



Funded by the European Union  
Horizon Europe  
(HORIZON-CL4-2021-HUMAN-01-27  
AI to fight disinformation)

[www.ai4trust.eu](http://www.ai4trust.eu)



**AI4TRUST**

**D7.2**

**Innovation, Exploitation  
and Sustainability v1**



FONDAZIONE  
BRUNO KESSLER



CERTH  
CENTRE FOR  
RESEARCH IN TECHNOLOGY  
HELLAS



UNIVERSITA  
DI TRENTO



NATIONAL CENTRE FOR  
SCIENTIFIC RESEARCH 'DEMOKRITOS'



CNRS  
CENTRE NATIONAL  
DE LA RECHERCHE  
SCIENTIFIQUE



GDI  
Global  
Information  
Index

sky



DEMAGOG

MALDITA.ES  
FUNDACION PARA UNA NUEVA CULTURA

ASTIKI MI KERDOSOPHIKI ETABRIA KENTRO  
KATHIPELENSIS TIS PARAPLIRAFORISIS /  
CIVIL NON-PROFIT COMPANY KENTRO  
KATHIPELENSIS TIS PARAPLIRAFORISIS

EURACTIV

ASSOCIATIA  
DIGITAL  
BRIDGE

EUROPEJSKIE  
MEDIA SP 200

FINCONS  
GROUP

UNIVERSITY OF  
CAMBRIDGE



Project acronym	AI4TRUST
Project full title:	AI-based-technologies for trustworthy solutions against disinformation
Grant info:	ID 101070190-AI4TRUST
Funding:	EU-funded under Digital, Industry, and space Overall budget € 5.950.682,50
Version:	1.5
Status	Final Version
Dissemination level:	Sensitive — limited under the conditions of the Grant Agreement
Due date	29/02/2024
Delivery date (re-submission)	04/06/2024
Work Package:	WP7 - Communication, Dissemination and Exploitation
Lead partner for this deliverable:	Fincons Group AG (FINCONS)
Partner(s) contributing:	All partners
Main author(s):	Marco Giovanelli (FINCONS), Gabriel Carraretto (FINCONS), Marcello Paolo Scipioni (FINCONS)
Contributor(s):	All partners



## Summary of modifications

VERSION	DATE	AUTHOR(S)	SUMMARY OF MAIN CHANGES
0.1	16/02/2024	Alessandra Arezza, Marcello Scipioni (FINCONS)	Conclusion of the collection of partners' contributions and first draft of the deliverable
0.2	24/02/2023	Serena Bressan, Riccardo Gallotti, Marco Guerini, Sara Tonelli (FBK)	Addition of further partner contributions and first review of the deliverable
0.3	27/02/2024	Alessandra Arezza, Marcello Scipioni (FINCONS)	Second draft of the deliverable
0.4	27/02/2024	Zoe Fourel (GDI)	Review and quality assurance of the deliverable
0.5	28/02/2024	Alessandra Arezza, Marcello Scipioni (FINCONS)	Final draft of the deliverable
1.0	29/02/2024	Serena Bressan (FBK)	Final review, formatting of the deliverable and submission
1.1	22/05/2024	Andrea Dambrosio, Nicola Bruno, Lara Ponzoni (SKY)	Added Market analysis chapter
1.2	24/05/2024	Marco Giovanelli, Gabriel Carraretto (FINCONS)	Added Overview of AI4TRUST Platform chapter, updated individual exploitation plans
1.3	28/05/2024	Marco Giovanelli, Gabriel Carraretto, Marcello Scipioni (FINCONS)	Revised structure of the document, methodology, and Initial Innovation, exploitation and sustainability strategy chapters



1.4	04/06/2024	All partners	Contributions to Individual Innovation, Exploitation & Sustainability Plan per Asset
1.5	04/06/2024	Marco Giovanelli, Gabriel Carraretto (FINCONS)	Finalisation of the revised deliverable

**Statement of originality** - This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation, or both.

The content represents the views of the author only and is their sole responsibility. The European Commission does not accept any responsibility for use that may be made of the information it contains.



## History of changes from V1.0 to V1.5

- Section 1 is updated according to the added material and the updated document structure
- Section 2 (with its sub-sections) is a newly added section that provides an overview about the market analysis
- Section 3 is revised, providing a more focused description of the methodological steps
- Section 4 (with its sub-sections) is a newly added section that describes the AI4TRUST platform goal, vision, main functionalities, key exploitable results and that embeds and extends the previous “Potential Benefits for the Stakeholders” section
- Section 5 is enriched with a more detailed description of the different type of partners in the AI4TRUST project and the newly added tools
- Section 6 is revised, providing a more detailed description of the Initial Innovation, Exploitation and Sustainability Strategy
- Section 7 is updated according to the added material and the updated document structure



# Table of contents

Summary of modifications.....	3
History of changes from V1.0 to V1.5.....	5
Table of contents.....	6
List of acronyms.....	8
List of figures.....	9
List of tables.....	9
Executive summary.....	10
1. Introduction .....	11
2. Market Analysis .....	12
2.1. Market needs and opportunities.....	15
2.2. Gains and Pains.....	16
2.3. AI4TRUST Market context and competitor analysis .....	17
2.4. Market positioning.....	21
2.5. Value proposition.....	25
3. Methodology.....	27
4. Overview of AI4TRUST Platform.....	31
4.1. Key Exploitable Results.....	31
4.2. Stakeholders and Potential Benefits.....	37
5. Individual Innovation, Exploitation & Sustainability Plan per Asset .....	41
5.1. Fondazione Bruno Kessler (FBK).....	42
5.2. ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS (CERTH) .....	45
5.3. UNIVERSITÀ DEGLI STUDI DI TRENTO (UNITN) .....	48
5.4. NATIONAL CENTER FOR SCIENTIFIC RESEARCH "DEMOKRITOS" (NCSR-D).....	51
5.5. CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS (CNRS).....	55
5.6. UNIVERSITATEA POLITEHNICA DIN BUCUREȘTI (UPB).....	61
5.7. SAHER (EUROPE) .....	64
5.8. GDI GLOBAL DISINFORMATION INDEX GUGHAFTUNGSBESCHRANKT (GDI).....	66
5.9. STOWARZYSZENIE DEMAGOG (DMGG).....	69
5.10. FUNDACION MALDITA.ES CONTRALA DESINFORMACION: PERIODISMOEDUCACION INVESTIGACIONY DATOS EN NUEVOS FORMATOS (MALDITA) .....	72
5.11. ASTIKI MI KERDOSKOPIKI ETAIRIA KENTRO KATAPOLEMISIS TIS PARAPLIROFORISIS / CIVIL NON-PROFIT COMPANY KENTRO KATAPOLEMISIS TIS PARAPLIROFORISIS (ELLINIKA) .....	75
5.12. EURACTIV MEDIA NETWORK B.V. (EURACTIV).....	77



5.13. SKYTG24.....	78
5.14. ASOCIATIA DIGITAL BRIDGE (ADB) .....	80
5.15. EUROPEJSKIE MEDIA SP ZOO (EMS).....	81
5.16. University of Cambridge (UCAM) .....	83
5.17. FINCONS.....	86
6. Initial Innovation, Exploitation and Sustainability Strategy .....	89
6.1. Innovation and Exploitation strategy.....	89
6.2. Sustainability Plan Outline.....	92
6.3. Engagement strategy.....	92
6.4. IPR Management plan.....	94
7. Conclusions .....	96



## List of acronyms

ACRONYMS	MEANING
AI	Artificial Intelligence
API	Application Programming Interface
CIB	Coordinated Inauthentic Behaviour
CSO	Civil Society Organization
DWS	Disinformation Warning System
GDPR	General Data Protection Regulation
IIT	Institute of Informatics and Telecommunications
IP	Intellectual Property
IPR	Intellectual Property Rights
KYC	Know Your Customer
ML	Machine Learning
MVP	Minimum Viable Product
NGO	Non-Governmental Organisation
NLP	Natural Language Processing
SW	Software
UI	User Interface
WP	Work Package



## List of figures

- **Figure 1:** Timeline of Major Initiatives Adopted by the EU Against Disinformation
- **Figure 2:** Proportion of main online news retrieval channel
- **Figure 3:** Weekly proportion of users that accessed online news by going direct to a news website or app
- **Figure 4:** Market segmentation of AI-based technologies for trustworthy solutions against disinformation
- **Figure 5:** Assets macro categories
- **Figure 6:** Functional Architecture
- **Figure 7:** Stakeholders categories

## List of tables

- **Table 1:** Comparative analysis results
- **Table 2:** Interview Structure
- **Table 3:** Key Exploitable Asset mapping
- **Table 4:** Partners' roles



## Executive summary

The current document represents the deliverable "**D7.2 - Innovation, Exploitation and Sustainability Plan v.1**" of the European project "**AI4TRUST - AI-based-technologies for trustworthy solutions against disinformation**" (hereafter also just 'AI4TRUST'). This is the second deliverable (D) of the **Work package (WP) 7 "Communication, Dissemination and Exploitation"** and outlines the first version of a comprehensive strategy for driving innovation, maximising the exploitation of research outcomes, and ensuring long-term sustainability of AI4TRUST solutions within the European context, as detailed in Task (T) 7.2 of WP7.



# 1. Introduction

The main goal of “**T7.2 – Exploitation strategy and innovation management**” is the development of a strategy for creating value from project results and bridging the gap between research and route to market.

This report is the first deliverable describing sustainable outcomes of the AI4TRUST project and it represents **preliminary exploitation paths for exploitable assets**, by focusing on each asset identified and related individual plan. The document presents a first version of the plan, that will be detailed and extended in the next Deliverable D7.4, that will contribute to the exploitation and sustainability of the AI4TRUST offering also beyond its lifespan.

This deliverable is organised with the following structure:

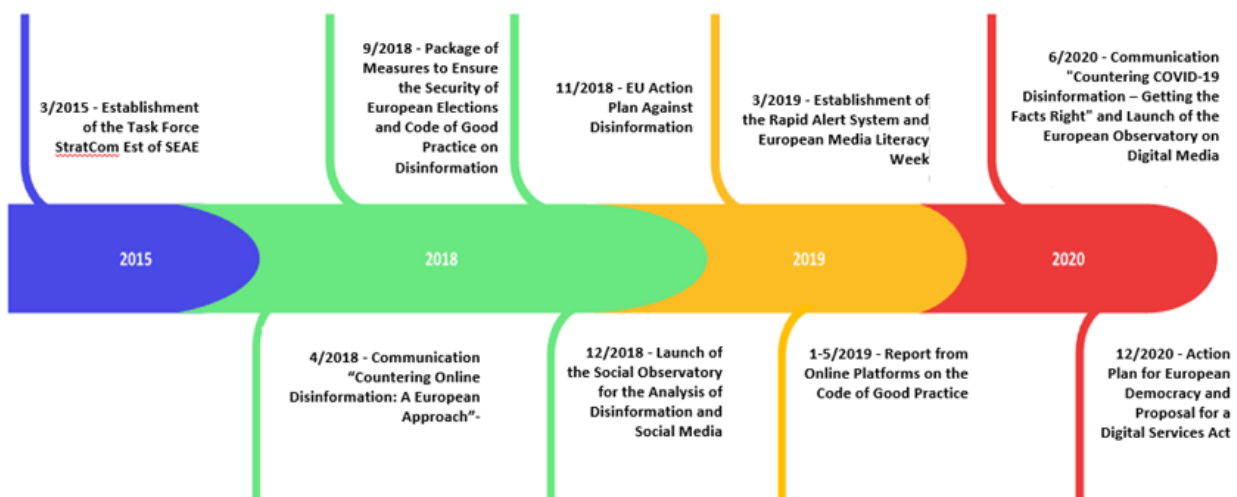
- The **second chapter** focuses on the market analysis and describes the market in which AI4TRUST aims to position itself, highlighting the needs and opportunities, the gains and pains, the context and competitors and the value proposition.
- The **third chapter** describes the methodology used for identifying the main functionalities of the AI4TRUST project, defining the preliminary benefit for the stakeholders, obtaining the individual innovation, exploitation, and sustainability plans for each consortium partner and outlining the initial innovation, exploitation, and sustainability plan.
- The **fourth chapter** provides an overview of the main functionalities of the AI4TRUST platform, defining the key exploitable assets and identifying the stakeholders and their potential benefits.
- The **fifth chapter** aims to describe individual innovation, exploitation, and sustainability plans for each asset of each consortium partner.
- The **sixth chapter** outlines at macro level the preliminary approach and methodology for the initial innovation, exploitation, and sustainability strategy of the whole AI4TRUST platform.
- The **seventh chapter** summarises the main takeaways.



## 2. Market Analysis

This section is dedicated to defining and describing the market in which AI4TRUST aims to position itself by developing the proposed solutions and all the integrated services within the project. The goal of this analysis is to describe the feasibility and potential of the project, which develops a hybrid system where machines cooperate with humans, relying on advanced AI solutions against sophisticated disinformation techniques to support media professionals and policymakers. The system will monitor various online social platforms almost in real-time, filtering out social noise and analysing multimodal content (text, audio, visual) in multiple languages (up to 70% coverage in the EU) with new AI algorithms. It will cooperate in an automated manner with an international network of human fact-checkers who will be periodically triggered and will frequently provide validated data to update the algorithms. The proposal, based on a human-centred approach to technology development and aligned with European social and ethical values, will be integrated into the standard toolbox of analysts working on disinformation.

Disinformation is a delicate issue that has captured the interest of institutions and governmental bodies striving to combat it for the collective good. The EU's commitment to fighting disinformation began as early as 2015 with the development of a strategic communication action plan to counter disinformation campaigns. Over the years, the EU's contribution has continued through actions and funding aimed at developing corrective solutions and mitigating the spread of disinformation, exacerbated by the growing popularity of social networks<sup>1</sup>(see Figure 1).



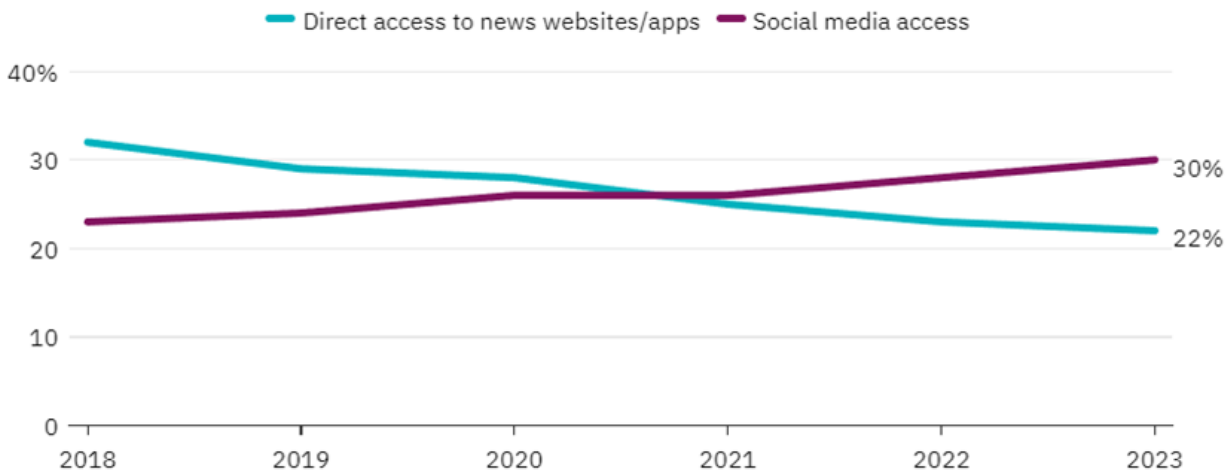
**Figure 1:** Timeline of Major Initiatives Adopted by the EU Against Disinformation

<sup>1</sup> [Special Report 09/2021: Disinformation affecting the EU: tackled but not tamed | European Court of Auditors](#)



To better analyse the concept of disinformation and develop solutions to counter it, it is necessary to consider how and which informational tools are used by users and, respectively, how news publishers have adapted. Over the past decade or more, information has been influenced by tech platforms and other intermediaries in the way news is made accessible. Although search and social media play different roles, news access has been dominated for some time by two major companies: Google and Facebook (now Meta). The growing popularity of digital audio and video is bringing new platforms into play, and in a certain sense, these changes represent a "new normal" in which publishers must navigate an even more complex platform environment, where attention is fragmented, trust is low, and participation is even less open and representative.

Every year, it is noticeable that direct access to apps and websites is becoming less important, while social media is becoming more important due to its ubiquity and convenience. At an aggregate level, we have reached a tipping point in recent years, with the preference for social media (30%) now far exceeding direct access to information (22%).<sup>2</sup>

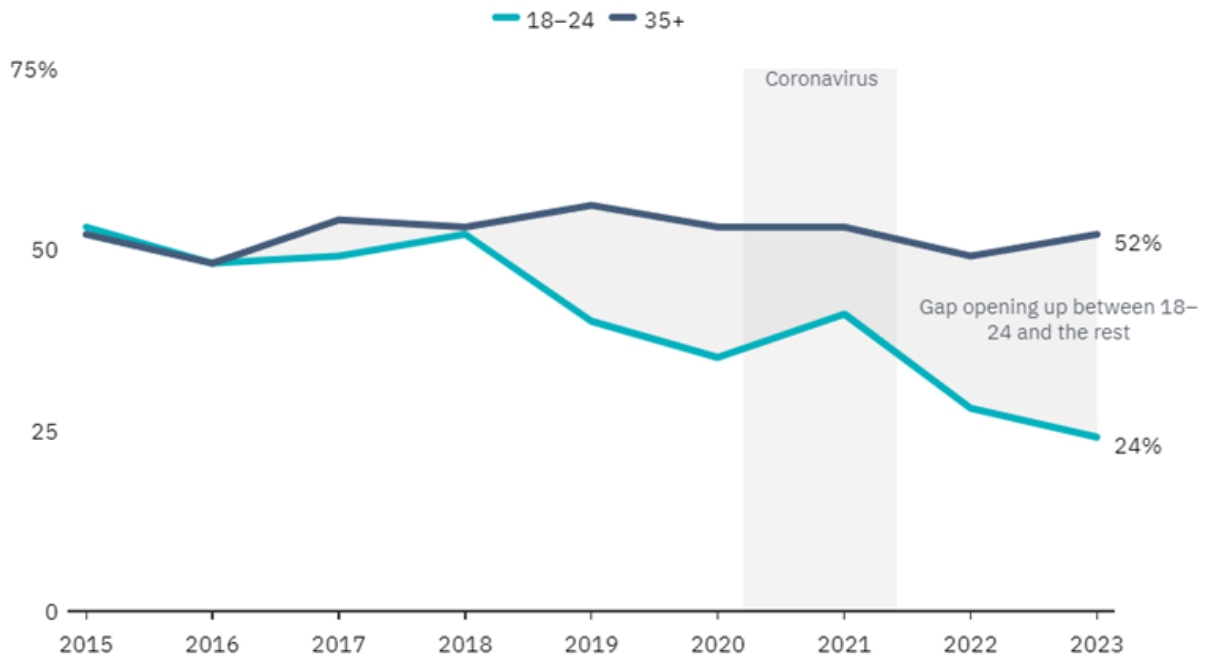


**Figure 2:** Proportion of main online news retrieval channel

Substantial differences are also observed by age group. It is highlighted that younger users are less likely to access a news website or app directly and more likely to use social media or other intermediaries. The annual changes noticed in the comparison between direct and social access seem to be less due to older individuals changing their habits and more due to emerging behaviours of younger groups. The following chart for the UK shows that over 35s have changed their access preferences little over time, while the 18-24 group has become significantly less likely to use a news website or app<sup>3</sup>.

