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

www.ai4trust.eu

AI4TRUST

AI4TRUST

D6.5 - PILOT PLANNING REPORT – REVISED VERSION

PARTNERS

FONDAZIONE BRUNO KESSLER
CERTH CENTRE FOR RESEARCH & TECHNOLOGY HELLAS
UNIVERSITÀ DI TRENTO
NATIONAL CENTRE FOR SCIENTIFIC RESEARCH "DEMOKRITOS"
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE
GDI Global Disinformation Index
sky
SAHER EUROPE
DEMAGOG
MALDITA.ES
ASTIKI MI KERDOSKOPIKI ETAIRIA KENTRO KATAPOLEMISIS TIS PARAPLIROFORISIS / CIVIL NON-PROFIT COMPANY KENTRO KATAPOLEMISIS TIS PARAPLIROFORISIS
EURACTIV
ASOCIATIA DIGITAL BRIDGE
EUROPEJSKIE MEDIA SP ZOO
FINCONS GROUP
UNIVERSITY OF CAMBRIDGE

Document information

Title	D6.5 - Pilot planning report - Revised	
Editor	NCSR - D	
Contributors	MALDITA, DEMAGOG, SKYTG24, ELLENIKA, EURACTIV, ADB, EMS, UCAM, POLITEHNICA, FBK, CERTH, UNITN, FINC	
Dissemination Level	☐ CO: Confidential, only for members of the consortium (including the Commission Services) ☐ RE: Restricted to a group specified by the consortium (including the Commission Services) ☐ PP: Restricted to other programme participants (including the Commission Services) ☒ PU: Public	
Reviewers	☒ FBK ☐ CERTH ☐ UNITN ☐ NCSR ☐ CNRS ☐ POLITEHNICA ☐ SAHER ☒ GDI	☐ SKYTG24 ☐ SKYTG24 ☐ ELLENIKA ☐ EURACTIV ☐ SKYTG24 ☐ ADB ☒ DEMAGOG ☒ MALDITA
Status	☐ Draft ☐ WP Manager accepted ☒ Coordinator accepted	
Due date	28 February 2025	
Delivery date (resubmission)	27 February 2025	

Summary of modifications

VERSION	DATE	AUTHOR(S)	SUMMARY OF MAIN CHANGES
0.1	22/06/2023	M. Gerosa (FBK), S. Bressan (FBK), M.V. Zucca (FBK)	First draft of D6.1
0.2	13/07/2023	A. Koukourikos (NCSR-D), P. Krokidas (NCSR-D), V. Karkaletsis (NCSR-D)	Second draft of D6.1
0.5	18/08/2023	M. Preoteasa (ADB), X. V. Barillas (Maldita), J. Hannay (Euractiv), K. Wysocka (EMS), A. Koutroumpelis (Ellenika), A. Dambrosio (Sky TG24), A. Szymkiewicz (Demagog)	Sections 2, 3
0.6	25/08/2023	G. Neff (UCAM)	Section 4
0.7	31/08/2023	A. Koukourikos (NCSR-D), P. Krokidas (NCSR-D), V. Karkaletsis (NCSR-D)	Draft finalisation
0.8	08/09/2023	Zoe Fourel (GDI)	Internal review
0.9	14/09/2023	A. Koukourikos (NCSR-D), P. Krokidas (NCSR-D), V. Karkaletsis (NCSR-D)	Deliverable finalisation
1.0	15/09/2023	S. Bressan (FBK)	Final review - Deliverable ready for submission
1.1	22/05/2024	A. Davvetas (NCSR-D)	Review Revision

