/.

[13] H. Alvari, E. Shaabani and P. Shakarian, “Early Identification of Pathogenic Social Media Accounts,” *2018 IEEE International Conference on Intelligence and Security Informatics (ISI),* pp. 169-174, 2018.

[14] Arizona State University, “Privacy protection systems and methods”. US Patent 202016782349, 6 September 2022.

[15] K. Shu, D. Mahudeswaran, S. Wang, D. Lee and H. Liu, “FakeNewsNet: A Data Repository with News Content, Social Context, and Spatiotemporal Information for Studying Fake News on Social Media,” *Big data,* vol. 8, no. 3, pp. 171-188, 2020.

[16] “PolitiFact,” [Online]. Available: https://www.politifact.com/.

[17] “GossipCop,” [Online]. Available: https://www.gossipcop.com/.

[18] Arizona State University, “Method and apparatus for collecting, detecting and visualizing fake news”. US Patent US 11494446 B2, 8 November 2022.

[19] P. A. Chew and J. Glicken Turnley, “SBP-BRiMS - Understanding Russian Information Operations Using Unsupervised Multilingual Topic Modeling,” *Social, Cultural, and Behavioral Modeling,* pp. 102-107, 2017.

[20] S. S. Bodrunova, I. S. Blekanov and M. Kukarkin, “Topics in the Russian Twitter and Relations between their Interpretability and Sentiment,” *2019 Sixth International Conference on Social Networks Analysis, Management and Security (SNAMS),* pp. 549-554.

[21] C. P. Alexander and Galisteo Consulting Group Inc, “Identifying propaganda in global social media”. US Patent US 10140289 B2, 27 November 2018.

[22] S. McCloskey and M. Albright, “Detecting GAN-Generated Imagery Using Saturation Cues,” *2019 IEEE International Conference on Image Processing (ICIP),* pp. 4584-4588, 2019.

[23] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford and X. Chen, “NIPS - Improved techniques for training GANs,” *Neural Information Processing Systems,* vol. 29, pp. 2234-2242, 2016.

[24] Ford Global Technologies Llc, “RCCC to RGB domain translation with deep neural networks”. CN, DE, US Patent US 11068749 B1, 20 July 2021.

[25] L. V. S. Lakshmanan, M. G. Simpson and S. Thirumuruganathan, “Combating fake news: a data management and mining perspective,” *Proceedings of the VLDB Endowment,* vol. 12, no. 12, pp. 1990-1993, 2019.

[26] U. N. Polytechnical, “Method for detecting false messages in social media”. CN Patent 202010921501, 18 December 2020.

[27] International Business Machines Corporation , “Assessment of inconsistent statements to a recipient group”. US Patent US 11443208 B2, 13 September 2022.

[28] Microsoft Technology Licensing Llc, “Data privacy pipeline providing collaborative intelligence and constraint computing”. WO, EP, US, CN Patent US 11455410 B2, 27 September 2022.

[29] F. Montori, L. Bedogni and L. Bononi, “A Collaborative Internet of Things Architecture for Smart Cities and Environmental Monitoring,” *IEEE Internet of Things Journal,* vol. 5, no. 2, pp. 592-605, 2018.

[30] University of Bologna, “SenSquare,” [Online]. Available: http://sensquare.disi.unibo.it/.

[31] S. K. Sharma and X. Wang, “Live Data Analytics With Collaborative Edge and Cloud Processing in Wireless IoT Networks,” *IEEE Access,* vol. 5, pp. 4621-4635, 2017.

[32] IBM Corp.;, “Mitigating misinformation in encrypted messaging networks”. US Patent 202016800361, 26 August 2021.

[33] Accenture Global Solutions Limited , “METHOD AND SYSTEM FOR DETECTION OF MISINFORMATION”. Patent 202117501195 , 1 December 2022.

[34] E. Pavlov, “System and Method for Detecting Misinformation and Fake News via Network Analysis”. US Patent 201917293748, 27 October 2022.

[35] IBM Corp., “Presenting thought-provoking questions and answers in response to misinformation”. 2022 Patent 202117184668, August 25 2022.

[36] Ericsson Telefon Ab L M, “Methods and devices for avoiding misinformation in machine learning”. WO Patent 2020066483 , 23 December 2021.

[37] E. Commission, “Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions tackling online disinformation: a European approach. COM/2018/236 final.,” 2018.

[38] Truthnest, “Truthnest,” [Online]. Available: https://www.truthnest.com/ .

[39] SOMA, “ D5.1 Impact Assessment Methodology,” [Online]. Available: https://www.disinfobservatory.org/wp-content/uploads/2020/05/D5.1-Impact-Assessment-Methodology.pdf.

[40] Truly, “Truly,” [Online]. Available: https://www.truly.media/.

[41] V. P´erez-Rosas, B. Kleinberg, A. Lefevre and M. R., “Automatic detection of fake news,” *Proceedings of the 27th International Conference on Computational Linguistics,* p. 3391–3401, 2018.

[42] R. Oshikawa, J. Qian and W. Wang, “A Survey on Natural Language Processing for Fake News Detection,” *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020),* pp. 6086-6093, 2020.

[43] M. Elhadad, K. Li and F. Gebali, “Fake News Detection on Social Media: A Systematic Survey,” *2019 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM),* pp. 1-8, 2019.

[44] D. Cer, “Universal sentence encoder,” *arXiv:1803.11175.*

[45] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems (NIPS),* 2017.

[46] J. Khan, M. Khondaker, S. Afroz, G. Uddin and A. Iqbal, “A benchmark study of machine learning models for online fake news detection,” *Machine Learning with Applications,* vol. 4, p. 100032, 2021.