<sup>2</sup> <https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023>

<sup>3</sup> [Digital News Report 2023 | Reuters Institute for the Study of Journalism](#)



**Figure 3:** Weekly proportion of users that accessed online news by going direct to a news website or app

The trends illustrated highlight the future direction emerging regarding the sharing, dissemination, and utilisation of news and the consequent adaptation that information providers must carry out. The increasing superficiality in news research, the spread of social media and social networks, and the conflicts arising from economic interests exacerbate the phenomenon of misinformation and increase the risk of disseminating incorrect information by information providers.

The market for anti-disinformation technology based on advanced human-machine hybrid solutions is an emerging segment that is rapidly gaining importance. These solutions are designed to detect, analyse, and counter the spread of false and misleading information to improve the quality of service offered and the reliability of transmitted information. The main trends characterising the market context are:

- **Increase in Disinformation:** Social platforms and digital media have seen a rise in the spread of fake news, often amplified by bots and algorithms.
- **Emerging Technologies:** Artificial Intelligence (AI) and Machine Learning (ML) are becoming crucial tools for analysing and recognizing disinformation patterns.
- **Multisectoral Collaboration:** Governments, non-governmental organisations, and the private sector are collaborating to develop solutions against disinformation.
- **Growth in Investments:** There is growing interest from investors in technologies that address disinformation, seen as crucial for the sustainability of modern democracies.



## 2.1. Market needs and opportunities

Media disinformation represents one of the most significant challenges of the digital era, with profound impacts on society, politics, and the economy. Citizens, consumers, businesses, and institutions are both authors and beneficiaries of media information and require the continuous evolution of the sector to ensure efficient service and high-quality standards.

The target market, being new and growing, presents multiple opportunities for actors interested in developing technological solutions and human-machine hybrid tools to address the needs and opportunities the market offers.

The drivers of market growth can be classified based on service users and information providers. The main aspects to consider include the proliferation of disinformation, characterised by the exponential increase in false information transmitted through social media and digital platforms, creating an urgent demand for effective solutions to counter the phenomenon. The growing production of material that fuels disinformation has led to increased public awareness of the damage caused by unreliable information, driving the introduction of stricter regulations and the adoption of new anti-disinformation technologies.

Advances in AI and Machine Learning technologies positively stimulate the sector's growth by enabling real-time analysis and accurate detection of false or manipulated content, significantly supporting work currently done manually. It is important to highlight that businesses, governments, and institutions require tools to monitor and protect their online reputation against disinformation campaigns to ensure the integrity of their social status.

The market currently expresses specific needs that include:

- **Accuracy of Information:** Users constantly seek reliable sources and verified information to make informed decisions and guide their social, commercial, and political choices. For instance, citizens may change their habits, and companies may establish their business strategies based on the information circulating globally.
- **Reputation Protection:** Businesses and institutions need tools to monitor and protect their online reputation, filter inappropriate content for dissemination, and identify false or altered news that could impact their operations.
- **Consumer Education:** There is a growing need to educate the public on recognizing disinformation and techniques to avoid it. Providing tools capable of filtering false content is a significant social advantage, which must be accompanied by proper user education, enabling them to independently identify valid and altered information.



- **Regulatory Compliance:** With the increase in regulations on online content, platforms must adhere to increasingly stringent regulations regarding the dissemination of false information. Authorities have already initiated various measures to regulate digital security, privacy, and information, establishing norms and guidelines to protect all user categories.

In recent times, the market has increasingly moved towards multimodal integration, seeking to generate more advanced solutions capable of integrating the analysis of text, audio, and video to improve the accuracy and completeness of disinformation detection. As previously mentioned, human-machine collaboration is essential, and the combined use of AI and human fact-checkers is becoming the norm to ensure that data is accurate and up to date. Two other aspects driving the market recently include the advanced automation of monitoring and analysis processes, which is improving the efficiency and scalability of anti-disinformation solutions, and linguistic expansion, which allows for broader language coverage to tackle disinformation in different regions and languages.

## 2.2. Gains and Pains

Advanced solutions in Artificial Intelligence (AI), Machine Learning (ML), human-machine hybrid solutions, and digital technologies in general address several critical challenges in combating disinformation, but at the same time, they also present numerous growth and development opportunities.

One of the main challenges is the constant evolution of disinformation techniques. Creators of disinformation use increasingly sophisticated methods, such as deep fakes, image manipulation, and automatically generated content, making their detection extremely difficult. This requires AI solutions to be continuously updated and improved to keep pace with these new techniques.

Technological barriers represent another significant obstacle. The accuracy of control and management algorithms heavily depends on the quality and quantity of available data. Without sufficiently representative and high-quality data, the algorithms can produce inaccurate results, compromising the effectiveness of the solutions. Moreover, the need to process real-time data on a large scale poses considerable technical challenges in terms of computing power and data management.

Ethical issues constitute another critical area. The use of AI and automated solutions to monitor online content raises concerns regarding user privacy and the risk of censorship. Balancing the need to combat disinformation with protecting users' rights is a delicate challenge that requires transparent solutions that comply with privacy regulations.



Despite these challenges, there are significant opportunities for human-machine hybrid technologies based on Artificial Intelligence and Machine Learning to combat disinformation. Strategic collaborations represent a promising avenue to enhance the effectiveness and adoption of these technologies. Partnering with social media platforms, academic institutions, and governments can facilitate access to high-quality data, improve algorithm accuracy, and increase the spread of these solutions.

Continuous innovation is essential to maintaining a competitive edge. Investing in research and development allows for the introduction of new features and improvements in solution efficiency. This includes developing more advanced algorithms for multimodal content analysis (text, audio, video), expanding recognized languages, and integrating new technologies such as machine learning.

Finally, emerging markets offer vast growth potential. In many regions, disinformation is particularly pervasive, and the demand for effective solutions is increasing. Expanding presence in these markets can not only help reduce the spread of fake news but also open new commercial opportunities for companies developing anti-disinformation technologies.

Digital technological solutions, specifically those based on the use of Artificial Intelligence to combat disinformation, must face significant challenges related to the evolution of disinformation techniques, technological barriers, and ethical issues. However, there are ample growth opportunities through strategic collaborations, continuous innovation, and expansion into emerging markets. Addressing these challenges with innovative and collaborative approaches is crucial for long-term success in this rapidly evolving sector.

## 2.3. AI4TRUST Market context and competitor analysis

The considered market, being emerging, of interest to various stakeholders, and rich in opportunities, is highly competitive and populated by numerous actors offering different types of solutions. The commercial proposal of AI4TRUST, based on advanced AI solutions against sophisticated disinformation techniques to support media professionals and policymakers, is supported by a comparative analysis conducted in the digital sector and social media management. The analysis has identified 8 tools that provide services similar to those offered by the AI4TRUST project, as we can see in **Table 1**.



**Table 1: Comparative analysis results**

Tool	Features
<p><a href="#">Truly Media</a> (Premium services)</p>	<p><b>Collaborative platform for verifying user-generated content</b></p> <ul style="list-style-type: none"> <li>• Tools for scrutinising text, images, videos for identification of misinformation and fake news</li> <li>• Examination of metadata, source details, timestamps, geolocation, to establish content origin and context</li> <li>• Shared workspaces</li> <li>• Monitoring and verification of content from popular social media platforms</li> <li>• Automated ad-hoc verification workflows</li> <li>• Multilingual support</li> <li>• Resources for training and guidance to enhance their media verification skills and optimise the tool's functionality</li> <li>• API Integration</li> </ul>
<p><a href="#">CrowdTangle</a> (Premium services) (Discontinued after <a href="#">14/08/24</a>)</p>	<p><b>Social media monitoring tool used to track performance and engagement of the content through various social media platforms.</b></p> <ul style="list-style-type: none"> <li>• Real-time social media monitoring based on keywords, hashtags, accounts, etc...</li> <li>• Monitoring of trending content and viral posts on multiple platforms (Facebook, Instagram, Twitter, Reddit)</li> <li>• Aggregated custom dashboard</li> <li>• Historical data (past trends, change of interests over time)</li> <li>• Statistics about engagement metrics such as likes, shares, comments, etc...</li> </ul>
<p><a href="#">Tweetdeck</a><sup>4</sup> (Premium services)</p>	<p><b>Social media management tool for twitter</b></p> <ul style="list-style-type: none"> <li>• Tracking of specific feeds, hashtags, mentions, etc...</li> <li>• Real-time updates</li> <li>• Advanced search and filtering for tracking trends</li> </ul>

<sup>4</sup> Now integrated in [Twitter Pro](#)



Tool	Features
<a href="#">Telemetrio</a> (Free services)	<b>Telegram chats rating based on subscribers and other statistics</b>
<a href="#">NewsWhip Spike</a> (Premium services)	<b>Media monitoring and analytics tool</b> <ul style="list-style-type: none"> <li>● Real-time trend analysis</li> <li>● Cross-platform social media tracking based on topics, keywords, etc...</li> <li>● Predictive insights about future popularity of content and emerging trends</li> <li>● Audience sentiment analysis</li> <li>● Historical data</li> </ul>
<a href="#">Check by Meedan</a> (Premium services)	<b>Tool for collaborative verification of online content</b> <ul style="list-style-type: none"> <li>● Real-time verification within multiple users</li> <li>● Multiple social media platforms integration</li> <li>● Verification workflows and structured check lists</li> <li>● Report generation</li> <li>● Geolocation and timeline analysis for context access</li> <li>● Access to training resources for users to improve verification skills</li> </ul>
<a href="#">Google Fact Check tools</a> (Explorer) (Free services)	<b>Web search engine for fact checked content</b> <ul style="list-style-type: none"> <li>● Debunked stories and images search</li> <li>● ClaimReview markup tool for adding</li> </ul>
<a href="#">Google Trends</a> (Free services)	<b>Web tool for tracking and analysing search query popularity and trends</b> <ul style="list-style-type: none"> <li>● (Almost) real-time topic tracking</li> <li>● Topic comparison</li> <li>● Filtering by language, geolocation, and time</li> </ul>



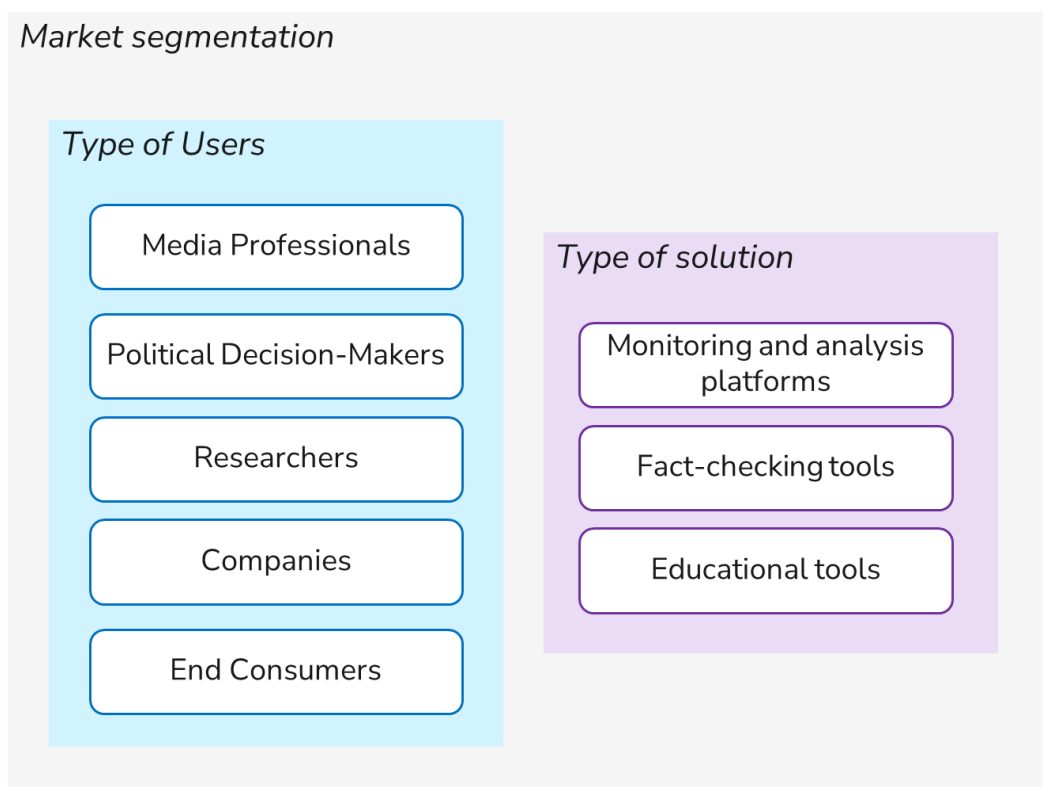
Tool	Features
<p><a href="#">Google API Cloud Vision</a> (Premium services)</p>	<p><b>Tool for extracting insights from image, videos and documents</b></p> <ul style="list-style-type: none"><li>● Image labelling</li><li>● Face recognition</li><li>● Points of interest recognition</li><li>● Optical Character Recognition (OCR)</li><li>● Tagging of explicit content</li><li>● Document categorisation</li><li>● Object tracking</li><li>● Activity recognition</li></ul>
<p><a href="#">Amazon Rekognition</a> (Premium services)</p>	<p><b>Tool for verification of images and videos</b></p> <ul style="list-style-type: none"><li>● Facial verification</li><li>● Face similarity</li><li>● Face attribute recognition</li><li>● Content moderation</li><li>● Custom object detection</li><li>● Text detection</li><li>● Object labelling</li><li>● Video segment detection</li><li>● Celebrity recognition</li></ul>
<p><a href="#">Geolocation</a> (Free services)</p>	<p><b>Tool for geolocation of IP address</b></p> <ul style="list-style-type: none"><li>● IP address Information extraction</li><li>● 2D map visualisation</li><li>● Real-time open source API</li></ul>



Tool	Features
<a href="#">ChatGPT</a> (Free services)	<b>Tool for generative AI</b> <ul style="list-style-type: none"><li>● Conversational capabilities</li><li>● Information retrieval</li><li>● Content creation</li><li>● Image and text analysis</li><li>● Language translation</li><li>● Programming assistance</li><li>● Learning support</li></ul>

## 2.4. Market positioning

The market for advanced Artificial Intelligence (AI) solutions to combat disinformation can be effectively segmented using various criteria, allowing for better identification and targeting of the specific needs of different customer groups. The main segmentation criteria include the type of user and the type of solution offered, as shown in Figure 4.





**Figure 4:** *Market segmentation of AI-based technologies for trustworthy solutions against disinformation*

The types of users can be divided into media professionals, policymakers, companies, researchers, and end consumers:

- **Media Professionals:** The media professionals segment includes journalists working in newsrooms press agencies operating in an environment characterised by a continuous flow of information and constant pressure to publish content quickly. In this context, the need for reliable tools to verify the accuracy of information before publication is crucial. The reputation and credibility of media organisations depend on their ability to provide accurate and timely news. These professionals require rapid and precise fact-checking solutions. The fast-paced nature of their work demands tools capable of handling large volumes of data in real-time, ensuring that information is accurately verified before dissemination. This includes the ability to analyse texts, images, videos, and other forms of multimedia content in multiple languages to identify any falsehoods or manipulations. Additionally, they need to quickly access reliable sources and verified fact databases to support their verification work. Advanced AI solutions offer significant value to media professionals by improving the accuracy and efficiency of the information verification process. These solutions use machine learning algorithms and natural language processing technologies to quickly analyse content and identify potential disinformation. By integrating such tools into their daily workflows, newsrooms can reduce the risk of publishing false news, thus protecting their reputation and maintaining reader trust. Moreover, AI solutions can automate part of the verification process, freeing journalists from repetitive tasks and allowing them to focus on quality journalism. With access to real-time reports and in-depth analyses, media professionals can make informed decisions more quickly, enhancing the timeliness of their publications without compromising accuracy.
- **Political Decision-Makers:** The political decision-makers segment includes government entities and institutions at national and international levels responsible for monitoring and countering disinformation. These entities operate in a complex environment where the spread of false information can have serious implications for national security, political stability, and public trust. The ability to identify and respond quickly to disinformation is crucial for maintaining public order and protecting citizens. Government entities and institutions require advanced tools to monitor information disseminated online and make informed decisions on regulation and crisis management. They need solutions that can analyse large amounts of data from various sources, including social media platforms, news websites, and blogs, to quickly identify potentially harmful information. Additionally, they must be able to assess the impact of disinformation and plan appropriate interventions. The need for real-time analysis and detailed reports is fundamental for the timeliness and effectiveness of their actions. Hybrid AI and Machine Learning-based solutions offer significant value to political decision-makers by enhancing their monitoring and analytical capabilities. These solutions use tailored algorithms to collect and analyse data in real-time,



providing a comprehensive and up-to-date overview of the information landscape. Through the processing of detailed reports, these technologies enable decision-makers to better understand the dynamics of disinformation, identify sources of false information, and assess their dissemination and impact. Advanced analytical capabilities support the development of more effective policies to counter disinformation. Decision-makers can use the data and analyses provided by the solutions proposed by the AI4TRUST project to develop more targeted communication strategies, introduce appropriate regulations, and coordinate rapid and informed responses during crises. Furthermore, the ability to continuously monitor the information environment allows for real-time policy adaptation, enhancing overall resilience against disinformation campaigns.

- **Companies:** The business segment encompasses sectors such as healthcare, finance, and technology, which are particularly vulnerable to reputational damage caused by the spread of false information. These companies operate in highly competitive and regulated markets, where customer and stakeholder trust is crucial for their success. Disinformation can not only compromise corporate reputation but also negatively impact financial performance and regulatory compliance. Companies require advanced tools to monitor brand mentions and information related to their products and services in real-time. They must be able to quickly identify potential threats and effectively manage reputational crises. This necessitates solutions that can analyse large volumes of data from various online sources, such as social media, forums, blogs, and news sites, to detect signals of disinformation and respond promptly. The solutions developed by the project offer significant value to companies by enhancing their monitoring and management of online reputation. These solutions use machine learning algorithms to continuously analyse the information landscape and provide timely alerts about potential threats. With these technologies, companies can receive immediate notifications when false or harmful information emerges, enabling them to intervene quickly to mitigate risks. The detailed analyses provided by the proposed solutions allow companies to better understand the dynamics of online conversations and identify sources of disinformation. This enables the development of more effective communication strategies to counter false news and restore customer and stakeholder trust. Furthermore, AI solutions can help predict potential reputational crises, providing companies with the time needed to prepare appropriate responses and coordinate preventive actions.
- **Researchers:** The researchers segment includes a large network of academic (and related) professionals working on various fields associated with dis/mis/malinformation. This group forms a vital segment within the larger effort to combat the spread of misleading content. Researchers operate in an academic environment, where a shared pool of knowledge and interdisciplinary collaboration is fundamental. The ongoing cycle of evaluation and publication of research material on academic journals and conferences, and the access to large amounts of data from different sources, can contribute substantially to the improvement of research related to fake news and mis/disinformation. Also, it aids the development of state-of-the-art solutions to tackle the spreading of this problem.