1.2	23/05/2024	R. Gallotti (FBK) X. V. Barillas (Maldita)	Internal Review
1.3	29/05/2024	E. Apostolidis (CERTH)	Internal Review
2.0	29/05/2024	A. Davvetas (NCSR-D), G. Petasis (NCSRD-D)	Deliverable finalisation
2.1	17/02/2025	A. Davvetas (NCSR-D), D. Katsamori (NCSR-D), G. Petasis (NCSRD-D)	D6.5 Revision
2.2	17/02/2025	Y. Delimaris (Ellinika) S. Felsberger (UCAM)	Input in Section 2, 4, and 5
2.3	21/02/2025	D. Giampiccolo (FBK) R. Gallotti (FBK) L. Livdane (GDI)	Internal review
2.4	26/02/2025	C. O. Abascal (Maldita)	Feedback in Section 2 (fact-checking state of the art)
2.5	27/02/2025	A. Davvetas (NCSR-D), D. Katsamori (NCSR-D), G. Petasis (NCSRD-D)	Deliverable finalisation
3.0	27/02/2025	R. Gallotti (FBK) S. Bressan (FBK)	Final review, formatting and checking by AI4TRUST Coordinator

History of changes

The deliverable **D6.1 - "Pilot Planning Report"** from Work Package 6 of the AI4TRUST project, originally due on 31 August 2023, has been conditionally accepted. However, it must undergo revisions and be resubmitted as **D6.5 – "Pilot Planning Report – Revised Version"** by 28 February 2025. These revisions must address several key recommendations outlined in the **GENERAL PROJECT REVIEW CONSOLIDATED REPORT (HE)**, dated 28 June 2024.

One of the primary concerns raised in the review is the unclear alignment of the **HLA** and its components with the different pilots, with the descriptions of the pilots being considered too brief and lacking detailed information. This has an impact on the assessment of the pilot's potential outcomes. Additionally, it was noted that the term **"fake news"** is used in D6.1, despite recommendations in **WP2** and **WP4** to avoid this terminology, given its imprecise nature.

The report also introduced **OpenAI tools** and **Large Language Models (LLMs)**, but more detailed elaboration is required on how these technologies are integrated into the project. Furthermore, while the number of testers is mentioned, there is no clear evidence provided regarding the exact number, which affects the credibility of the testing plan.

Another issue is the lack of clarity regarding the time duration of the phases in all pilots. Furthermore, while the pilot descriptions focus on the participants and their business contributions, they provide little detail on the actual activities and functional aspects of the pilots themselves.

It was also recommended that an **integration leader** be appointed to ensure a clear definition and management of the various actions required across the project. Moreover, it was suggested that the **ISO/IEC 25010:2011** model be adapted to better suit the specific needs of the AI4TRUST platform.

The report was also criticised for not adequately developing **Key Performance Indicators (KPIs)** to measure the success of the platform's features. It is important that these KPIs focus on tangible outcomes, and their development should be closely linked to the **Key Exploitable Results (KERs)**. In conclusion, the revision of D6.1 will need to address these points to ensure the pilot planning and execution phases are well-defined and properly aligned with the project's objectives and goals.

History of changes from V1.0 to V1.1

- **Section 1** has been revised to incorporate newly added material.
- **Section 2** is now a unified and summarised version of the previous state-of-the-art sub-sections (2.1.1, 2.2.3, 2.3.1, 2.4.1, 2.5.3, 2.6.3, and 2.7.3).
- **Section 3.1** consolidates and summarises the previous stakeholder engagement sub-sections (2.1.3, 2.2.2, 2.3.3, 2.4.3, 2.5.2, 2.6.2, and 2.7.2), along with the introductory paragraphs from sections 2.1 to 2.7.
- **Section 3.2** is a newly added section, primarily based on the high-level requirements outlined in the previous sub-sections (2.1.2, 2.2.1, 2.3.2, 2.4.2, 2.5.1, 2.6.1, and 2.7.1).
- **Sections 4 and 5** are newly introduced to address reviewer comments.
- **Section 6** is now a unified and summarised version of the previous Section 3, as well as individual sub-sections 2.1.4, 2.2.4, 2.3.4, 2.4.4, 2.5.4, 2.6.4, and 2.7.4.
- **Section 7** is newly added.