[47] R. Zellers, A. Holtzman, H. Rashkin, Y. Bisk, A. Farhadi, F. Roesner and Y. Choi, “Defending Against Neural Fake News,” 2020.

[48] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei and I. Sutskever, “Language models are unsupervised multitask learners,” OpenAI blog, 2019. [Online].

[49] I. Solaiman, M. Brundage, J. Clark, A. Askell, A. Herbert-Voss, J. Wu, A. Radford and J. Wang, “Release strategies and the social impacts of language models,” *OpenAI Report,* 2019.

[50] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen and V. R. Stoyanov, “A robustly optimized bert pretraining approach,” 2019.

[51] H. Jwa, D. Oh, K. Park, J. Kang and H. Lim, “exBAKE: Automatic Fake News Detection Model Based on Bidirectional Encoder Representations from Transformers (BERT),” *Applied Sciences,* vol. 9, p. 4062, 2019.

[52] W. Marcellino, “Detecting Conspiracy Theories on Social Media Improving Machine Learning to Detect and Understand Online Conspiracy Theories,” RAND CORP, Santa Monica CA, 2021.

[53] T. Fagni, F. Falchi, M. Gambini, A. Martella and M. Tesconi, “TweepFake: About detecting deepfake tweets,” *PLoS ONE,* vol. 16, 2021.

[54] R. Gordon, “Better fact-checking for fake news,” MIT News, 2019. [Online]. Available: https://news.mit.edu/2019/better-fact-checking-fake-news-1017.

[55] S. Ghosh and C. Shah, “Towards automatic fake news classification,” *Proceedings of the Association for Information Science and Technology,* vol. 55, no. 1, pp. 805-807, 2018.

[56] S. Shaar, N. Babulkov, G. Da San Martino and P. Nakov, “That is a Known Lie: Detecting Previously Fact-Checked Claims,” *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL),* p. 3607–3618, 2020.

[57] Repustar, “FactSparrow,” [Online]. Available: https://factsparrow.repustar.com.

[58] G. Cloud, “Analyzing Entity Sentiment,” [Online]. Available: https://aws.amazon.com/it/machine-learning/automl/.

[59] Neo4j, “Build a Knowledge Graph using NLP and Ontologies,” [Online]. Available: https://neo4j.com/developer/graph-data-science/build-knowledge-graph-nlp-ontologies.

[60] Google, “AutoML solutions,” [Online]. Available: https://cloud.google.com/automl?hl=it.

[61] Microsoft, “Automated machine learning,” [Online]. Available: https://azure.microsoft.com/en-us/products/machine-learning/automatedml/.

[62] Amazon, “AutoML solutions,” [Online]. Available: https://aws.amazon.com/it/machine-learning/automl/.

[63] G. Bond and A. Lee, *Applied Cognitive Psychology,* vol. 19, pp. 313 - 329, 2005.

[64] M. Frank, M. Menasco and M. O'Sullivan, “Human Behavior and Deception Detection,” in *Wiley Handbook of Science and Technology for Homeland Security*, 2008.

[65] B. Ghanem, S. P. Ponzetto and P. Rosso, “FacTweet: profiling fake news twitter accounts,” *International Conference on Statistical Language and Speech Processing,* p. 35–45, 2020.

[66] M. Oakes, “Literary Detective Work on the Computer,” *Natural Language Processing,* vol. 12, p. 283, 2014.

[67] W. Ce, Y. Hongzhi and W. Fucheng, “Information Retrieval Technology Based on Knowledge Graph,” *Advances in Engineering Research,* vol. 162.

[68] Kaggle, “Fake News repository,” [Online]. Available: https://www.kaggle.com/c/fake-news/data.

[69] WeFakeNews, “repository,” [Online]. Available: https://zenodo.org/record/4561253.

[70] FN44k, “repository,” [Online]. Available: https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news-dataset.

[71] FN20k, “repository,” [Online]. Available: https://www.kaggle.com/c/fake-news/data?select=train.csv .

[72] Brainwave Science, “World leaders for Intelligence solutions,” [Online]. Available: https://www.linkedin.com/company/brainwave-science-inc. [Accessed 22 March 2023].

[73] Gulf Today, “Dubai Police crack murder case using ‘brain fingerprint’ technology,” [Online]. Available: https://www.gulftoday.ae/news/2021/01/25/dubai-police-crack-murder-case-using-brain-fingerprint-technology. [Accessed 22 March 2023].

[74] EC, “Horizon-IA HORIZON Innovation Actions,” [Online]. Available: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/topic-details/horizon-cl3-2024-fct-01-06;callCode=null;freeTextSearchKeyword=;matchWholeText=false;typeCodes=1,2,8;statusCodes=31094501,31094502;programmePeriod=2021%20-%2.

[75] L. Nicholas, Q. Mu, O. Y. Jeremy, V. M. Affonso, P. C. Santos and D. P. R. Abreu, “Presenting thought-provoking questions and answers in response to misinformation,” 2022.

[76] V. F. D. Santana, M. A. Vasconcelos, M. C. Pichiliani and H. C. D. S. P. Candello, “Mediating between social networks and paid curated content producers in misinformative content mitigation,” 2022.

[77] W.-Y. Wang and W.-C. Peng, “Team Yao at Factify 2022: Utilizing Pre-trained models and Co-attention networks for multi-modal fact verification,” *AAAI 2022 De-Factify Workshop: First Workshop on Multimodal Fact-Checking and Hate Speech Detection,* 2022.

[78] R. Sultana, M. K. Hassan, M. R. Hassan, S. R. Sourav, M. A. Huraira and d. S. Ahmed, “An Effective Fake News Detection on Social Media and Online News,” *Australian journal of engineering and innovative technology,* vol. 5, no. 4, pp. 109-120, 2022.