- **End Consumers:** The end consumer segment encompasses a wide range of users, including social media users, students, educators, and the public. In an era characterised by an incessant flow of information through various digital channels, these groups are frequently exposed to disinformation and manipulated content. The ability to distinguish between true and false information has become an essential skill for navigating the contemporary information landscape. End consumers require tools that help them improve their digital literacy, i.e., the ability to critically evaluate online information and recognize fake news. These tools must be accessible and easy to use, providing support in identifying and verifying information. Advanced AI solutions to address misinformation offer significant value to end consumers through user-friendly applications that educate the public and offer simple methods for fact-checking. These applications may include features such as automatic fact-checking, alerts of potential misinformation, and guides on how to assess source credibility. By integrating these technologies into everyday life, users can become more informed and resilient to misinformation. Additionally, AI solutions can analyse misinformation patterns and provide useful data for public awareness campaigns, enhancing individuals' ability to distinguish between true and false information.

Market segmentation based on the type of solution offered allows companies to develop and market targeted products that meet the specific needs of different segments, improving the overall effectiveness of the solutions. The solutions currently available on the market include:

- **Monitoring and Analysis Platforms:** Systems designed to monitor social platforms in real-time and analyse content to identify disinformation. Solutions include continuous real-time monitoring, data analysis with advanced algorithms, and visualisation tools to quickly identify disinformation trends. These platforms are typically ideal for media professionals, policymakers, and businesses requiring constant and detailed monitoring. They can assist journalists in fact-checking news, policymakers in managing disinformation crises, and businesses in protecting their online reputation.
- **Fact-Checking Tools:** Solutions designed to assist human fact-checkers in identifying and validating false information. These tools are characterised by AI algorithms for natural language processing, databases of verified facts, and integration with news platforms. They are essential for journalists and newsrooms needing to quickly verify information before publication. They are also useful for governmental entities and non-governmental organisations working to combat disinformation on a large scale.
- **Educational Tools:** Designed to educate the public on recognizing disinformation and on the importance of digital literacy, these tools include interactive educational content, teaching materials, and resources to increase users' awareness and critical thinking skills. They are generally aimed primarily at end consumers who can benefit most from learning the skills necessary to navigate the contemporary information landscape.

To effectively position ourselves in the sector described in the preceding paragraphs, it is necessary to address market needs and meet consumer demands. From the analyses conducted, it emerges



that proper positioning is based on technological innovation, ease of use, reliability, accuracy, as well as strategic collaborations. These combined elements allow us to offer advanced and effective solutions to combat disinformation and promote a more reliable and secure information ecosystem.

To this end, the project aims to distinguish itself through a series of key factors that address market needs and challenges of digital disinformation. Firstly, Technological Innovation is identified; the project is committed to offering cutting-edge Artificial Intelligence (AI) solutions for identifying and mitigating disinformation by filtering social noise and analysing multimodal content. Using advanced algorithms and machine learning technologies, the goal is to provide state-of-the-art tools capable of quickly analysing large volumes of data and accurately identifying false and manipulated content. Another aspect to consider is ease of use; therefore, the project is committed to providing intuitive and accessible tools for a wide range of users. However, the priority is to ensure the reliability and accuracy of the solutions. The project aims to provide high precision in detecting disinformation, using sophisticated algorithms and credible verification sources. We collaborate with experts in the field of fact-checking and information sciences to ensure that the solutions are reliable and comply with the highest standards. Finally, the commitment of the AI4TRUST Project is to create a partnership of strategic collaborations to combat disinformation and cooperate to generate an effective and efficient tool.

## 2.5. Value proposition

The project's value proposition revolves around several key points, each aimed at providing advanced and effective solutions to counter disinformation and promote a more reliable and secure information ecosystem. By utilising Artificial Intelligence algorithms, the system can quickly identify and flag potentially false content, enabling a prompt response from users and digital platforms. The ability to detect and verify information in real-time is crucial for combating the rapid spread of online disinformation.

Through Machine Learning techniques, the system can predict disinformation trends and assist digital platforms in preventing the dissemination of false content before it goes viral. This predictive analysis capability allows for effective preventive measures to be taken and enables anticipation of the strategies of malicious actors.

The technology can be easily integrated with other digital platforms and scaled to handle large volumes of data and users. This flexibility and scalability allow organisations to tailor the solution to their specific needs and expand its usage as they grow and evolve over time.

In summary, the value proposition is based on offering comprehensive and innovative solutions to counter disinformation, including real-time detection and verification, predictive analysis, education



and awareness, as well as integration and scalability. These combined elements enable us to offer a holistic and effective approach to address the challenges of disinformation in the digital age.



### 3. Methodology

The overall objective of the “T7.2 – Exploitation strategy and innovation management” is to guarantee a **comprehensive and strategic approach to innovation, exploitation, and sustainability**. To achieve this objective, the **methodology** employed during the initial phase of the activity has outlined four main phases:

- **Overall overview and key exploitable results of the AI4TRUST Project.** In this activity, the **main functionalities** of the AI4TRUST platform are identified and grouped to obtain a clear overview of the platform capabilities. This has involved multiple partners in the definition of the specific functionalities and in their harmonisation in order to provide the end-users with a consistent and effective solution for their specific needs. This activity has been fundamental for sharpening the goal and the vision of the AI4TRUST platform, but also for the **mapping between the described functionalities and the key exploitable results**, that have been identified and analysed in the subsequent activities. The results of this activity can be found in the “Overview of AI4TRUST Platform” section (sec. 4) and “Key Exploitable Results” section (sec. 4.1).
- **Preliminary Benefits for Stakeholders of the AI4TRUST Project.** In this activity, the **preliminary benefits** that stakeholders can expect from the AI4TRUST project in terms of innovation, exploitation, and sustainability, were evaluated and articulated, highlighting specific advantages and outcomes that will result from the successful implementation of the project. Through this activity, it is possible to have **visibility on specific advantages** that will arise from the successful execution of the project in terms of creation/innovation of assets. The results of this activity can be found in the “Stakeholders and Potential Benefits” section (sec. 4.2).
- **Individual Innovation, Exploitation, and Sustainability Plan for Each Asset of Each Partner.** This activity aims to **classify the individual assets based on relevance to the AI4TRUST project** with potential impact on innovation, exploitation, and sustainability. The work carried out was based on a comprehensive evaluation process through the **consultation of all consortium partners** for discerning and documenting the specific assets of interest for each of them. This detailed assessment aimed to identify the diverse resources and capabilities of each entity involved in the project. Thanks to those interviews, the activity has highlighted not only **tangible assets** but also **intangible knowledge**, expertise, and unique contributions that can significantly impact the success of the overall exploitation strategy approach, by underlying strategic advantages that can be leveraged for optimal project outcomes. This **interview-based assessment** represents a dynamic tool that will follow the project lifecycle and will enable a continuous and adaptive optimization of resources and capabilities for the overall success of project outcomes. The following Table 2 represents the interview structure and information requested to partners. Assets



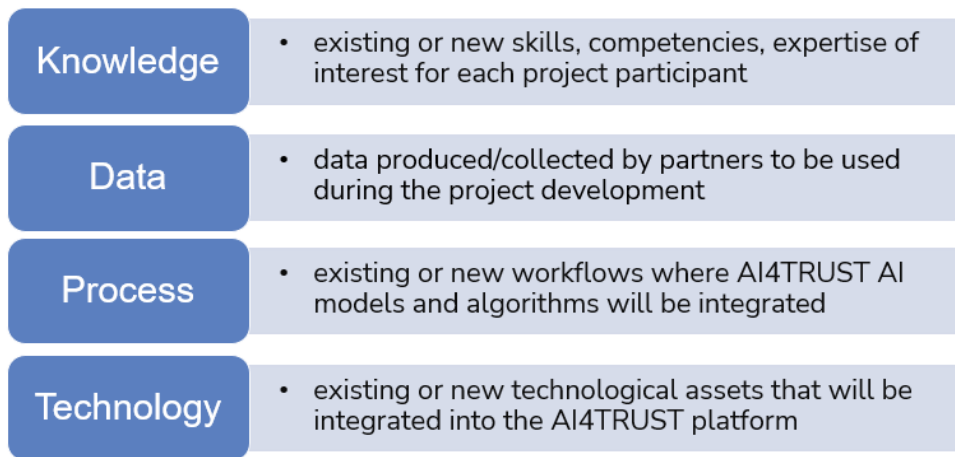
identified by interviews with consortium partners have been classified in four main categories (see Figure 5).

*Table 2: Interview Structure*

<b>QUESTIONS</b>	
	<p><b>1. Exploitable assets and results:</b> Describe what assets you intend to exploit (whether this involves specific components, tools, knowledge, methodologies, skills, etc.).</p> <p><b>ASSET 1:</b></p> <ul style="list-style-type: none"> <li>● Description</li> <li>● Target Groups / beneficiaries</li> <li>● Innovation, Exploitation &amp; Sustainability Plan: (explain innovation outcomes from the identified asset in terms of results or knowledge that could be exploited and describe the timeline plan you have for using the asset. Provide concrete actions)</li> </ul> <p>...</p> <p><b>ASSET N:</b></p> <ul style="list-style-type: none"> <li>● Description</li> <li>● Target Groups / beneficiaries</li> <li>● Innovation, Exploitation &amp; Sustainability Plan: (explain innovation outcomes from the identified asset in terms of results or knowledge that could be exploited and describe the timeline plan you have for using the asset. Provide concrete actions)</li> </ul>
	<p><b>2. Partner expertise toward assets innovation:</b> Explain in detail how your skills align with the previously mentioned assets and articulate how your expertise will drive the innovation of project outcomes associated with those assets.</p>
	<p><b>3. Rationale:</b> Explanation of why you are interested on those assets (the added-value they provide), how do you plan to exploit them (academically or industrially: e.g., provide as commercial solution, certification services, standardisation, consultancy, further R&amp;D, positioning).</p>
	<p><b>4. Opportunity which appeared/appears:</b> your participation is the result of the real need of your customers (for industrial partners) or internal needs (for user partners). For academic partners, mention if AI4TRUST is in line with other projects (continuation) and reuse of know-how. Are there any other opportunities in the pipeline when the project finishes?</p>



QUESTIONS	
	<p><b>5. Your Value Proposition towards Joint Exploitation of AI4TRUST outcomes:</b> what do you expect from project partners, what benefits will you deliver to the rest of the consortium, what components/interest do you share with other partners.</p>



*Figure 5: Assets macro categories*

After the identification of the individual assets, each partner was asked for **customised plans for innovation, exploitation, and sustainability for each identified asset**, to maximise the potential of each asset within the project consortium resources. This outlined strategies for enhancing innovation, maximising exploitation potential, and ensuring long-term sustainability for each asset. The purpose is to understand the unique characteristics of each asset, involving not only identifying the technological advancements but also recognizing the innovative potential inherent in each asset. The potential exploitation of assets needs to involve strategically capitalising on the identified assets **to maximise their potential values**. This entails **drawing a pathway and a go-to market strategy**, underlining **opportunities**, and establishing **strategic partnerships and business models**, starting from the exploitation strategy of each asset. Finally, the integration of sustainability into the plans is **crucial for long-term viability**. Sustainability considerations go beyond the immediate financial gains and require a **holistic approach** ensuring that the project outcomes will contribute to a sustainable and ethically responsible future. The results of this activity can be found in the “Individual Innovation, Exploitation & Sustainability Plan per Asset” section (sec. 5).



- **Initial Innovation, Exploitation and Sustainability Strategy.** This activity gives a preliminary overview of the **initial innovation, exploitation and sustainability plan**. The objective is to formulate a **comprehensive macro strategy** that integrates innovation, exploitation, and sustainability efforts across all assets and partners, taking into account the different segments that stakeholders belong to, to ensure exploring various market segments and needs and different possible business models. An engagement strategy has been devised to collect feedback from internal and external stakeholders, fostering **synergies and opportunities for collaboration**, and promoting the results obtained within the project, enhancing its expected impacts. Moreover, an Intellectual Property Rights (IPR) management plan is presented to ensure proper management of intellectual property rights related with the development of platform components. The envisioned strategy is presented in the “Initial Innovation, Exploitation and Sustainability Strategy” section (sec. 6).



## 4. Overview of AI4TRUST Platform

The goal of AI4TRUST is to tackle mis/disinformation with **human validated content items** and **state-of-the-art tools** for monitoring, debunking and obtaining analytics. This is offered through a cross-country joint effort to build a platform that is trustworthy, ethical, respectful of data ownership and of user's privacy in adherence with EU values and guidelines. In particular, the AI4TRUST platform will empower: (i) media practitioners in debunking and monitoring of news items; (ii) policy makers with analytics and insights; (iii) researchers in data mining and development of AI models.

This goal is well coupled with **the AI4TRUST vision** of extending and improving the fact-checking activities, supporting the definition of rules and counter-measures, and bringing generative AI knowledge to the next frontier, with the final overall goal of **defending EU citizens from mis/disinformation and manipulation**.

To this end, the following sections provide an overview of the functionalities offered by the AI4TRUST platform, the key assets selected for exploitation and the stakeholders and their potential benefits identified at this stage of the project.

### 4.1. Key Exploitable Results

The AI4TRUST platform envisages multiple assets of which the ones with higher potential to be exploited can be grouped in the following functional areas (see Figure 6):

- **Data Acquisition and Normalisation**, for the assets responsible for the automatic acquisition of single news items from different sources and the collection of the respective human validations;
- **Analysis and Monitoring of Single News Items**, for the assets responsible for the textual, visual and speech analysis of single news items and the subsequent automatic ranking according to the manipulation risk evaluated by the Disinformation Warning System;
- **Collective Analysis of Social Media Actors and Items**, for the assets responsible for higher-level analysis across the multiple news items collected.

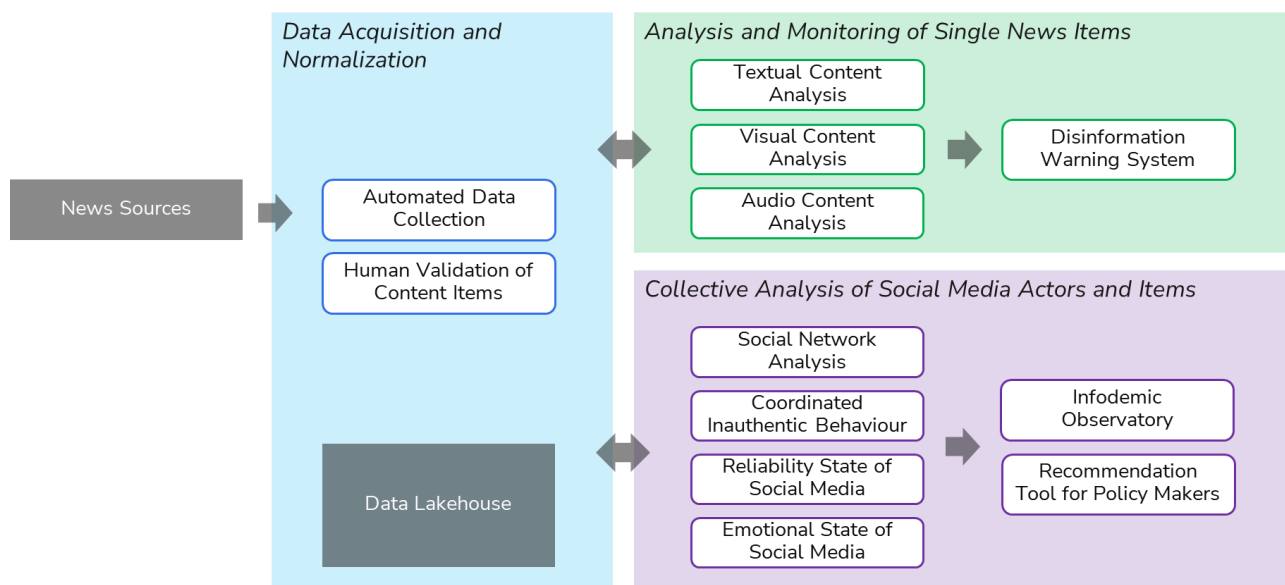


Figure 6: Functional Architecture

For what concerns the **Data Acquisition and Normalisation**, the main functionalities are:

- **Automated Data Collection**, which obtains single news items from the envisaged social media (i.e., YouTube and Telegram);
- **Human Validation of Content Items**, which enable fact-checkers to provide manual characterisation of analysed single news items.

Data acquisition assets are key to the whole platform, as all tools in the AI4TRUST platform are data-driven. As a matter of fact, analysis tools focusing on single media items and aggregated analysis tools all depend on the data that have been collected by data acquisition tools.

The automated data collection tool is configured in AI4TRUST to collect relevant data within three well-defined topics, as the amount of data in input to the platform has to be limited due to computational and storage constraints. However, this filtering is parametric and can be tuned according to specific market needs, which makes the AI4trust platform versatile to be potentially used in the future for topics and in contexts different from those identified within the project.

**Automated data collection** plays a crucial role for media professionals, political decision-makers and businesses which require an analysis over large amounts of data from diverse sources, formats, and languages. **Human validation** ensures the reliability and accuracy of these data sources, enabling journalists, newsrooms and other media users to sustain their verification processes effectively.

In relation to the **Analysis and Monitoring of Single News Items**, the main functionalities are:



- **Textual, Visual and Speech analysis**, which consist of a series of modules that enable the user to:
  - **detect disinformation signals**, segmenting the given input text with associated labels (i.e., “hate speech”, “clickbait”, “offensive”) and providing a confidence score;
  - **detect the check-worthiness of a claim**, labelling the given input text as “check-worthy” (i.e., factual and verifiable text that appears to be false, may be of public interest or of impact to the public, or may cause harm to the society, entities, groups, or individuals) and providing a confidence score;
  - **retrieve previously fact-checked claims**, searching the given input text amongst archived fact-checked claims and providing a similarity score;
  - **detect image deepfakes**, reporting the probability scores of different manipulation techniques for the given image;
  - **detect video deepfakes**, reporting the probability scores of different manipulation techniques for the given video;
  - **reverse search videos on the web**, extracting the most relevant frames from the given video, automatically performing image-based search on the web and returning a list of visually-matching online sources and similar videos;
  - **detect audio deepfakes**, reporting the probability scores of manipulation for the given audio;
  - **detect audio anomalies**, providing a list of splicing points with the relative confidence scores;
  - **transcribe speech to text**, enabling textual analysis on the given audio content.
- **Disinformation Warning System**, which automatically labels single news items with a disinformation risk score, based on the results of the Textual, Visual and Speech analysis.

All these functionalities allow the evaluation of single content items in a multimodal way, acquiring automated indications regarding specific characteristics. This is especially useful for fact checkers and media professionals, who need to daily assess large quantities of content items in a context in which disinformation is growing in quantity and becoming more and more sophisticated, given the availability of AI-based tools that allow producing increasingly convincing disinformation content.

Monitoring features, like the **Disinformation Warning System**, are aimed at assisting political decision-makers, media experts and consumers in overseeing online information dissemination, which involves understanding disinformation patterns, assessing their social impact, and identifying their sources.