History of changes from V1.1 to V1.2

- **Section 7.1.1** has been revised to remove the term “fake news,” recognising it as an oxymoron, and to address the relevant reviewer comment.
- **Section 2** has been updated to provide additional context on EH’s use of LLMs, with rewording for improved clarity and to address the reviewers’ feedback.
- **Section 7.1.1** has been further revised to incorporate the reviewers’ comments regarding the estimated number of participants.
- The description of the **"Trust"** characteristic in the AI4TRUST Platform quality model has been expanded.
- The **evaluation scenario** description in **Section 5.1** has been enhanced.
- The subsections for **Phase 1 and Phase 2** have been further developed.

Table of contents

Executive summary

This deliverable of the **"AI4TRUST - AI-based technologies for trustworthy solutions against disinformation"** project, titled **"D6.5 – Piloting Planning Report -Revised version"**, is an updated version of the previously submitted "D6.1 – Piloting Planning Report", **incorporating the recommendations received on 28 June 2024** from the experts and the PO following the project's first Review Meeting. This deliverable is part of **Work Package 6 "Piloting, Assessment & Fact-checking"** (hereinafter referred to as WP6) led by DEMOKRITOS (NCSR-D).

Piloting is a key component of the AI4TRUST project, designed to address challenges across diverse contexts, countries, and types of misinformation and disinformation. A well-structured piloting plan is essential to meeting the needs of various stakeholders while ensuring that pilot findings translate into a coherent set of requirements for the AI4TRUST Platform.

This document first presents a **state-of-the-art analysis**, which serves as the foundation for pilot design by providing a comprehensive understanding of existing technologies, methodologies, and best practices within the AI4TRUST ecosystem. This analysis helps identify gaps and opportunities, enabling project partners to build upon current knowledge and advance relevant solutions. The identified gaps inform the formulation of **evaluation requirements**, which serve as guiding principles throughout the pilot execution to ensure alignment with the project's objectives.

Next, the document outlines the **pilot specifications**, detailing the pilot requirements that highlight existing gaps and necessary improvements for the AI4TRUST Platform. It also introduces the **AI4TRUST Platform Quality Model**, which defines key quality characteristics and sub-characteristics, along with the corresponding **Assessment Plan**. The deliverable further specifies the methodologies for evaluating these characteristics, use case scenarios for **quality assessment**, and indicative **KPIs** to measure the success of the piloting sessions.

Additionally, the document describes activities related to the **assessment and validation** of the AI4TRUST Platform. These include **preparatory activities** to ensure a smooth piloting phase, **execution activities** detailing the piloting process, and **post-pilot activities** aimed at stakeholder engagement and community-building.

The proposed piloting strategy, along with **end-users and participants recruitment plans**, provides responsible partners with a structured framework to effectively organise their preparatory efforts. Therefore, the AI4TRUST Platform will undergo **testing, validation,**

and evaluation by a targeted user group of journalists and fact-checkers from partner organisations, including DEMAGOG, MALDITA, ELLINIKA, SKY TG24, EURACTIV, ADB, and EMS. Additionally, further end-users will be engaged through **the partners' extensive networks**, ensuring coverage across multiple countries and EU languages. Following an **iterative approach**, the Platform will be continuously refined based on feedback from end-users after **three testing rounds**. Furthermore, it will be **co-designed and assessed** by **policymakers and government officials** through dedicated workshops organised by UCAM (WP6).

The piloting strategy not only establishes a **foundation for the project's technical developments** but also the creation of essential documentation to **support pilot partners in engaging with relevant stakeholders**. This, in turn, contributes to the **sustainability of the project** and its output in the short, medium, and long term. The AI4TRUST Platform can serve as a valuable resource beyond the project's duration, providing media professionals and fact-checkers with advanced tools to detect and analyse misinformation and disinformation (hereinafter also mis/disinformation). Additionally, policy makers and other potentially interested stakeholders can leverage the platform's insights to inform regulatory decisions and shape evidence-based policies. For a more detailed analysis of the project's innovation, exploitation and sustainability strategy, please refer to **D7.4 - Innovation, Exploitation, and Sustainability Plan v2**.