Companies can benefit from monitoring as well. By tracking data related to brands and products, businesses can deepen their understanding of social interactions, allowing them to promptly mitigate potential reputational-related threats.



Collectively, **textual, visual and speech analysis** can be implied by all customer segments, giving means to rapidly validating content accuracy.

As regards the **Collective Analysis of Social Media Actors and Items**, the main functionalities are:

- **Social Network analysis**, which evaluates how much hierarchical, unequal or biased the social network is.
- **Coordinated Inauthentic Behaviour**, which allows to explore coordination dynamics around same/similar pieces of content across different digital spaces by checking simultaneously the status of specific pieces of contents within one platform and the status of same/similar content for how they are present on different platforms.
- **Reliability State of Social Media**, which provides the end-user with synthetical quantitative indices of the disinformation risk presented in given areas of social media, to bring attention to areas of interest where the risk of disinformation flowing is higher;
- **Emotional state of Social Media**, which provides a simple measure of the share of endorsement/rejection of given pieces of content, which can also be computed at different levels (e.g., user, channel, etc.);
- **Infodemic observatory**, a map-based tool which provides disinformation statistics and indices, such as the number of unreliable news circulating for a given country, the social media volume or the risk index in a given country, regarding a certain topic, and in a given timeframe,
- **Recommendation tool for policy makers**, which draws on different functionalities of the AI4TRUST platform, namely the detection and analysis of disinformation signals (e.g. claim validity and Social Network Analysis), to link aggregated inputs from the AI4TRUST platform with a classification of their severity level and a guide towards mitigation measures. The hybrid character is conferred by both the human-centred design of the tool and the human supervision of the recommendation inputs, particularly when dealing with disinformation classified as a systemic risk.

The evaluation and **analysis of items and social media actors** are directly related to the needs of political decision makers and companies. Through these processes, coupled with monitoring, users can acquire analytical capabilities, access source reliability, and anticipate potential threats.

Each of these functionalities has been mapped to specific Key Exploitable Assets, as shown in Table 3, and each asset is described in detail with its “Individual Innovation, Exploitation & Sustainability Plan” in the following sections (from sec. 5.3 to sec. 5.19), according to the involved partner.



**Table 3: Key Exploitable Asset mapping**

Functionality	Key Exploitable Asset	Partner	Type of Asset	Individual Innovation, Exploitation & Sustainability Plan
Human Validation of Content Items	Human Validation Tool	FINCONS	Foreground	See sec. 5.17
Automated Data Collection	Social listening data stream	FBK	Foreground	See sec. 5.1
Textual Content Analysis	AI model for check worthiness of textual claims	FBK	Foreground	See sec. 5.1
Textual Content Analysis	AI model for retrieval of previously fact-checked claims	FBK	Foreground	See sec. 5.1
Textual Content Analysis	Document intelligence – Technology	NCRS-D	Foreground	See sec. 5.4
Textual/Visual Content Analysis	AI model for visual-text misalignment detection	CERTH	Foreground	See sec. 5.2
Visual Content Analysis	Tool for reverse video search on the Web	CERTH	Background	See sec. 5.2
Visual Content Analysis	Deepfake image/video detection	CERTH	Background	See sec. 5.2



Functionality	Key Exploitable Asset	Partner	Type of Asset	Individual Innovation, Exploitation & Sustainability Plan
Visual Content Analysis	AI models for sensational content detection	CERTH	Foreground	See sec. 5.2
Visual Content Analysis (no final user)	Tool for video anomaly detection	UNITN	Foreground	See sec. 5.3
Speech Content Analysis	Speech-to-text technology and web service	UPB	Foreground	See sec. 5.6
Speech Content Analysis	Audio deep fake detection technology and web service	UPB	Foreground	See sec. 5.6
Speech Content Analysis	Audio anomaly detection technology and web service	UPB	Foreground	See sec. 5.6
Disinformation warning system	Disinformation warning system (DWS)	GDI	Foreground	See sec. 5.8
Social Network Analysis	Social Network Analysis Tool	CNRS	Foreground	See sec. 5.5
Coordinated Inauthentic Behaviour	Coordinated Inauthentic Behaviour Tool	CNRS	Foreground	See sec. 5.5



Functionality	Key Exploitable Asset	Partner	Type of Asset	Individual Innovation, Exploitation & Sustainability Plan
Reliability State of Social Media	Reliability State of Social Media Tool	CNRS	Foreground	See sec. 5.5
Emotional State of Social Media	Emotional State of Social Media Tool	CNRS	Foreground	See sec. 5.5
Infodemic Observatory	Infodemic Observatory Tool	CNRS	Foreground	See sec. 5.5
Recommendation Tool for Policy Makers	Hybrid Recommendation Tool	UCAM	Foreground	See sec. 5.16

## 4.2. Stakeholders and Potential Benefits

Starting from the analysis of individual assets per partner, this chapter aims **to identify potential benefits for the project stakeholders**, current and future. Based on the interviews with project partners and on the market analysis, different groups of stakeholders emerged, as described in Figure 7.



**Figure 7: Stakeholders categories**

To summarise, the current analysis presents a **range of potential benefits for diverse stakeholders**:

- **Media professionals** will be supported in **tackling the overwhelming volumes of mis/dis/malinformation** content that circulates online, detecting and **monitoring contents, understanding the impact that innovative AI tools** could have in their workflow and **boosting their knowledge**. This will be possible through the set of tools provided by the AI4TRUST platform. In particular, the **Analysis and Monitoring of Single News Items**, in which the **Disinformation Warning System** will be able to automatically label news items according to the disinformation risk, will provide media professionals with an additional way to monitor and filter the contents that need to be manually verified, and support them in this process. The **Textual/Visual and Speech Analysis** of single news items will also empower the media professionals with a set of tools that can improve their workflow and speed up their activities while maintaining high qualitative standards, thanks to the enhanced capabilities of handling multimedia content and providing detailed analysis about tampering and misuse.
- **Political Decision-Makers** will benefit from the AI4TRUST platform: a **streamlined information verification** process, the access to **reliable information**, novel avenues for



leveraging **AI algorithms** and the possibility to **explore practices** of other European organisations in tackling mis/dis/malinformation. This will be possible through the **Collective Analysis of Social Media Actors and Items** offered by the AI4TRUST platform, that will enable Political Decision-Makers to evaluate the social network through the **Social Network Analysis**, to assess the **Reliability and Emotional State of Social Media**, to verify the presence of **Coordinated Inauthentic Behaviours**, to analyse geographically grouped disinformation statistics and indices through the **Infodemic Observatory** and to obtain semi-automated reports with guidelines to mitigate the spread of disinformation through the **Recommendation Tool**.

- **Researchers**, through the AI4TRUST platform, will have the possibility to access the **human validated results** and the **dataset of collected news items** from the social listening data streams. Moreover, they will be involved in an **interdisciplinary context** where they could access a **shared pool of knowledge** and **collaborate with industry partners**. This will create the perfect environment to enhance the network between researchers, experiment and generate new advanced technological solutions, improve the capacity to deliver impactful outcomes and to publish the results in relevant journals and conferences, influencing the forthcoming project and funding opportunities.
- **Companies** will be able to leverage the automated collection of data through the **social listening data stream**, and the **infodemic observatory tool** to monitor consumer opinions and concerns regarding their products, brands, and overall reputation. Furthermore, **emotional state of social media tool** and **sensational content detection** will be useful for monitoring content endorsement. By utilising sentiment analysis and content classification, companies will gain insights into users' emotional responses to specific content. Fake news and mis/disinformation related to a specific product or the company will be addressed with a wide range of textual, visual and speech content analysis tools. Functionalities such as **reverse video search on the web** and **visual-text misalignment detection** will help to debunk false associations between videos and products which may lead to defamation and harm to the company's image. Additionally, **deepfake image, video and audio detection** will uncover attempts to discredit testimonials, company representatives and individuals associated with a brand or product.
- **End consumers** will have extensive access to automatic fact-checking tools to verify the authenticity of online content. Tools such as **deepfake image, audio, and speech detection, document intelligence, visual-text misalignment detection, reverse video search on the web** and others will enable users to distinguish between authentic and false information, especially on social media. The use of these fact-checking tools, combined with access to information on the **reliability and emotional state of social media**, and the ability to monitor the spread of misinformation and disinformation through the **infodemic observatory tool**, will enhance digital knowledge, educate users about potential concerns, and raise awareness about the reliability of online content. The **retrieval of previously fact-checked**



**claims** will enable users to directly verify if a given claim has already been debunked or if there are any related debunking articles for a particular topic. By providing a user-friendly introduction to fact-checking, users will stay informed about the latest topics related to mis/disinformation, capacitating and improving the ability to raise consciousness on these issues.



## 5. Individual Innovation, Exploitation & Sustainability Plan per Asset

This section of the document describes each above-mentioned asset and will explain **how the Consortium partner is planning to exploit the AI4TRUST results**. The information has been gathered thanks to interviews internally to the consortium to understand the assets to exploit and an initial plan to drive the route to the market.

Each partner of the AI4TRUST project has a different role in the valorisation of the project results, based on the asset owned and the mission. Three groups have been defined according to the role:

- **Scientific partners** ensure knowledge generation and capacity of providing technologies beyond the state-of-the-art.
- **Business partners** develop and commercially exploit technological solutions based on project results.
- **End-user partners** benefit from the adoption of project results in their daily operations and duties.

The assignment of each partner to these categories is reported in the table below (Table 4). As shown, different partners have multiple roles within the project, highlighting the opportunity to exploit the AI4TRUST platform from different perspectives, not only the business one.

**Table 4: Partners' roles**

Scientific	Business	End-user
<ul style="list-style-type: none"> <li>● FBK</li> <li>● CERTH</li> <li>● UNITN</li> <li>● NCSR - D</li> <li>● CNRS</li> <li>● UPB</li> <li>● UCAM</li> </ul>	<ul style="list-style-type: none"> <li>● FIN</li> <li>● GDI</li> <li>● MALDITA</li> <li>● SAHER</li> </ul>	<ul style="list-style-type: none"> <li>● GDI</li> <li>● DEMAGOG</li> <li>● MALDITA</li> <li>● ELLINIKA</li> <li>● EURACTIV</li> <li>● SKYTG24</li> <li>● ADB</li> <li>● EMS</li> </ul>



## 5.1. Fondazione Bruno Kessler (FBK)

**Fondazione Bruno Kessler (FBK)** is bringing to this project an extensive set of experience in the field of **Data Engineering, Data Science, Social Network Analysis (SNA), Natural Language Processing (NLP), and the development of Large Language Models (LLMs)**.

The AI4TRUST platform builds upon the **experience of the Covid19 Infodemics Observatory<sup>5</sup>** and other important technological **assets developed by FBK's Digital Society Centre**. FBK aims to widely expand from this early experience creating a **new social listening data stream platform** for automated collection and processing of social and news media, compliant with EU-GDPR and other related European and national legislation, and tailored to the needs of scientific research in the field of mis/disinformation.

FBK also develops **AI models for assessing the check-worthiness of textual claims** in multiple languages, and for retrieving previously fact-checked claims across languages. Additionally, FBK pioneers AI models for generating verdicts on claim veracity, addressing pragmatic aspects like style and emotion.

Finally, the research experience and expertise that will be gained thanks to the dynamic partnerships within AI4TRUST with media and fact-checking organisations not only enhances FBK capacity for **technology transfer** but also fuels its ability to continually generate **fresh research ideas**, driving innovation and keeping this institute at the forefront of the fields of Artificial Intelligence (AI) and the study of Online Communication, strengthening its position in future projects.

Asset	<b>Social listening data stream</b>
Type	Technology/Data
Description	FBK will develop a social listening platform that automatically collects and processes social media and news from multiple sources. This platform will elaborate on the raw data, transforming them into GDPR-compliant data while capturing text, shared audiovisual content, and social interactions, making it suitable for scientific use.
Target Groups/Beneficiaries	Researchers
Innovation, Exploitation and Sustainability Plan	The Data Stream platform will be an asset also after the end of the AI4TRUST project, as it could continue to generate relevant

<sup>5</sup> <https://covid19obs.fbk.eu>



Asset	<b>Social listening data stream</b>
	<p>data for scientific research and could be further adapted and expanded to cover other data sources.</p> <p>The data collected will inform the scientific work of the AI4TRUST consortium. Post project, part of the data collected will be, in aggregated form, shared with the whole scientific community to ensure further future research developments.</p>

Asset	<b>AI model for check worthiness of textual claims</b>
Type	Knowledge/Technology
Description	<p>FBK will develop an AI model that is able to assess whether a textual claim is check-worthy or not in different languages. The model will be trained either on existing datasets for check-worthiness already available to the Natural Language Processing (NLP) community as well as on a dataset that has been specifically created during AI4TRUST for the Italian language.</p>
Target Groups/Beneficiaries	Researchers and professionals
Innovation, Exploitation and Sustainability Plan	<p>In AI4TRUST, FBK will build on (monolingual and multilingual) LLMs and fine-tune them both on existing datasets and on the Italian data that FBK has specifically created for the task. Multilingual solutions will also be explored and evaluated to cover as many languages as possible. The trained models will be exposed through APIs and will allow researchers and professionals to filter claims that may be worth checking for veracity, before further processing them. The trained model will be made available too for further reuse and improvement both from professional fact-checkers and researchers.</p>

Asset	<b>AI model for retrieval of previously fact-checked claims</b>
Type	Technology



Description	FBK will develop an AI model that, given a claim, provides in output a ranked list of previously fact-checked claims that are like the one of interest. The tool can work cross-lingually and can be adapted to include previously fact-checked claims of professional fact-checkers.
Target Groups/Beneficiaries	Researchers and professionals
Innovation, Exploitation and Sustainability Plan	In AI4TRUST, FBK will build on works related to previously fact-checked claim retrieval and extend them to make them more robust and accurate, also cross-lingually. The source code of the application will be released, so that fact-checking companies will have the possibility to use the tool beyond the project ends and add their previously fact-checked claims to the pool of existing ones on their own servers.

Asset	<b>AI models for verdict generation</b>
Type	Knowledge/Technology
Description	A set of trained Language models for generating short texts that discuss the veracity of a claim provided a fact-checking article connected to the claim. The input of these models is textual (the claim and the debunking article concatenated together) while the output is (1) a text that can be used by stakeholders for performing social correction on social media platforms or similar tasks and (2) a list of the most relevant sentences from the input article. These language models are meant to address not only the veracity of the claim but also to consider pragmatic aspects such as style and emotions, that are crucial in an effective communication, especially on social media platforms.
Target Groups/Beneficiaries	Fact-checkers/journalists, citizens.  In general, whoever needs to verify discusses the veracity of a claim from an argumentative point of view and not only from a classification perspective.



<p>Innovation, Exploitation and Sustainability Plan</p>	<p>In AI4TRUST, FBK will build on existing LLMs that are already in place and fine-tune them on data that FBK has specifically created for the task. The trained models will be exposed through API and run internally to the platform, to provide suggestions to claims to which the veracity needs to be assessed by the users.</p>
---	---

## 5.2. ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS (CERTH)

The CERTH team has substantial experience in the field of **media analysis and verification**, and more specifically on methods for:

- video fragmentation and keyframe selection;
- detection of deepfake and AI synthesised content;
- visual content classification;
- multimodal representation learning.

CERTH’s research experience and expertise in the area in combination with the technology requirements by the fact-checking and media organisations of AI4TRUST will help us drive new advances in our existing solutions and potentially lead to new methods that considerably **improve the performance of our methods and services**.

Asset	<b>Tool for reverse video search on the Web</b>
Type	Technology
Description	<p>A web-based technology that allows a user to temporally segment a video into visually-coherent fragments, extract a set of representative keyframes and use them to perform keyframe-based search for this video on the Web. This search can be made through various search engines (Google, Yandex, Bing, Baidu) and image search tools (TinEye). The detection of one or more near-duplicates of the query video that have been posted on the Web in the past, indicates a video re-use; then, based on the claim that is associated with the video, the fact-checker can decide on whether the video has been used for misleading the viewers about an event, or not.</p>



<p>Target Groups/Beneficiaries</p>	<p>Fact-checkers/journalists (to seek assistance in content verification).</p> <p>Media integrity teams at various companies/organisations (to verify content e.g., in a Know Your Customer (KYC) application setting), Researchers (for comparing their approach).</p>
<p>Innovation, Exploitation and Sustainability Plan</p>	<p>In AI4TRUST, the partner will build on a tool for reverse video search on the Web that is already in place in CERTH, and advance it by: a) automating interaction with multiple search engines and thus facilitating the identification of near-duplicates of the video (Year 1), b) integrating a more advanced method for keyframe selection (Year 2), and c) further improving the robustness of the service (Year 2). This tool will be used both as a stand-alone technology via its web-based UI, as well as an integrated component of the AI4TRUST platform. The stand-alone version of the tool will be hosted at CERTH and will be publicly usable for free.</p>

<p>Asset</p>	<p><b>Deepfake image/video detection</b></p>
<p>Type</p>	<p>Technology</p>
<p>Description</p>	<p>An image or video file is classified as being real or generated using one of the popular deep fake generation models. A score between 0-100 is also produced and expresses the confidence of the decision. In addition, in the case of videos, distinct scores are assigned per video shot and keyframe to help users localise the deep fake in the video.</p>
<p>Target Groups/Beneficiaries</p>	<p>Fact-checkers/journalists (to seek assistance in content verification).</p> <p>Media integrity teams at various companies/organisations (to verify content e.g., in a Know Your Customer (KYC) application setting), Researchers (for comparing their approach).</p>
<p>Innovation, Exploitation and Sustainability Plan</p>	<p>In AI4TRUST, the partner will build on a deepfake image/video detection service that is already in place in CERTH, and through pertinent research, we will improve it by: a) increasing the detection accuracy and reducing the number of false positives in</p>



	<p>existing models (Year 1), b) developing new models that can detect content coming from recent generative AI models (Year 1 and 2), c) providing explanations about the decisions of the developed models (Year 2 and beyond), and d) further improving the robustness of the service (Year 3). Given that deepfake image and video detection are strategic research topics for our team, we will be seeking additional funding from different sources (e.g., research grants, commercial contracts, and licensing) to ensure maintenance and improvement of the service beyond the end of AI4TRUST.</p>
--	--

<b>Asset</b>	<b>AI models for sensational content detection</b>
<b>Type</b>	Technology
<b>Description</b>	<p>A set of trained AI models for annotating images and videos according to the existence of sensational (e.g., shocking, scary, disgusting, excessively violent) visual content, that is typically found in disinformation campaigns. The output of these models (a score in the range [0, 1], with 0 / 1 indicating low / high probability of sensational content) will be used as evidence (in addition to the output of other data analysis technologies of AI4TRUST) for the check-worthiness of the news item or claim that is associated with the analysed image/video, and the prioritisation of its analysis with the different tools for trust-worthiness/reliability evaluation.</p>
<b>Target Groups/Beneficiaries</b>	<p>Fact-checkers/journalists (to seek assistance in content verification).</p> <p>Media integrity teams at various companies/organisations (to verify content e.g., in a Know Your Customer (KYC) application setting), Researchers (for comparing their approach).</p>
<b>Innovation, Exploitation and Sustainability Plan</b>	<p>In AI4TRUST, CERTH will build on visual classifiers that are already in place in CERTH, and we will re-train and fine-tune them (using appropriate data) to effectively detect the existence of various types of sensational visual content (Year 2 and half of Year 3). The trained models will be exposed through an API and run internally to the platform, to annotate the collected visual</p>



	data by the social media listening module of the AI4TRUST platform.
--	---

Asset	<b>AI model for visual-text misalignment detection</b>
Type	Technology
Description	An AI model that gets as input an image/video and a text describing the visual content and makes an estimate about their contextual alignment. The output of this model (a score in the range [0, 1], with 0 / 1 indicating poor / strong alignment between the visual and textual content) will be used as evidence (in addition to the output of other data analysis technologies of AI4TRUST) for the check-worthiness of the news item or claim that is associated with the analysed pair of image/video and text, and the prioritisation of its analysis with the different tools for trust-worthiness/reliability evaluation.
Target Groups/Beneficiaries	Fact-checkers/journalists (to seek assistance in content verification).  Media integrity teams at various companies/organisations (to verify content e.g., in a Know Your Customer (KYC) application setting), Researchers (for comparing their approach).
Innovation, Exploitation and Sustainability Plan	In AI4 TRUST we will build on models for multimodal representation learning that are already in place in CERTH, and we will re-train and fine-tune them (using appropriate data) to effectively make decisions about the level of misalignment between the content of an image/video and the associated textual description (Years 2 and 3). The trained model will be exposed through an API and run internally to the platform, to annotate the collected visual data by the social media listening module of the AI4TRUST platform

### 5.3. UNIVERSITÀ DEGLI STUDI DI TRENTO (UNITN)

The researchers of UNITN have a long-standing experience in the field of **computer vision and multimedia analysis**. Specifically, UNITN has expertise on methods for: a) generating synthetic visual contents with deep generative models, b) visual content classification, d) video analysis and



d) large multimodal models. The collaboration between UNITN and the partners of AI4TRUST has the potential to not only enhance the performance of existing methods developed by UNITN, but also pave the way for the **development of new and more effective approaches**.