1. Introduction

The primary objective of **WP6 "Piloting, Assessment & Fact-checking"** is to pilot and validate the **AI4TRUST Platform**, as well as to design and implement effective fact-checking and validation activities for media professionals, researchers, and policymakers. These activities aim to enhance their ability to **monitor, detect, and document mis/disinformation** across online social media and traditional media, while also facilitating the **creation and dissemination of reliable information**. To achieve this, a well-defined **piloting plan**, supported by a structured feedback collection and evaluation methodology, is essential for assessing the **quality and effectiveness** of the AI4TRUST solutions.

As part of this effort, **Task T6.1 – Piloting Requirements, Coordination, and KPI Definition** plays a central role in guiding all WP6 activities. It establishes the **scope and objectives** of the piloting sessions, defines the mechanisms for **collecting and processing feedback**, and sets the **framework for evaluation**, ensuring that the AI4TRUST Platform evolves in alignment with user needs and project goals.

Furthermore, this task introduces the **AI4TRUST Platform Quality Model**, which **adapts key characteristics** from the widely recognised **ISO/IEC 25010 standard**. By outlining **evaluation scenarios and defining key performance indicators (KPIs)** for future piloting sessions, it sets the foundation for **testing methodologies and success criteria**. In addition, the document presents **complementary activities** beyond testing, such as **preparatory actions**, to ensure the seamless execution of the piloting process.

This deliverable, titled **"D6.5 – Piloting Planning Report – Revised"**, is part of the **AI4TRUST** project. It serves as an updated version of the previously submitted **"D6.1 – Piloting Planning Report"**, integrating the recommendations provided by the experts and the Project Officer (PO) on 29 July 2024, following the project's first Review Meeting. The document is structured as follows:

- **Section 2** examines the **current state of fact-checking**, reviewing best practices and relevant literature.
- **Section 3** defines the **pilot specifications**, situating them within the broader context and extracting **pilot requirements** from the fact-checking landscape.
- **Section 4** presents the **AI4TRUST Platform Quality Assessment**, including an introduction to **ISO/IEC 25010** and its adaptation into the **AI4TRUST Platform Quality Model**.
- **Section 5** details the **testing plan**, outlining evaluation methods for platform characteristics, **test scenarios**, and **KPIs**, which serve as success criteria.

- **Section 6** describes **additional piloting-related activities**, covering the distinct **preparation, execution, and post-pilot phases**.
- **Section 7** lays out the **piloting and stakeholder engagement plan**, including the project timeline.
- **Section 8** presents the **conclusions** of the report.

This structured approach ensures that the piloting activities are conducted effectively, providing valuable insights that will contribute to the refinement and validation of the AI4TRUST Platform.

2. Fact-checking: Analysis of State-Of-The Art and Best Practices

Fact-checking methodologies can be broadly classified into **two distinct approaches**:

1. **Human-Centric Fact-Checking**, which relies on extensive **collaboration among media professionals**. This **manual approach** requires media professionals to **cross-check claims against available evidence**, benefiting from their **deep understanding and expertise**. Journalists verify information manually, often consulting **databases of known disinformation**. However, this **time-consuming and labour-intensive** process can become a **bottleneck** in an era of rapid information dissemination, as workloads may increase exponentially.
2. **Automated Fact-Checking**, which leverages **technological tools and algorithms** to improve efficiency. Fully automated solutions use **statistical models, network analysis, and machine learning** to address the scalability challenges of human-centric approaches. These tools can **process vast amounts of data simultaneously** and identify **patterns of disinformation**, enabling **automated content flagging**. However, they **lack human judgment**, are prone to errors, and often produce results that can be **difficult to interpret**.

According to the literature and the practitioners involved in the AI4TRUST consortium, the **ideal approach** is a **semi-automated model**. Most media organisations adopt a **hybrid strategy**, combining **human expertise** with **technological tools** to enhance **speed** while maintaining **accuracy** and **reliability**. This method allows **fact-checkers and journalists** to leverage **automation** for greater **efficiency**, while ensuring that **human judgement** remains central to the **verification process**.