UNITN aims to enhance its research expertise, focusing on image and video generation methods to support the development of more powerful tools for deep fake detection useful for fact-checkers and journalists in combating disinformation. The outcomes of UNITN scientific knowledge and technological expertise will be shared through **high-impact scientific papers**. Additionally, UNITN is **open to commercial opportunities**, considering licensing of developed tools, services, or providing consulting services to interested third parties.

UNITN envisions substantial advantages in participating in AI4TRUST by working on real-world cases and data provided by fact-checking and journalist partners. Additionally, the partnership with CERTH will allow the development of **more robust generative models**. Similarly, the collaboration with partners working on audio processing and Natural language processing (NLP) will allow to **improve the module of video anomaly detection with a multimodal pipeline**.

Asset	<b>Tool for image generation</b>
Type	Technology
Description	A technology that allows the generation of images. Given as input an image of a face and the indication of a certain attribute (e.g., colour of the hair) the software produces as output another image with the face and changed attributes.
Target Groups/Beneficiaries	Journalists and media (to create visual contents.) Researchers (for comparing their approach).
Innovation, Exploitation and Sustainability Plan	The AI4TRUST team will build on an existing model developed by UNITN for image generation and will target the further development of the software to handle arbitrary images - e.g., with different head poses, image clutter. The software will be made publicly available online (GitHub).

Asset	<b>Tool for video generation</b>
Type	Technology
Description	A software that permits generating short video clips. It takes as input a video and a segmentation mask which indicates the object in the video that needs to be modified and outputs



	another video where the indicated object is modified in appearance while maintaining motion.
Target Groups/Beneficiaries	Journalists and media (to create visual contents). Researchers (for comparing their approach).
Innovation, Exploitation and Sustainability Plan	Innovation, Exploitation & Sustainability Plan: In AI4TRUST we will build on a state-of-the-art text-to-image generation network available from the research community and will modify the network architecture with temporal layers and with conditioning input thanks to a ControlNet like architecture. Such modifications will improve the software, allowing the users to specify the object that needs to be edited with fine granularity (Year 1). In Year 2 and 3 we plan to integrate other control signals, e.g., textual inputs to further improve the software. The software will be made publicly available online (GitHub).

Asset	<b>Tool for video anomaly detection</b>
Type	Technology
Description	An AI model that gets as input an image/video and estimates if a certain frame corresponds to a normal or an abnormal situation in a video . The output of this model (a score in the range [0, 1], with 0 / 1 indicating normal/abnormal. The software will be made publicly available online (GitHub).
Target Groups/Beneficiaries	Fact-checkers/journalists (to seek assistance in content verification). Researchers (for comparing their approach).
Innovation, Exploitation and Sustainability Plan	Innovation, Exploitation & Sustainability Plan: In AI4TRUST, we will build on models for video anomaly detection that are already in place at UNITN, and we will re-train and fine-tune them to consider data relevant for the AI4TRUST project. We also expect to investigate other foundation models instead of CLIP (Years 2 and 3). The trained model will be exposed through an API and run internally to the AI4TRUST platform.



## 5.4. NATIONAL CENTER FOR SCIENTIFIC RESEARCH "DEMOKRITOS" (NCSR-D)

**NCSR-D** and especially the Institute of Informatics and Telecommunications (IIT) consists of multiple researchers, Post-Docs, PhD, and MSc students, as well as practitioners which are active in the fields of **natural language processing and machine/deep learning**. Lately, IIT has been focused on multidisciplinary research by involving partners from various fields beyond computer science majors, such as the disciplines encompassing **linguistics, law, and ethics**. Furthermore, NCSR-D has been a part of multiple relevant European projects, which provide experience in turning project and research outcomes into a competitive edge and opportunities for innovation. Furthermore, the results of **document intelligence** hold significant importance for the forward-looking digital innovation hub. This hub possesses expertise in advancing innovation outcomes and leveraging them independently. Moreover, it aims to amplify project innovations and contribute to the innovation hub sector, all while extending its reach to the broader research community. In turn the experience, the expertise and multidisciplinary nature of the Institute allow for actions, which will drive innovation such as:

- Leveraging the latest advances and techniques in Natural Language Processing (NLP) and deep learning to develop document intelligence;
- Process and analyse textual datasets and extract valuable knowledge;
- Design and implement scalable and robust systems;
- Communicate and collaborate with other researchers, stakeholders, and users;
- Identify and address ethical and social issues and implications of research driven by development of document intelligence.

Asset	<b>Document intelligence - Technology</b>
Type	Technology
Description	<p>Detecting disinformation is a complex task that often requires analysis beyond the surface level. An orchestrated approach to understand, clearly define, and detect these signals is essential. Document intelligence is an amalgamation of data-driven methods for detecting disinformation signals in textual content. The asset aims to deploy Natural Language Processing (NLP) methods to identify textual content produced for dis/mis/malinformation. Furthermore, it includes extensive analysis of disinformation signals. In turn, this analysis allows for a more verbose identification of disinformation. The verbosity of the asset enriches beneficiaries and stakeholders with informative results and permits deeper understanding of the</p>



	<p>disinformation agenda, while at the same time builds expertise and practical skills on dealing with such signals.</p> <p>From a technological point of view, document intelligence includes:</p> <p>(1) AI techniques that minimise human supervision by relying on a semi-supervised approach, complement existing tools with robust natural language processing and utilise self-learning and active learning paradigms. Furthermore, enriching these neural models with cross lingual fine-tuning enables repurposing of these techniques in multiple use cases and beneficiaries, while improving their generalisation and performance.</p> <p>(2) Analysis, identification, and definition of relevant dis/mis/malinformation signals, whose derivation requires deeper understanding of the content semantics that goes beyond superficial analysis, such as argument mining, fact detection, warning signals for truthfulness, credibility, veracity, and authenticity — detection of bot/AI-generated text, etc.</p>
<p>Target Groups/Beneficiaries</p>	<p>Any platform or organisation that involves user-generated text, especially in domains with disinformation prone themes such as media outlets, researchers, fact-checkers, public health or climate change experts, policymakers, educators, etc.</p>
<p>Innovation, Exploitation and Sustainability Plan</p>	<p>The innovation outcomes of the asset mainly involve neural Natural Language Processing (NLP) models trained with up to date, fact-checked and unique data provided by the fact-checking organisations involved in the project’s consortium. Besides, it includes any analytics or data products that may come up during training of these models, as well as expertise on dealing with disinformation signals. Furthermore, the fine-grained identification of disinformation signals, involvement of multilingual textual content and incorporating knowledge from domain experts of fact-checking organisations through carefully curated data instances and disinformation examples, as well as, through interactive consulting within the AI4TRUST project, allow for competitive advantage.</p> <p>As NCSR-D is a scientific research centre, the main direction for utilising the asset is mainly for academic exploitation. The focus</p>



	<p>of the exploitation would be to publish research outcomes in relevant journals and conferences with the intent to provide a competitive advantage and a specialised expertise to NCSR-D that will advance its collaborations with other institutes and further establish its position in relevant projects. Furthermore, the asset will be exploited within the AHEDD digital innovation hub either by providing the outcomes of the asset (e.g., trained neural models) within the services of the hub or providing expertise and know-how for the involved parties.</p> <p>The plan to sustain the asset is to maintain it for a period up to 3 years after the project ends in collaboration and in agreement with the rest of the consortium. Adjustments and modifications are to be expected regarding the outcomes of the asset in terms of performance or including additional data resources to increase the generalisation and robustness of the trained models. Sustainability of document intelligence also depends on the maintenance of the rest of the tools, such as the AI4TRUST platform. Adjustments in the integration of the asset within the platform such as updating of connection protocols, maintaining the containerisation of the asset, or updating and maintaining technological dependencies are within the scope of the sustainability plan.</p>
--	---

Asset	<b>Document intelligence - Knowledge</b>
Type	Knowledge
Description	Document intelligence includes expertise, practical skills, and knowledge of disinformation signals, such as understanding their features, translating this knowledge into building and training models to classify such signals. Such expertise ensures sustainability and effectiveness of combined methods over time, while they are encapsulated in the AI4TRUST platform.
Target Groups/Beneficiaries	Any platform or organisation that involves user-generated text, especially in domains with disinformation prone themes such as media outlets, researchers, fact-checkers, public health or climate change experts, policymakers, educators, etc.



Innovation, Exploitation  
and Sustainability Plan

As NCSR-D is a scientific research centre, the main direction for utilising the asset is mainly for academic exploitation. The focus of the exploitation would be to publish research outcomes in relevant journals and conferences with the intent to provide a competitive advantage and a specialised expertise to NCSR-D that will advance its collaborations with other institutes and further establish its position in relevant projects. The exploitation of the technological asset within the AHEDD digital innovation hub will provide expertise and know-how for the involved parties.



## 5.5. CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE CNRS (CNRS)

The **CNRS team** involved in AI4TRUST has extensive experience researching social networks and their effects on organisations, markets, and political issues. Team members previously developed a **model of how actors use ‘appropriateness judgments’** to give meaning to information and elaborate it interactively with their networks. Judgments depend on people’s identification to reference groups, recognition of authorities, and alignment with priority norms. Adoption of mis/dis/malinformation should not be taken for granted and can rather be hypothesised to increase when judgments are similar and signalled as such in communication networks. AI4TRUST could build on these ideas and flag such signals to help users in their contextualization and interpretation of the phenomena described. The CNRS team also developed new research **mixing network and content analysis (‘socio-semantic networks’)** in response to findings on how network structures heavily affect the propagation of mis/dis/malinformation, regardless of the characteristics of individuals in these networks. Additionally, CNRS members have in-depth knowledge of cross-country, shadow market of paid engagements that contributes to injecting fabricated behavioural traces into social media, and can influence the **speed and direction of propagation of false or misleading contents**.

Asset	<b>Expertise in social science, especially social network analysis</b>
Type	Knowledge
Description	CNRS brings to the project expertise in social science, with a distinctive focus on social network analysis
Target Groups/Beneficiaries	Researchers and policymakers
Innovation, Exploitation and Sustainability Plan	The CNRS team can provide insight into the motivations, constraints, and social pressures that lead actors, under uncertainty and embedded in their social networks, to interactively elaborate information and participate in the production and dissemination of harmful contents. Understanding the social dynamics that drive the production and diffusion of online information and contents is essential to identify forms of mis/dis/malinformation and to explain how some of them spread, while others do not. On this basis, CNRS can help to identify behavioural and social signals that, in addition to semantic signals, support the future AI4TRUST platform in its effort to detect mis/dis/malinformation. CNRS can



	<p>also provide suggestions on how to improve platform design to account for both social and semantic signals and therefore, to improve its performance. The team’s deep knowledge of the broad social media environment can offer insights into the political, economic and organisational factors that can impact observed outcomes as well as ethical issues specific to this type of setting.</p> <p>The large and rich network datasets that can be accessed through the project can help the CNRS team to test and refine social network theories at unprecedented scale and level of detail. The expected outputs have potential to significantly enrich CNRS knowledge base and to stimulate further research in this area. Further, the collaboration with non-academic partners and the close contact with policymakers can provide invaluable insights to pursue CNRS’s social impact goals.</p>
--	---

Asset	<b>Social Network Analysis Tool</b>
Type	Technology
Description	<p>This Social Statistical Network Analysis tool will aim to visualize the distribution and dissemination of artifacts (such as URLs) or thematic elements (like embeddings) within a particular system, such as Telegram channels focused on specific topics. The tool will collect data by gathering artifacts from messages exchanged within specified Telegram channels and themes derived from content using methods like word embeddings or topic modeling. Visualizations will be created using an interactive mapping tool akin to Gephi, with nodes representing channels and edges indicating the relationships between them based on the frequency of information exchange. Spatialization of nodes will likely be determined by algorithms considering connection strength, often employing force-directed layouts to highlight clusters of closely related nodes. Different colors will be used to represent various themes or types of artifacts, with intensity possibly indicating frequency or importance within the network. Principles such as degree (measuring node connections), strength (measuring connection weight), and clustering</p>



	(grouping nodes into communities) might ideally be employed to elucidate the important element of the network, possibly of its dynamics. Users will be able to interact with the map, zooming, panning, and accessing detailed information via tooltips.
Target Groups/Beneficiaries	Project partners and future users of the platform (i.e., fact-checkers, media workers, policymakers)
Innovation, Exploitation and Sustainability Plan	The joint visualization of the network of interconnection across online channels and various semantic features related to disinformation would offer a novel tool that CNRS would rely to analyze and showcase the connection between topological structure and misinformation dissemination. Such a progress would be further useful for CNRS broader understanding of joint social and semantic dynamics, for instance in terms of assessing whether there exist specific, assortative clusters of social groups which are feeding on the same type of information or not. This would also make it possible to hope for a comprehensive understanding of content dissemination dynamics on group-based platforms, such as Telegram.

Asset	<b>Coordinated Inauthentic Behaviour Tool</b>
Type	Technology
Description	This functionality will rely on a re-elaboration of quantitative analysis of the reconstructed networks of social interaction on each social media platform mapped throughout the project. The intended use is to allow end-users to explore how dis/misinformation circulates within and across platforms thanks to common languages/discourses. It is shaped as a mini-dashboard that allows end-users to visualise some qualitative and quantitative indicators about the prominence of some contents within and across platforms.
Target Groups/Beneficiaries	Project partners and future users of the platform (i.e., fact-checkers, media workers, policymakers)



<p>Innovation, Exploitation and Sustainability Plan</p>	<p>This functionality exploits the results of the pre-processing phase via Social Network Analysis while rendering them visible and useful for end-users. The innovation in the simple indicators provided by the tool will rely on their unique computation from the combination of rich social media data, automatic labelling tools for disinformation detection, and social media analysis which is rarely combined with the two former. This component is as sustainable as the data and technologies are, but needs to be recomputed each time the data is updated.</p>
---	---

<p>Asset</p>	<p><b>Reliability State of Social Media Tool</b></p>
<p>Type</p>	<p>Technology</p>
<p>Description</p>	<p>This functionality will rely on a mostly quantitative analysis of the reconstructed network of social interactions on each social media. The intended use of this component is to provide the end-user with synthetical quantitative indices of the disinformation risk presented in given areas of social media that the platform is able to map, to bring attention to areas of interest where the risk of disinformation flowing is higher. This simple tool will be beneficial to any category of end-user, as it will paint a straightforward picture of the reliability state of interactions under study.</p>
<p>Target Groups/Beneficiaries</p>	<p>Project partners and future users of the platform (i.e., fact-checkers, media workers, policymakers)</p>
<p>Innovation, Exploitation and Sustainability Plan</p>	<p>The CNRS team, using data and technologies provided in WP2 and WP3, combined with Social Network Analysis (SNA), can help provide synthetical measures of disinformation risk in areas of interaction between social media users, to assist the consortium and the end-user of the platform in identifying less reliable areas. The innovation in these simple measures will rely on their unique computation from the combination of rich social media data, automatic labelling tools for disinformation detection, and social media analysis which is rarely combined with the two former. This component is as sustainable as the</p>



	data and technologies are, but needs to be recomputed each time the data is updated.
--	--

<b>Asset</b>	<b>Emotional State of Social Media Tool</b>
<b>Type</b>	Technology
<b>Description</b>	<p>The partners in WP4 will explore an analysis of the emotional state of social media. This analysis could be twofold:</p> <ul style="list-style-type: none"> <li>- statistics of emoji reacts to content/users/... when available (depending on the social media API), in the format of a synthetical (e.g. binary) score of endorsement for each piece of content/users/...</li> <li>- statistics of sentiment analysis in textual reactions to content</li> </ul>
<b>Target Groups/Beneficiaries</b>	Project partners and future users of the platform (i.e., fact-checkers, media workers, policymakers)
<b>Innovation, Exploitation and Sustainability Plan</b>	<p>This account of emotional online expression will be made from the social media data collected in WP2. This functionality requires reactions (likes, dislikes, ...), body of text (comments, messages, ...) including emojis if applicable, to be kept during data collection. By incorporating a component of emotional analysis at scale, both straightforwardly drawn from social media features and indirectly produced using sentiment analysis, the AI4Trust platform will be the first automated tool to combine information tracking and checking with information on the emotional reception of said information, at least to our knowledge.</p> <p>This module will help researchers understand better how the use of emotion is leveraged by disinformation spreaders to maximise the endorsement of falsified content. Furthermore, it will help journalists and policy-makers contextualise the social reception of information.</p> <p>To guarantee sustainability and long-term exploitability at scale, this component will be best applied on relatively viral pieces of content, which gather many reactions and comments if</p>



	<p>applicable, and are associated with a sufficient volume of text data. Only with this information will we be able to draw a relevant account of the emotional state in certain areas of social media platforms. It is important to note that for sentiment analysis, the contribution of expert technical partners such as computer scientists from WP3 will be needed.</p>
--	---

Asset	<b>Infodemic Observatory Tool</b>
Type	Technology
Description	<p>The “Infodemic Observatory” will be capable of tracking aggregated statistical information on the quantity of misleading news circulating in the past week on different topics and across various social media platforms, both in absolute terms and relative exposure to the public. The observatory design is inspired by the COVID-19 Infodemics Observatory platform developed by FBK. To characterise the risk for social media users, aggregated information will be made available to media practitioners, fact-checkers, journalists and policy makers end-users:</p> <ul style="list-style-type: none"> <li>- The total number of news collected in the platform;</li> <li>- The total volume of messages sharing unreliable news items</li> <li>- The total exposure these messages are expected to have on social media;</li> <li>- An infodemic risk index illustrating how likely it is to encounter unreliable news at that moment in time.</li> </ul> <p>This information will be provided as a time series illustrating the evolution in time of these indicators in the different languages, topics and platforms. At the same time, once aggregated over all different languages, these indicators will be also grouped geographically and represented as a map evolving over time.</p>
Target Groups/Beneficiaries	Project partners and future users of the platform (i.e., fact-checkers, media workers, policymakers)



<p>Innovation, Exploitation and Sustainability Plan</p>	<p>This tool will rely on social media data collected in WP2, fact-checking conducted by social media partners, and leverage disinformation predicting tools developed in WP3.</p> <p>The innovation lies in the unique alliance of fact-checking and technology to produce a novel tool for risk estimation.</p> <p>This component is as sustainable as the data and technologies are, and will be optimally developed in a pipeline that continuously collects and analyses social media data.</p>
---	--

## 5.6. UNIVERSITATEA POLITEHNICA DIN BUCUREȘTI (UPB)

The team of UPB has substantial experience in the field of **audio and speech processing**, and more specifically methods for:

- speech to text transcription;
- text to speech synthesis;
- detecting audio deep fakes and AI synthesised content;
- detecting emotions from speech;
- speaker identification.

UPB research experience and expertise in combination with technology requirements by fact-checking and media organisations of the AI4TRUST consortium, will drive new advances in UPB existing solutions and potentially lead to new methods that considerably improve the performance of UPB methods and services.



Asset	<b>Speech-to-text technology and web service</b>
Type	Technology
Description	A technology that allows a user to transcribe the spoken content of a multimedia file (audio or video that contains speech). The technology is deployed as a service that allows content upload, transcription progress assessment and transcription download. The technology is and will be adapted for transcribing news, press conferences and media content in general and is and will be available for various European languages, such as RO, EN, ES, PL, DE, EL, and IT.
Target Groups/Beneficiaries	Media companies (to transcribe media content for writing news articles or reports); Fact checkers (to access faster the spoken content of long multimedia files).
Innovation, Exploitation and Sustainability Plan	UPB will bring into the project a tool for speech to text already in place in UPB workflows, and it's planning to upgrade it by: a) updating the AI transcription models for RO and EN (Year 1), b) adding new AI models for other European languages (Year 2), and c) extending the API to support transcription of audio/video files based on their URL (Year 2). This tool will be used both as a stand-alone service as well as an integrated component of the AI4TRUST platform. The service will be further used by UPB in other research projects.

Asset	<b>Audio deep fake detection technology and web service</b>
Type	Technology
Description	A technology that allows a user to assess whether the audio stream in a multimedia file (audio or video that contains speech) is computer generated or manipulated by cutting and pasting various real or computer-generated content.
Target Groups/Beneficiaries	Fact-checkers/journalists (to assess content veracity)



<p>Innovation, Exploitation and Sustainability Plan</p>	<p>In AI4TRUST, UPB will build this technology and the related service from scratch, as follows: a) the base technology is developed for English (Year 1), b) the technology is updated to support AI explainability (Year 2) and c) a web service wrapping the technology and exposing it to end users is created (year 3). The technology and service will be used stand-alone and integrated in the AI4TRUST platform. They will be further used by UPB in other research projects.</p>
---	--

<p>Asset</p>	<p><b>Audio anomaly detection technology and web service</b></p>
<p>Type</p>	<p>Technology</p>
<p>Description</p>	<p>A technology that allows a user to assess whether the audio stream in a multimedia file (audio or video that contains speech) contains anomalies. At this point it is mandatory to decide what can be considered an anomaly from the user's point of view (possible candidates could be: speech slowed down or bomb sounds were added in the background).</p>
<p>Target Groups/Beneficiaries</p>	<p>Fact-checkers/journalists (to assess content veracity)</p>
<p>Innovation, Exploitation and Sustainability Plan</p>	<p>In AI4TRUST UPB will build this technology and the service from scratch, as follows: a) the base technology is developed (Year 1), b) the technology is updated to support AI explainability (Year 2) and c) a web service wrapping the technology and exposing it to end users is created (year 3). The technology and web service will be used stand-alone and integrated in the AI4TRUST platform. They will be further used by UPB in other research projects.</p>

<p>Asset</p>	<p><b>Audio deep fake generation technology</b></p>
<p>Type</p>	<p>Technology</p>
<p>Description</p>	<p>A technology that allows a user to input a text message, select the voice and generate the spoken message corresponding to the input text (text-to-speech technology).</p>



<p>Target Groups/Beneficiaries</p>	<p>AI4TRUST research team (to generate audio deepfakes to further improve the detection technology);  Media companies (to create spoken news starting from text news).</p>
<p>Innovation, Exploitation and Sustainability Plan</p>	<p>In AI4TRUST, UPB will improve a tool for text-to-speech that is already in place in UPB by: a) creating newer, more natural sounding AI models (Year 1), b) evaluating the impact of various vocoders on speech naturalness (Year 2). This tool will be used as a stand-alone technology in UPB to create deepfakes for further improving the detection technology. The technology will be further used by UPB in other research projects.</p>

## 5.7. SAHER (EUROPE)

As a company, SAHER (EUROPE) is highly experienced and skilled in **creating and exploiting its professional networks to maximise project impacts**. The knowledge that SAHER gains throughout the project is both used as a foundation for its commercial activity and for using it as a platform for future collaborative work. Furthermore, SAHER with its renowned expertise in **ethical and legal aspects** related to the development and implementation of AI-based technologies is a key partner for the **release of trustworthy and EU-GDPR compliant solutions**.

<p>Asset</p>	<p><b>Networking through stakeholders</b></p>
<p>Type</p>	<p>Knowledge</p>
<p>Description</p>	<p>Thanks to its expertise, SAHER will undertake key action to disseminate AI4TRUST outcomes to maximise the impact and reach out future academic, research and industrial partners in the project. This will involve attending domain-specific events and gatherings and seeking opportunities to promote the AI4TRUST network and the results generated by it. This networking will be undertaken by:</p> <ul style="list-style-type: none"> <li>● individual discussion and negotiations with pre-selected partners both face to face and virtually;</li> <li>● cooperation on research projects, by inputting to other networks and seeking to create new networks that will add value;</li> </ul>



	<ul style="list-style-type: none"> <li>pro-active networking to attract new partners and commercial opportunities to exploit based on the knowledge gained by participating in the AI4TRUST network and by contributing to platforms, EU initiatives and alongside activities to exploit the network through writing publications and contributions using online and new media opportunities such as LinkedIn posts.</li> </ul>
Target Groups/Beneficiaries	Researchers and Professionals
Innovation, Exploitation and Sustainability Plan	<p>The dissemination activity will be based on to three specific short-term goals:</p> <ul style="list-style-type: none"> <li>to build a network of potential academic partners (European universities). The potential academic partners will, in the first phase, be selected according to the strongest performing fields of research within each institute. This will be part of our knowledge transfer and knowledge exchange activities with the company;</li> <li>to establish a network of potential industrial partners;</li> <li>to create a network of potential research partners (including end-users).</li> </ul> <p>The empowerment of the communication in different networks and related activities will support the exploitation of the AI4TRUST results in a broad range of domains and stakeholders.</p>

Asset	<b>Extensive expertise in ethical and legal aspects</b>
Type	Knowledge
Description	The SAHER(EUROPE) strategy for implementing security innovations is reinforced by incorporating internationally recognised research legal and ethics expertise, as well as extensive security policy review and evaluation at both national and international levels.
Target Groups/Beneficiaries	Researchers and Professionals



<p>Innovation, Exploitation and Sustainability Plan</p>	<p>The Legal &amp; Ethics Team of SAHER comprises internationally recognised professionals who contribute a wealth of expertise to guide and advise on complex research and innovation programs. With a deep understanding of developing legal and ethical frameworks for international research, the team possesses unique proficiency in areas such as data protection, privacy, human rights, and the ethical considerations surrounding biometrics, facial recognition, big data analytics, and the utilisation of AI.</p> <p>Emphasising and empowering the ethical and legal aspects will facilitate the exploitation of AI4TRUST outcomes across various domains and stakeholders in a manner that fully complies with the EU-GDPR and aligns with the requirements of the new AI Act.</p>
---	---

## 5.8. GDI GLOBAL DISINFORMATION INDEX GUGHAFTUNGSBESCHRANKT (GDI)

**GDI** has in-depth knowledge in the development of **technology which can automatically detect signs of online disinformation**. Also, GDI’s team has subject matter expertise on disinformation **in various geographical contexts**. GDI leverages recent advances in **Natural Language Processing (NLP)** to detect domains which are at risk of sharing disinformation in various languages. The list of domains included in **GDI’s data platform** relies on a review process performed by a team of trained intelligence analysts. The Manual Review is run across sites identified by its machine learning classifiers as carrying the highest potential disinformation risk. Websites are analysed to identify the presence of adversarial narratives<sup>6</sup> tracked by GDI. Each website is reviewed by a minimum of two intelligence analysts who perform a “blind” review meaning that they do not see each other’s rating.

Asset	<b>Data used for ground truth</b>
Type	Data
Description	GDI will share a dataset to contribute to the effort of Work Package 2-”Methodological design, data gathering and pre-processing” to collect ground truth data. The data shared by GDI will include domains in French, English, German, Spanish and

<sup>6</sup> <https://www.disinformationindex.org/blog/2022-06-22-disinformation-as-adversarial-narrative-conflict/>



	Italian. This dataset will be shared with consortium partners; these domains have all been labelled by GDI as sharing disinformation.
Target Groups/Beneficiaries	Professional and Researchers
Innovation, Exploitation and Sustainability Plan	<p>The dataset provided is the result of GDI’s work using both machine learning and the expertise of GDI’s intelligence team. To constitute this dataset, GDI’s team has leveraged its technological capabilities to scrape domains from the open web and assess domains which are likely to share disinformation. Additionally, GDI implements a review process performed by a team of trained intelligence analysts. The manual review is run across sites identified by GDI technology as carrying the highest potential disinformation risk.</p> <p>This dataset will be used by consortium partners, especially technical partners developing technology components in Work Package 3-“AI-driven data analysis methods”. The timeline for the use will be regulated by a consortium’s data sharing agreement.</p> <p>The creation and exploitation of this asset by GDI pre-exists and is independent of the AI4TRUST project. In that regard, this asset will continue to be exploited by GDI for research/ commercial/ educational purposes during and after the AI4TRUST project.</p>

Asset	<b>Disinformation warning system (DWS)</b>
Type	Technology
Description	As part of Work Package 3-“AI-driven data analysis methods”, GDI is going to develop a new model alongside other partners of Task 3.4. This new model will aim to detect whether a piece of content is likely to contain disinformation or not, also providing a confidence score.
Target Groups/Beneficiaries	Professional and Researchers



<p>Innovation, Exploitation and Sustainability Plan</p>	<p>This model will be specifically designed by GDI and other partners involved in T3.4 to integrate outputs from technologies developed in WP3, as well as GDI’s data platform. This model will be developed to respond to the specific needs of the AI4TRUST consortium, in terms of language and type of content reviewed by the tool (text, image, audio, multimodal). The timeline for this asset is tied to the delivery of technologies developed in T3.1-T3.3 given the disinformation warning system will integrate the outputs of these technologies.</p> <p>This asset will serve the consortium’s work and help in the detection of disinformation. This asset is specially designed to meet the needs of the consortium, and will be tailored to integrate technologies from T3.1-T3.3. In that regard, GDI does not plan to exploit this asset outside of AI4TRUST consortium work.</p>
---	--

<p>Asset</p>	<p><b>Data platform for indexing and documenting verified and manipulated content</b></p>
<p>Type</p>	<p>Technology</p>
<p>Description</p>	<p>GDI will use its existing data platform for indexing and documenting verified and manipulated content, which will constitute a feature in the disinformation warning system. This data platform is the result of pre-existing work carried out by GDI, including the identification and labelling of domains which regularly share disinformation across various languages of the AI4TRUST project.</p>
<p>Target Groups/Beneficiaries</p>	<p>Professional and Researchers</p>
<p>Innovation, Exploitation and Sustainability Plan</p>	<p>The database used by GDI will be made up of domains that have been previously labelled by GDI as “worst offenders” in terms of spreading disinformation in the languages used in the AI4TRUST project. The identification of these domains is the combined result of GDI’s technology capabilities, and expertise of GDI’s intelligence team. Each domain within this dataset has been manually reviewed by GDI’s analysts. The exploitation of this</p>



	asset in the AI4TRUST project will be completed by GDI during the development of the model of the disinformation warning system. The creation and exploitation of this asset by GDI pre-exists and is independent of the AI4TRUST project. In that regard, this asset will continue to be exploited by GDI for research/ commercial/ educational purposes during and after the AI4TRUST project.
--	--

## 5.9. STOWARZYSZENIE DEMAGOG (DMGG)

**Demagog Association (DMGG)** has a nuanced knowledge, experience, and understanding of the challenges that surround the process of verifying information on social media. Demagog is the first **fact-checking organisation** in Poland. For ten years, the organisation has been verifying politicians' statements. In 2019, DMGG joined the Third-Party Meta program, and in 2023, it began collaborating with TikTok. The association also belongs to the International Fact-Checking Network (since 2019) and the European Fact-Checking Standards Network (DMGG was heavily involved in the creation of the entire organisation, moreover, Pawel Terpilowski, editor-in-chief of the portal sits on the EFCSN board).

The first resource DMGG brings to the project is a **strong expertise in fact-checking and recognising misinformation, disinformation, malinformation, and conspiracy theories**. The organisation specialises in recognising disinformation on health, war, and climate, targeting minority groups, technology & AI, conspiracy theories and hoaxes. Also, DMGG tracks actors who financially gain from spreading disinformation. These resources are fundamental for the AI4TRUST project. Moreover, DMGG will share **ground-truth data** with the Consortium. This data is a key asset for the project as it will be used **to train AI models** that will be integrated into the AI4TRUST platform. The data will be shared with the Consortium only for the project's lifecycle and objectives, while any eventual further commercial use will be **agreed upon in terms of legitimate interests and protection of intellectual value**.

Asset	<b>Fact-checking analyses data of misleading information published on social media</b>
Type	Data
Description	Demagog Association (DMGG) provides a database containing fact-checking analyses of misleading information published on social media, including fake content generated by artificial intelligence. The database is essential for training the models



	<p>used in the AI4TRUST project to establish a platform to limit false information in the digital space.</p> <p>DMGG has verified more than 2,200 fake news stories in the social media and digital space during its activities. This data is going to be collected and shared as part of the AI4TRUST project to train the AI model that will power the platform. Precise and transparent methodology, reliance on primary sources, correction policies, and financial transparency that are validated by the IFCN and EFCSN ensure that the shared database is representative and reliable.</p>
Target Groups/Beneficiaries	Technical partners and Researchers
Innovation, Exploitation and Sustainability Plan	The data provided by DMGG is a valuable resource to develop more accurate and efficient fact-checking/ disinformation identification tools using artificial intelligence. By M14, this data will be available to the consortium, while DMGG plans to continually update and expand these datasets to ensure their continued relevance and usefulness, even after the end of the project, to guarantee sustainability of the overall results.

Asset	<b>Fact-checking and detecting misinformation, disinformation, and malinformation on social media</b>
Type	Knowledge
Description	DMGG shares critical insights and expertise in terms of fact-checking and identifying misinformation, disinformation, malinformation, and conspiracy theories. DMGG's assistance translates into improving the effectiveness of the tools implemented by technology partners, ensuring that the work results in a platform that will be a useful, practical, and an effective tool in the fight against disinformation.
Target Groups/Beneficiaries	Technical partners of the project and developers of the platform
Innovation, Exploitation and Sustainability Plan	Similarly to other fact-checking organisations involved in AI4TRUST, DMGG's expertise plays a key role in driving



	<p>innovation within the project. DMGG is committed to ongoing collaboration with technology partners to refine and improve AI tools and approaches, ensuring their relevance and applicability. Once the project is completed, there is an opportunity to continue DMGG’s involvement, contributing to the adaptation and enhancement of the platform in future evolutions.</p>
--	--

Asset	<b>Fact-checkers everyday workflow</b>
Type	Process
Description	DMGG is involved in the process of testing the AI4TRUST platform pilot. For this purpose, the process of using the platform by fact-checkers and media professionals in their daily work will be designed, along with reporting on the level of satisfaction and usefulness of the pilot version of the AI4TRUST platform.
Target Groups/Beneficiaries	Project partners and future users of the platform (i.e., fact-checkers, media workers, policymakers)
Innovation, Exploitation and Sustainability Plan	The pilot testing is a key phase to the success of the project, providing immediate and practical feedback to technical partners. This information will be used to refine and optimise the AI4TRUST platform. After the end of the project, further use and feedback will contribute to the platform’s long-term improvement and adaptation to the dynamic and changing needs of fact-checking.

## 5.10. FUNDACION MALDITA.ES CONTRALA DESINFORMACION: PERIODISMO EDUCACION INVESTIGACION Y DATOS EN NUEVOS FORMATOS (MALDITA)

**MALDITA** brings to the table a wealth of experience and expertise in the field of **mis/disinformation detection and fact-checking**. Its team comprises skilled journalists, fact-checkers, and technology product and development experts who have a proven track record in combating dis/mis/malinformation and promoting media literacy. MALDITA’s expertise aligns



seamlessly with the assets outlined, particularly the integration of AI4TRUST platform AI tools into our Disinformation Management System. MALDITA has a comprehensive understanding of the complexities and nuances of the disinformation landscape.

MALDITA's experience in debunking and analysing mis/disinformation patterns positions us as a valuable partner in fine-tuning AI solutions to effectively detect and combat mis/disinformation. We **bring firsthand knowledge of the challenges faced by fact-checkers**, allowing us to provide insights and feedback that are essential for the development and optimisation of AI tools. Leveraging its editorial team's expertise, MALDITA will actively **contribute to integrating AI4TRUST platform AI tools into its news production processes**. MALDITA's journalists possess a deep understanding of reporting standards and journalistic ethics, ensuring that the advanced tools are seamlessly **integrated into MALDITA's workflows while upholding accuracy and integrity**.

MALDITA is committed to driving innovation in the fight against mis/disinformation while upholding the highest **standards of quality journalism**. MALDITA's dedication to accuracy, transparency, and accountability ensures that the integration of AI solutions into MALDITA's fact-checking processes is conducted with the utmost rigour and integrity. MALDITA actively seeks to **innovate and improve its practices**, continually refining its approaches to better serve its audience and combat dis/mis/malinformation effectively.

As a fact-checking partner, **MALDITA is well-positioned to deliver significant benefits to the consortium:**

- **Real-world scenarios and practical use cases:** MALDITA brings firsthand experience in combating mis/disinformation in real-world contexts. By providing access to **MALDITA's extensive database of mis/disinformation cases and fact-checking methodologies**, MALDITA offers valuable insights into the practical challenges and opportunities associated with disinformation detection and debunking. These real-world scenarios serve as essential test cases for evaluating the effectiveness and reliability of the AI4TRUST outcomes in real-world applications.
- **Validation and testing:** as a trusted fact-checking organisation, MALDITA provides **credibility and validation to the AI4TRUST outcomes**. MALDITA offers its platform and resources for testing and validation, ensuring that the AI tools effectively address the needs and requirements of fact-checkers. By rigorously **testing the AI4TRUST solutions in a real-world environment**, MALDITA contributes to enhancing their **accuracy, reliability, and usability**, ultimately improving their **value proposition** for the Consortium.
- **Shared interest in combating mis/disinformation:** MALDITA shares a common interest with other project partners in combating disinformation and promoting **media integrity**. By collaborating closely with academic and technology partners within the Consortium, MALDITA contributes to a **holistic approach to mis/disinformation mitigation**, leveraging its respective expertise and resources to achieve shared objectives. Together, MALDITA



aims to develop **innovative solutions and methodologies** that effectively address the challenges posed by disinformation in today's digital landscape.

Asset	<b>Data used for ground truth</b>
Type	Data
Description	Maldita will share a dataset to contribute to the effort of Work Package 2-”Methodological design, data gathering and pre-processing” to collect ground truth data. This dataset will be shared with consortium partners.
Target Groups/Beneficiaries	Professionals and Researchers
Innovation, Exploitation and Sustainability Plan	<p>The dataset provided is the result of the expertise of Maldita’s team. To constitute this dataset, Maldita’s team has included reports of contents with possible disinformation from users on WhatsApp and other social media such as Twitter, Facebook, Instagram, and Telegram.</p> <p>This dataset will be used by consortium partners, especially technical partners developing technology components in Work Package 3-“AI-driven data analysis methods”. The timeline for the use will be regulated by a consortium’s data sharing agreement.</p> <p>The creation and exploitation of this asset by Maldita pre-exists and is independent of the AI4TRUST project. In that regard, this asset will continue to be exploited by Maldita for research/ commercial/ educational purposes during and after the AI4TRUST project.</p>

Asset	<b>API connection of AI4TRUST tools to Maldita’s Disinformation Management System</b>
Type	Technology
Description	The AI4TRUST AI tool will be integrated through APIs connection to the Maldita’s Disinformation management system to test and validate the AI4TRUST models and technologies on



	<p>an existing, in production and every day used disinformation management framework and related data.</p>
<p>Target Groups/Beneficiaries</p>	<p>Maldita’s newsroom team</p>
<p>Innovation, Exploitation and Sustainability Plan</p>	<p>The combination of AI4TRUST platform AI tools with Maldita's Disinformation Management System represents a significant enhancement to our fact-checking capabilities. By leveraging the AI capabilities offered by AI4TRUST, we aim to improve several aspects of the fact-checking process, thereby improving the efficiency and effectiveness of our operations.</p> <p>The integration of AI4TRUST tools in daily activities will lead to a more efficient fact-checking workflow for our newsroom team. For instance, we anticipate significant time savings for our fact-checkers, enabling them to focus more of their time and energy on higher-value tasks, such as verifying the accuracy of information and producing high-quality debunk articles.</p> <p>One of the primary benefits of streamlining our fact-checking workflow is the ability to respond more quickly to user inquiries and reports of disinformation. By reducing the time it takes to identify and debunk false information, we can improve our overall response time and provide more timely and accurate information to our audience. This will enhance trust in our organisation and increase engagement with our fact-checking services.</p> <p>Moreover, we aim at being better positioned to scale our fact-checking operations to handle larger volumes of dis/mis/malinformation. This scalability is essential for meeting the growing demand for fact-checking services and ensuring that we can effectively combat disinformation across various platforms and channels. Additionally, we are laying the foundation for long-term sustainability and continued innovation in our fact-checking efforts.</p>



## 5.11. ASTIKI MI KERDOSKOPIKI ETAIRIA KENTRO KATAPOLEMISIS TIS PARAPLIROFORISIS / CIVIL NON-PROFIT COMPANY KENTRO KATAPOLEMISIS TIS PARAPLIROFORISIS (ELLINIKA)

ELLINIKA's organisation's expertise in fact-checking is deeply intertwined with each of the assets of the project. With years of experience in **identifying, analysing, and debunking misleading information**, ELLINIKA brings a nuanced understanding of the challenges and requirements in this field. This expertise is vital in shaping the development and application of the AI4TRUST platform. With an almost 10-year active presence in Greece, ELLINIKA has fact-checked numerous misleading claims and has produced more than 2,000 debunking articles. These articles constitute our **database**, which has now been shared with the Consortium to be used for the development of the AI4TRUST platform. Our high journalistic standards, aligned with the international fact-checking standards suggested by the International Fact-Checking Network, of which we are a signatory, ensure that our data are not only extensive but also accurately representative. Below are the main expertise ELLINIKA brings to the project:

- **Expertise in fact-checking:** Our experience and skills in the fact-checking field are directly transferred to the development of the AI4TRUST platform. We have already and will continue to provide critical insights and feedback to the tech partners, ensuring that the tools developed are not only technologically advanced but also practically applicable and effective.
- **Pilot testing of the AI4TRUST platform and database:** ELLINIKA's expertise extends to the practical application of the platform, where, once it is developed, we intend to critically assess its performance in real-world scenarios. Our experience will allow us to provide valuable feedback on the platform's functionality and usability, directly influencing its refinement and optimization. ELLINIKA brings important strengths to the AI4TRUST project, including our expertise and resources in tackling mis/disinformation. By providing datasets and engaging in the pilot testing of the platform, ELLINIKA brings significant benefits to the consortium and aligns closely with the project's goals. ELLINIKA's most important benefits offered to the consortium are data and the ability to test the platform. Our datasets, which include all our fact-checking articles and cases containing AI-generated disinformation, serve as a valuable resource for all partners involved. At the same time ELLINIKA's commitment alongside other fact-checking organisations to pilot test the platform is of great importance for the project as providing direct, actionable feedback on the platform's efficiency is crucial for the platform's iterative improvement. This real-life testing of the platform ensures that the tools developed are not only technologically advanced but also practically applicable and effective in real-world fact-checking scenarios.



Asset	<b>Data related to fact-checking articles and dataset with generated AI content</b>
Type	Data
Description	In collaboration with other fact-checking partners in the project, ELLINIKA provides essential data, including all our fact-checking articles and datasets containing AI-generated disinformation. This data is crucial for training the AI tools of the AI4TRUST platform.
Target Groups/Beneficiaries	Technical partners and academic researchers in the field of dis/mis/malinformation
Innovation, Exploitation and Sustainability Plan	These datasets represent a valuable resource for developing more accurate and efficient AI tools. By M14, these datasets will have already been shared with the consortium, and we plan to continually update and expand these datasets to ensure their ongoing relevance and usefulness.

Asset	<b>Fact-checking background and expertise</b>
Type	Knowledge
Description	ELLINIKA contributes expert knowledge and experience in the field of fact-checking. This expertise is essential to guide technical partners in the development of AI tools, ensuring they are accurately aligned with the typical, everyday requirements of fact-checking operations.
Target Groups/Beneficiaries	Project consortium and future users of the platform.
Innovation, Exploitation and Sustainability Plan	As with all fact-checking organisations involved in the project, ELLINIKA's expertise plays a critical role in driving innovation within the project. We are committed to ongoing collaboration with tech partners to refine and enhance AI tools and approaches, ensuring their relevance and applicability. Post-project, we plan to continue our engagement, contributing to the platform's adaptation and improvement if needed.



## 5.12. EURACTIV MEDIA NETWORK B.V. (EURACTIV)

**Euractiv’s** expertise derives from the 20 years of experience in **EU policy news** and the wide range of journalists with varying years of experience (junior and senior journalists who have different workflows and priorities). The skills offered by Euractiv will support the development of the AI tool every step of the way - providing **knowledge on journalists’ workflows, testing pilot versions and ultimately the final product**. Euractiv offers **journalists in English, French and German** – aligning with the **multilingual angle of the project**.

Euractiv’s involvement is a real need both internally and externally from the project. Internally, its expertise is required to **properly test the platform** before it is made public. Externally, **Euractiv’s journalists will continue to combat disinformation** and increase the public’s trust in news organisations by generating trustworthy content both during and after the project has finished. **Euractiv will disseminate the results of the project through its channels** (i.e., website, social media, events) through editorial and multimedia content. This will increase the reach of AI4TRUST, complimenting Euractiv’s communication work which focuses on sharing general information on Artificial Intelligence.

Asset	<b>Expertise in journalism</b>
Type	Knowledge
Description	The Euractiv Media Network’s newsroom hosts a wide range of expertise in Brussels, Paris, and Berlin. The journalists will offer their knowledge in what is required to create news content; testing the efficacy and efficiency of the AI4TRUST fact checking tools.
Target Groups/Beneficiaries	Professionals
Innovation, Exploitation and Sustainability Plan	In a business perspective, Euractiv will benefit from the AI4TRUST platform in the long term as its journalists will be able to use it to further enhance the public’s trust in public institutions. By testing the platform, Euractiv’s skills will continue to develop in creating trustworthy news.

Asset	<b>Access to content generated by Euractiv</b>
-------	--



Type	Data
Description	<p>Euractiv Media Network is an independent pan-European media organisation, with expertise across Europe. The newsroom covers 8 different hubs, ranging from Artificial Intelligence to Climate Change and European Politics.</p> <p>AI4TRUST and Fact-checkers will have access to content generated by Euractiv on these wide-ranging topics, as well as multi-faceted formats including editorial, video, infographics and podcasts.</p>
Target Groups/Beneficiaries	Professionals and Researchers
Innovation, Exploitation and Sustainability Plan	Euractiv does not plan to exploit this asset outside of AI4TRUST consortium work.

## 5.13. SKYTG24

**Sky TG24** pr offers comprehensive **daily news coverage**, encompassing all relevant events, both domestic and international, on air as on its website and social media. With more than 7 thousand hours of live content each year, and in-depth reporting, Sky TG24 can also provide **expertise in the countering of disinformation**. Over the years, it has acquired extensive experience with **TV programs like "Numeri" , "Pillole di vaccino" and "Impact"**, aimed to spread high quality information and data literacy on economic and political topics, vaccination campaigns and climate change, tackling dis/mis/malinformation (as per D6.1, 2.3.1 section).

Sky TG24 can provide state-of-the-art expertise in **analysing and testing the AI4TRUST tools** from the perspective of a media company and journalists. The daily activity in journalism, carried out through the complexity of the news landscape, can provide a **wider point of view on relevant issues for media companies** in the fight against dis/mis/malinformation.

Asset	<b>Expertise in journalism field</b>
Type	Knowledge
Description	<p>Sky TG24 can provide its knowledge of day-to-day news coverage and expertise in covering all relevant news: this can help to analyse if/to what extent the platform under development can meet the needs of journalists and media companies. Sky TG24 can also provide suggestions - if helpful -</p>



	on how to change the platform to make it more useful for journalists and media companies. Sky TG24 knowledge of the media ecosystem can also offer useful insights on the journalism field to inform all the phases of the project.
Target Groups/Beneficiaries	Journalists and Media Companies
Innovation, Exploitation and Sustainability Plan	Working alongside other media companies can help Sky TG24 better understand the needs of the sector, exchange valuable knowledge and expertise, and discuss different points of views on the best ways to tackle disinformation. Moreover, working alongside both industrial and academic partners allows us to build connections between studies on social media, journalism and technological research and development. Working with academic and industrial partners can broaden our view on different approaches and needs, further augmenting Sky TG24 abilities to understand the reality around us. Sky TG24 can provide to consortium partners its expertise in the field of journalism, allowing them to better understand what the needs of a media company in tackling dis/mis/malinformation are online.

## 5.14. ASOCIATIA DIGITAL BRIDGE (ADB)

**Facilitating fact-checking and debunking** in Romanian newsrooms is crucial, given the rarity and inconsistency of such activities in Romania. An AI solution has the potential to expedite the development of fact-checking endeavours. However, larger newsrooms often lack the flexibility to experiment with various platforms and tools. **ADB Euractiv Romania**, structured as an NGO with the specialisation of a media outlet, possesses the capability to **test the AI4TRUST platform** within a more specialised domain, specifically in the field of European policies.

Asset	<b>Expertise in Journalism</b>
Type	Knowledge
Description	ADB - Euractiv Romania boasts a small yet experienced team of journalists capable of thoroughly testing the platform and offering feedback on both technical responsiveness and the quality of AI-generated results.



Target Groups/Beneficiaries	Journalists and Media Companies
Innovation, Exploitation and Sustainability Plan	The team of journalists involved possess the expertise and flexibility necessary for engaging in projects and testing prototypes aimed at enhancing the quality of journalism and streamlining the journalistic workflow in the Romanian language newsroom. The testing timeline will be aligned with the project's requirements. Once a piece of information is checked, ADB - Euractiv Romania will publish the respective item in Romanian language in the section Facts, not Fake, on euractiv.ro.

Asset	<b>Partnership with key media organisations and fact-checking NGOs / companies</b>
Type	Knowledge
Description	The core team of journalists can collaborate with ADB - Euractiv Romania's media partners, including newsrooms, fact-checking NGOs, or academics specialised in debunking, to expand the platform testing. ADB - Euractiv Romania can offer feedback and suggestions based on their own testing as well as any testing extended to partners.
Target Groups/Beneficiaries	Fact-checkers
Innovation, Exploitation and Sustainability Plan	ADB - Euractiv Romania aims to use the AI4TRUST outcomes and gains expertise to enlarge and differentiate its network.

Asset	<b>Native knowledge of Romanian language</b>
Type	Knowledge
Description	ADB - Euractiv Romania foresees the possibility to use the AI4TRUST results to improve the Romanian language's transcriptions tools.
Target Groups/Beneficiaries	Researchers and Professionals



<p>Innovation, Exploitation and Sustainability Plan</p>	<p>The ADB - Euractiv Romania team of journalists has already begun using the transcription app, which is a component of the AI4TRUST project. While utilising the app for transcriptions, they also make corrections within the app for the returned results in Romanian.</p>
---	--

## 5.15. EUROPEJSKIE MEDIA SP ZOO (EMS)

As a media organisation, **EMS** delivers benefits to the Consortium by providing **real-world scenarios, practical use cases**, and access to its **extensive network** for validation and testing. The EMS organisation possesses a strong foundation in **EU news reporting, policy analysis, and debunking false claims**. The editorial team's expertise aligns seamlessly with the AI4TRUST outcomes, ensuring a deep understanding of the disinformation challenges in Poland. Leveraging its experience, EMS will actively **contribute to the development and fine-tuning of the AI4TRUST AI solutions, integrating them effectively into its news production processes**. EMS commitment to accuracy and journalistic integrity will drive the innovation of project outcomes, ensuring that the advanced tools **not only detect disinformation but also enhance the overall quality and reliability of its news content**. This synergy between EMS organisational skills and the project assets positions EMS as a key driver in achieving the Consortium's objectives.

<p>Asset</p>	<p><b>Knowledge of journalistic, editorial team as regards AI tools, disinformation, news technologies</b></p>
<p>Type</p>	<p>Knowledge</p>
<p>Description</p>	<p>As a media organisation, EMS provides to the project consortium competencies in terms of EU news reporting, policy analysis, and debunking false information. In this case, the asset brought to the project is the skills and expertise of the editorial teams towards challenges belonging to every aspect of disinformation.</p>
<p>Target Groups/Beneficiaries</p>	<p>Professionals and Researchers</p>
<p>Innovation, Exploitation and Sustainability Plan</p>	<p>EMS aims to empower its team's expertise with enhanced capabilities in content verification, source analysis, and disinformation detection. This sort of training on the job will also help EMS to improve competency and familiarity with every day AI technologies (like AI translation robots, AI content production</p>



	and generation – i.e., ChatGPT). This innovation will enable more efficient debunking of fake news, ensuring the dissemination of accurate and trustworthy information to our audience.
--	---

Asset	<b>News process and workflows</b>
Type	Processes
Description	EMS will provide the project the possibility to test new AI tools into its existing news process. The skills of its teams will be used to understand the impact of the AI4TRUST models on existing processes while training of the teams on new technologies and tools will empower global expertise of the organisation.
Target Groups/Beneficiaries	Professionals and Researchers
Innovation, Exploitation and Sustainability Plan	The advanced AI solutions offer a unique added value by providing robust tools to combat disinformation effectively within the EU landscape. EMS plan is to exploit these assets both industrially and academically. Industrially, we aim to integrate the AI solutions into their news production pipeline, offering a more reliable and fact-checked news service to our audience. Simultaneously, we see an opportunity to position ourselves as a provider of commercial solutions, offering expertise and tools to other media organisations facing similar challenges. Academically, we envision contributing to ongoing research and development efforts, fostering collaboration within the consortium and beyond.

## 5.16. University of Cambridge (UCAM)

The Minderoo Centre for Technology and Democracy consists of an independent team of academic researchers affiliated with the **University of Cambridge (UCAM)**. They are engaged in fundamentally reimagining the dynamics of power between digital technologies, society, and the planet.

In today's context, **understanding the collective challenges arising from power, technology, and democracy** is more crucial than ever. UCAM is of the belief that discussions regarding **technology and regulation** often revolve around comprehending, interpreting, and implementing the ideals of



democracy, particularly within complex modern states heavily reliant on technologies that are often unfamiliar to the majority of citizens. This understanding is **vital for citizens** to effectively govern themselves and shape their collective state. UCAM's research is centred around four primary objectives:

1. **Improving public comprehension of digital technologies** and their societal impacts;
2. **Highlighting the global environmental repercussions** stemming from digital technology;
3. **Proposing solutions to address the adverse effects of digital technologies** on citizens' rights;
4. **Cultivating informed trust in digital technology** while emphasising the significance of democratic values over corporate interests.

Asset	<b>Knowledge of policy networks and human-centred AI techniques</b>
Type	Knowledge
Description	Policy and Human-Centred AI techniques knowledge includes expertise, practical skills, and knowledge of how people use systems and how policy makers and civil society use and navigate systems and entails translating this knowledge into building tools to serve such users. Such expertise ensures the 'human-centredness' of the AI4TRUST platform and the relevance of the platform for its intended users.
Target Groups/Beneficiaries	Any platform or organisation that involves user-generated text, especially in domains with disinformation prone themes such as media outlets, researchers, fact-checkers, public health or climate change experts, policymakers, educators, etc.
Innovation, Exploitation and Sustainability Plan	The knowledge developed for the AI4TRUST model has the potential for reuse in providing services for potential users and UCAM reserves the right to provide further services and commercial exploitation of the knowledge developed for the project.

Asset	<b>Hybrid Recommendation Tool</b>
Type	Technology



<p>Description</p>	<p>This tool can help end users to overcome their information gaps about the current state of the information ecosystem with platforms walling off access to data and can help identify systemic challenges and risks (e.g. public health, fundamental rights)</p> <p>To do so, the recommendation tool draws on different functionalities of the AI4TRUST platform, namely the detection and analysis of disinformation signals (e.g. claim validity and Social Network Analysis). The purpose of the tool is to link aggregated inputs from the AI4TRUST platform with a classification of their severity level and a guide towards mitigation measures. The hybrid character is conferred by both the human-centred design of the tool and the human supervision of the recommendation inputs, particularly when dealing with disinformation classified as a systemic risk.</p> <p>Based on aggregate outputs from the AI4Trust Platform, such as Social Network Analysis (SNA), Coordinated Inauthentic Behaviour (CIB), reliability and emotional state of social media, and the classification of mis- or disinformation according to levels of severity, the platform can produce semi-automated reports tailored to end users (e.g. policy makers, journalists, researchers). These reports can compile the human-led analyses of spread of disinformation campaigns, their virality, their emotional state and reaction to it on social media platforms, and visualise the spread of disinformation. The reports will not list automatic suggestions for mitigation measures but can present a scientific overview of mitigation measures and their shortcomings. Journalists and fact-checkers can use the reports to identify relevant, viral disinformation and thus better allocate their scarce resources. Policy makers will be interested to inquire about specific topics, time frames, and locations/language (national level or EU level).</p>
<p>Target Groups/Beneficiaries</p>	<p>Project partners and future users of the platform (i.e., fact-checkers, media workers, policymakers)</p>
<p>Innovation, Exploitation and Sustainability Plan</p>	<p>Innovation resides in the hybrid component combining automated disinformation signals with human supervision and</p>



	<p>analysis, which draw on some of the principles and techniques developed in WP4. Although widely desirable, this hybrid design is rarely deployed in AI-based platforms.</p> <p>The sustainability of the tool is assured by the constant access to platform data streams as well as the existence of a critical mass of researchers, who can access the data and either analyse the output (e.g. with SNA tools) or verify the validity of prior claims. In this sense, the exploitation of the hybrid recommendation tool rests on two pillars: contextual relevance and scale. On the one side, we expect to promote the relevance and use value of the hybrid recommendation tool to end users within the context of their own work. On the other side, by scaling up the number of end users, the hybrid recommendation tool has the potential to become an independent resource for both information mediators (journalists, fact-checkers) and policy makers. The publication of scientific research outputs and policy reports will further contribute to expose and exploit the potential of the hybrid recommendation tool.</p>
--	---

## 5.17. FINCONS

**FINCONS** is a leading player in **international business consulting and system integration** in support of technological and digital transformation and in the AI4TRUST is responsible for the integration of the project Platform and **build a strategy to bring it to the market**, bringing to the projects its competencies and expertise on system integration, IoT, Big Data and Blockchain. FINCONS is leader of WP5, dedicated to the **design and development of the AI4TRUST Platform** and provision of **suitable technical solutions for assuring EU-GDPR and ethics compliance**.

Asset	<b>Fight dis/mis/malinformation technological issues</b>
Type	Knowledge
Description	Fincons provides to the project its expertise on media domain in terms of platform, regulations, and data. This knowledge will be enriched by specific topics, barriers, and facilitators (not only technological) dedicated to dis/mis/malinformation.



Target Groups/Beneficiaries	Researchers and professionals
Innovation, Exploitation and Sustainability Plan	As an industrial partner, FINCONS will use the acquired knowledge from the AI4TRUST outcomes to support researchers and customers in facing dis/mis/malinformation challenges, first integration into existing workflows of new AI technologies and tools.

Asset	<b>AI4TRUST Platform integration</b>
Type	Technology
Description	Integrating the AI4TRUST Platform means to integrate novel AI models, technologies and data in a unique framework dedicated to different stakeholders. The technical skills of FINCONS as a system integrator will provide the correct methodologies and instruments to create a consistent and reliable platform.
Target Groups/Beneficiaries	Professionals
Innovation, Exploitation and Sustainability Plan	As other industrial companies in the Consortium (EURACTIV, SKYTG24, ADB, EMS), FINCONS will exploit the pilots results and deployed solutions to increase the level of quality of product and services of their portfolio, to fight dis/mis/malinformation, and to empower scientific researchers, media practitioners and policy makers with advanced AI-based technologies.

Asset	<b>Human Validation Tool</b>
Type	Technology
Description	The Human Validation Tool integrated in the web application of the AI4TRUST platform will complete the fact-checkers workflow, allowing them to provide manual feedback about the news items passed through their monitoring and analysis activities. This enables an hybrid approach where the AI modules



	provide filters and insights for fact-checking, while the fact-checkers can manually evaluate the news items, ensuring the high level of reliability guaranteed by the human-in-the-loop.
Target Groups/Beneficiaries	Fact-checkers/journalists
Innovation, Exploitation and Sustainability Plan	As an industrial company, FINCONS will exploit this asset in conjunction with the other results of the pilots, in order to enhance the quality of its products and services, empowering media practitioners, fact-checkers and journalists to fight dis/mis/malinformation.



## 6. Initial Innovation, Exploitation and Sustainability Strategy

This section describes the initial strategies for innovation, exploitation, sustainability, engagement, and intellectual property rights management adopted by the AI4TRUST project.

The defined innovation and exploitation strategy is based on the internal and external stakeholder feedback implementation to enhance marketability and sustainability. This is achieved by adapting different approaches for professionals, researchers, and policy-makers to address unique needs and barriers.

The proposed sustainability plan foresees a long-term analysis of threats/opportunities, cost evaluation and proposals of integrated solutions. This plan is supported by an engagement strategy through which a constant dissemination of knowledge, establishment of partnerships, and multi-project initiatives are planned. Moreover, Intellectual Property Rights (IPR) management is granted by a transparent and ethical utilisation of project outcomes by mutually agreeing on terms for asset use and conducting systematic Intellectual Property (IP) reviews.

Overall, these strategies are designed to maximise the impact of the AI4TRUST project, fostering collaboration, responsible utilisation of project outcomes, and a long-term sustainability within relevant communities.

### 6.1. Innovation and Exploitation strategy

This activity involves **the integration of relevant feedback from stakeholders** to enhance the **exploitation and marketability of the results achieved** through the AI4TRUST project to ensure sustainability of the results over the time. This process includes conducting both **internal and external interviews** to gather insights, opinions, and suggestions from key stakeholders such as professionals, researchers, and policymakers. Internal interviews may involve discussions with project team members, researchers, and developers involved in the AI4TRUST project. These interviews aim to gather perspectives on the strengths, weaknesses, opportunities, and challenges associated with the developed tools, technologies, and strategies within the project. External interviews, on the other hand, involve engaging with external stakeholders such as professionals in the media industry, fact-checking organisations, policymakers, and other relevant entities. The goal is to obtain external viewpoints on the practicality, effectiveness, and market demand for the outcomes of the AI4TRUST project. By integrating feedback coming from internal and external interviews, it will be possible to identify **barriers** that may fight the successful exploitation and adoption of the project results, as well as the **facilitators** that can enhance results marketability. The **outcomes of interviews** conducted with internal and external stakeholders, coupled with the identification of barriers and facilitators, mark the **initial stage in driving the exploitation and**



**innovation plan** at the project's overarching level. It's crucial to monitor this activity throughout the project duration to enable **necessary adjustments, enhancements, or customisations** based on real-world feedback received. This **iterative process** ultimately enhances the likelihood of successful exploitation and ensures the long-term sustainability of project outcomes.

This way, the AI4TRUST project has designed a tailored approach to cater to the unique needs of these varied groups, guaranteeing that the results are optimally advantageous for each segment:

- For **professionals** involved in the media industry, fact-checking organisations, and journalism, the AI4TRUST project focuses on providing tangible tools and solutions that **enhance their capabilities in combating disinformation, misinformation, and malinformation**. This includes the development and deployment of advanced AI technologies, such as debunking tools and content verification methods, which can be seamlessly **integrated into their existing workflows**. Moreover, **APIs and plugins** will be provided by the AI4TRUST project to facilitate easy integration into partners' news production pipelines, allowing professionals to incorporate AI-driven solutions without significant adjustments to their established processes. The aim is to empower professionals to navigate the challenges posed by dis/mis/malinformation effectively, thereby elevating the overall quality and reliability of their work and offering a more dependable and fact-checked news service to a worldwide audience.
- For **researchers**, the project prioritises the advancement of **research, educational, and programs**. For instance, through the integration of AI4TRUST results into academic curricula, the project endeavours to furnish upcoming researchers with state-of-the-art knowledge and capabilities in the realm of dis/mis/malinformation detection. This methodology guarantees a steady stream of proficient researchers who contribute to the sustained endeavours in countering dis/mis/malinformation. Furthermore, the partnership with academic partners promotes **interdisciplinary research**, nurturing a more profound comprehension of the collaborative and evolving mechanisms involved in generating and disseminating problematic online content. Moreover, the academic partners will materialise the gained scientific knowledge and technological experience through **new scientific papers** in high-impact venues for the project. Finally, academic institutions play a pivotal role in disseminating knowledge and **training future professionals and researchers**. Simultaneously, the project aims to enhance the **technological transfer** potential for local companies. This ensures that the knowledge produced by the project is not only shared through **educational channels** but also implemented in **real-world contexts**.
- For **policymakers**, acknowledging the influential role played in shaping the **regulatory framework concerning disinformation, misinformation, and malinformation**, the project strives to offer invaluable **insights and recommendations**. This entails consolidating research discoveries and policy implications, rendering them easily understandable and relevant for policymakers. Through active involvement with policymakers, the project



endeavours to aid in the formulation of well-informed and impactful policies aimed at tackling the issues presented by dis/mis/malinformation in the digital space.

In addition to these, the market analysis and the interaction with business experts will help foster concrete exploitation opportunities by identifying specific business needs and gaps, and discussing how AI4TRUST technologies can deliver value to specific customer segments. The AI4TRUST project in fact, presents an important chance for Consortium partners interested in pursuing commercial exploitation opportunities by licensing the developed tools and offering their expertise to other media organisations encountering similar challenges. Based on these analyses, specific business cases will be explored to be able to identify a Minimum Viable Product (MVP) that can be further shaped for adopting a commercial strategy.

The interaction with internal and external experts is essential also to identify a viable high-level business model to be adopted for commercialization. Iterative refinements and assessments will be essential to ensure that the identified solution can be applied in specific markets and that a clear value proposition is developed for different customer segments, leveraging the key features developed throughout the project.

Besides developing an MVP, a range of viable business models will be evaluated to identify possible ways to sustainably monetize it. Possible business models may be based on different strategies, such as:

- pay per use;
- licence fees;
- commissions on transactions.

Such business model strategies are extensively used in products and services available nowadays. Examples of services that use a pay per use strategy are cloud providers, such as Amazon AWS or Microsoft Azure platforms; licence fee model are entertainment platforms for movie streaming, such as Netflix or Disney Plus, while example services that use the commissions on transactions model are concert ticket booking platforms that apply a pre-sales commission to the price of concert tickets. Similar strategies can be applied to the AI4TRUST platform as a whole, or on specific components, depending on the specific solution.

## 6.2. Sustainability Plan Outline

The sustainability plan will delineate key measures aimed at ensuring the project's results and outputs remain beneficial for long-term sustainability beyond the project's duration, as well as strategies to amplify their impact over time.



The plan will encompass various components, including an analysis of potential threats and opportunities concerning post-project sustainability, the evaluation of the costs, the marketability of the solution and recommendations for the Consortium to ensure sustainability and maximise project outcomes.

Furthermore, the plan will also investigate the future integration challenges that may arise beyond the duration of the AI4TRUST project, such as:

- the **volume of data** to be processed, which depends on the selected sources, topics and filtering criteria;
- the needed **computational power**, which is related to the amount of data and the type of selected analyses and that might be optimised according to specific requirements;
- the integration with **other data sources**, which is needed to extend the monitoring capabilities of the platform, but that also depends on API availability and legal requirements that might change over time;
- the **modularity** of the system, which would enable the possibility to deploy customised versions of the platform or exploit a sub-group of components;
- the **remuneration** of the involved parties, which shall be carefully agreed in order to find a balance that address the specific needs of the involved parties but that also enable enough flexibility for the exploitation;
- the **interoperability** with third-party applications, which requires the analysis of different technical aspects such as the use of common data models, the authentication and authorisation procedures, the approaches for monetisation, etc.

These integration challenges will be then analysed and the potential mitigations and improvements will be proposed and possibly implemented during the AI4TRUST project, to increase the impact and pave the way for the future developments.

### 6.3. Engagement strategy

The **long-term engagement** of AI4TRUST implies a constant and targeted sharing of dissemination knowledge and sensibilization through the **networks and communication channels** defined within the project, together with the support from related projects such as the **Horizon Europe project AICODE**<sup>7</sup> (featuring FBK, CERTH, and EURACTIV as partners), the **CERV project HATEDEMICS**<sup>8</sup> (led by FBK with involvement from MALDITA, DMGG, and SAHER), and series of

---

<sup>7</sup> [AI-CODE](#)

<sup>8</sup> [Hatedemics](#)



forthcoming project proposals set for submission under the **Horizon Europe Clusters 2<sup>9</sup>, 3<sup>10</sup>, and 4<sup>11</sup>** funding calls. Partners can contribute effectively to the overall engagement by:

- Sharing knowledge and insights from AI4TRUST with its existing company network, which includes end-users, industry professionals, and academic institutions from Europe and beyond. This activity will be done proactively by forwarding constantly the information on the outputs and deliverables as they evolve over the duration of the project.
- Leveraging the advantages of engaging with this network to establish new partnerships and collaborations with companies that share a mutual interest and expertise in the AI4TRUST domain, fostering opportunities for future research and operational benefits.

The objective is to create new opportunities for the future project's active involvement in **various research and innovation projects**, as well as **other forms of collaboration** with research and development institutions and industrial partners in Europe and globally. This goal is broken down into three specific short-term objectives:

- to build a **network of potential academic and research partners** (European universities and research centres). The potential academic partners will, in the first phase, be selected according to the strongest performing fields of research within each institute.
- to establish a **network of potential industrial partners**.
- to create a **network of potential professional partners and policy makers** (including end-users)

With this perspective, the project provides a unique opportunity to **engage with other European organisations** and discuss their practices aimed at reducing the presence of false information in the digital space.

Additional initiatives such as the connection to other European projects can enhance the dissemination efforts. For instance, projects like **TITAN<sup>12</sup>** and **Vera.ai<sup>13</sup>** can provide valuable insights into the innovative approaches and the unique value of the AI4TRUST outcomes. In this context, it is important to highlight the **emerging synergy between projects** addressing similar issues, which we hope will be expanded to create a solid network of collaborators, facilitating knowledge-sharing, dissemination activities and results among interested entities. This is the central objective of **“Meet the Future of AI”**, a Horizon Europe Policy and Innovation initiative

---

<sup>9</sup> [Cluster 2: Culture, Creativity and Inclusive society - European Commission](#)

<sup>10</sup> [Cluster 3: Civil security for society - European Commission](#)

<sup>11</sup> [Cluster 4: Digital, Industry and Space - European Commission](#)

<sup>12</sup> [TITAN Project](#)

<sup>13</sup> [Vera.ai](#)



organised by "sister projects" such as **AI4TRUST**<sup>14</sup>, **AI4Media**<sup>15</sup>, **AI4Debunk**<sup>16</sup>, **AI-CODE**<sup>17</sup>, and the previously mentioned **TITAN** and **Vera.ai** projects. This initiative has already resulted in a dedicated event "**Meet the Future of AI: Countering Sophisticated & Advanced Disinformation**"<sup>18</sup> held in Brussels in June 2023 and the publication of a **White Paper titled "Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities"**<sup>19</sup> in February 2024.

In addition, as part of AI4TRUST's engagement efforts, a **Community of Practice** will be established until M38. This community will capitalise on the project's accumulation of knowledge and practical experiences to enhance its expected impacts. FINCONS as task leader, with the support of EURACTIV as WP7 Leader, will lead this initiative by identifying stakeholders interested in AI4TRUST's topics and outputs both during and beyond the project duration. Leveraging the AI4TRUST Europe-wide network as well as the extensive networks of the Consortium partners, FINCONS will foster the engagement of media professionals, fact-checkers, IT companies, NGOs/CSOs, and their networks across numerous countries, in order to collect feedback on AI4TRUST based on their experiences and to understand their potential interests in the platform and/or its components.

## 6.4. IPR Management plan

The plan recognises the significance of **addressing legal and Intellectual Property Rights (IPR)** matters at the project level. The IPR policy is crafted to address potential challenges and facilitate the responsible and equitable utilisation of project outcomes. By establishing a framework for legal considerations and IPR tracking, the project ensures a transparent and ethical approach to innovation and knowledge dissemination.

For all assets generated throughout the project, it is important to come to a joint agreement on terms and conditions for the use of generated results. In particular, further collaboration activities will be carried out with all partners to list and describe IPR to avoid conflicts and at the same time to support individual or joint exploitation plans for all consortium members.

To do so, a systematic review of all generated assets will be conducted, clarifying for each asset its owner(s), the software components implementing the related functionalities, and the dependencies from other components. This step will be carried out with a collaborative approach,

---

<sup>14</sup> <https://ai4trust.eu/>

<sup>15</sup> <https://www.ai4media.eu/>

<sup>16</sup> <https://ai4debunk.eu/>

<sup>17</sup> <https://aicode-project.eu/>

<sup>18</sup> [Meet the Future of AI programme](#)

<sup>19</sup> [Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities](#)



so that all partner members can have a clear view of the IP generated throughout the project and of the terms and conditions for use and access of project assets.



## 7. Conclusions

This **Deliverable D7.2** is the **first deliverable outlining sustainable outcomes of the AI4TRUST project**, presenting preliminary exploitation paths for identified assets and their individual plans. It paves the way for the final plan, namely **Deliverable D7.4**, which will serve as a guideline for the innovation, exploitation, and sustainability of AI4TRUST offerings beyond task 7.2 lifespan.

**Chapter 2** has provided the background market analysis and an initial positioning where the AI4TRUST platform aims to position itself. **Chapter 3** has described the methodology that was used for identifying the more detailed plans discussed in Chapters 4, 5 and 6. **Chapter 4** has provided an high-level functional overview of the AI4TRUST platform, clarifying its goals, its vision, its main functionalities, its key exploitable assets and the preliminary benefits identified for the stakeholders. **Chapter 5** delved into individual plans for innovation, exploitation, and sustainability for each asset of each consortium partner. **Chapter 6** outlined a macro-level approach and methodology for the initial innovation, exploitation, and sustainability strategy of the project results, while **Chapter 7** provides a summary of conclusions.

The overview of the AI4Trust platform in Chapter 4 presents the main functionalities of the AI4TRUST project in light of the market analysis presented in Chapter 2, clarifying the clear potential that such kinds of solutions have in the current market context. Key exploitable results are presented and linked with potential benefits for different stakeholders groups. Furthermore, individual assets for each partner are presented in Chapter 5, and individual innovation, exploitation, and sustainability plans are outlined systematically, providing a clear roadmap for the exploitation of project outcomes. Preliminary benefits for stakeholders have been identified, contributing to a **better understanding of the potential impact of the AI4TRUST project**.

Besides individual plans, an overall initial innovation, exploitation and sustainability strategy is then presented in Chapter 6 to evaluate marketability opportunities and to ensure sustainability of obtained results. An engagement strategy is outlined to foster continuous collaboration and networking between internal and external stakeholders, contributing to increase the impact of the project, complemented by an Intellectual Property Rights (IPR) management plan to support individual and joint exploitation of project results.

In conclusion, this **framework for the innovation, exploitation, and sustainability strategy for future outcomes of AI4TRUST** provides a forward-thinking perspective by presenting **initial guidelines and methodologies** for the overarching project innovation, exploration, and sustainability plan that will be followed throughout Task 7.2, contributing to the impact maximisation in the fight against disinformation, misinformation, and malinformation.