Within the **AI4TRUST** consortium, different members implement various **fact-checking strategies**, ranging from **semi-automated in-house solutions** to **big tech-supported approaches** and fully **manual verification methods**. The following examples illustrate how consortium members integrate these **methods** to assess the **veracity of information** effectively. The following examples illustrate the **different fact-checking approaches** used by AI4TRUST consortium members:

- **In-house Semi-Automated Approach:** Maldita and Demagog;
- **Big Tech-Supported Semi-Automated Approach:** SkyTG24 and ELLINIKA;
- **Fully Manual Approach:** EURACTIV, ADB and EMS.

1. Maldita

Maldita employs a **semi-automated fact-checking system** through its **WhatsApp chatbot** and **Disinformation Management System (DMS) database tool**. These tools collect and process **potential disinformation reported by users**, allowing for **real-time monitoring**. However, the actual fact-checking remains a **manual process**, adhering to **strict editorial standards** and involving collaboration with **verified experts**.

2. Demagog

Demagog focuses on the **development of fact-checking technologies**, such as an **educational platform** and a **climate chatbot** for verifying climate change information. Despite these tools, the **media scanning, information verification, and analysis** are performed **manually by trained analysts** following **rigorous editorial standards**.

3. ELLINIKA

ELLINIKA utilises **Meta's Workplace** to **collect claims requiring verification**, aggregating data from **community messages, manual searches, and Meta's TPFC platform**. Their **hybrid approach** involves fact-checkers assessing claims through **Google Search, Reverse Image Search, Maps, Translate, and AFP resources**. Additionally, **Large Language Models (LLMs)** assist in **semantic searches** and **data analysis** to guide fact-checkers.

4. SkyTG24

SkyTG24 integrates **fact-checking into its TV programmes and digital journalism**, though not on a continuous basis. It performs **debunking activities** on major news topics, such as **COVID-19 misinformation** and the **war in Ukraine**. This process relies on **specialised agencies** (e.g., Storyful, F5) and **big tech tools** such as **Google Search, TinEye, CrowdTangle, and Whois**.

5. EURACTIV

As a **media network**, EURACTIV conducts **continuous fact-checking activities**. With the rise of **LLMs in journalism**, it explores their use for **digital content verification**. However, due to **concerns over transparency and copyright**, their deployment is carefully balanced to **mitigate risks** while enhancing content accuracy.

6. ADB

ADB highlights the **gap between journalism and fact-checking**, performing **fact-checking activities intermittently**. Since 2019, it has run the **"Facts, Not Fake"** section, exposing **major disinformation narratives** in **Romania and Moldova**. ADB collaborates with **fact-checking organisations** (e.g., Funky Citizens, Misreport, Stop Fals) and **monitors EUvsDisinfo**. Its **primary fact-checking approach remains manual**, relying on **journalistic methods**.

7. EMS

EMS does not conduct **continuous fact-checking** but acknowledges the **complexity of countering disinformation**. A study by the **Reuters Institute** found that **44% of Polish internet users encounter false news at least once a week**, highlighting the ongoing challenge of combating misinformation. When EMS engages in fact-checking, it primarily **relies on manual methods**.

The **fact-checking landscape** is characterised by a mix of **manual, semi-automated, and fully automated approaches**. While human expertise ensures accuracy and depth, technological tools enhance efficiency and scalability. The AI4TRUST project leverages **a hybrid fact-checking model** to **combine the strengths of both methodologies**, ensuring **high-quality verification processes** across multiple EU countries and languages.

3. Pilot Specifications

This section examines the **fundamental components** that underpin a **comprehensive and effective pilot design**. Building upon the **state-of-the-art** in fact-checking, the design establishes a set of **pilot requirements** for the validation of the **AI4TRUST Platform**, whose **technical feasibility** will subsequently be assessed by the **consortium end-users**. A more in-depth analysis of the **pilot participants, pilot scenarios**, and **expected outcomes** is presented in **Sections 6 and 7**.

3.1. Pilot Requirements

The analysis of the **state-of-the-art in fact-checking**, alongside **oral and written communication** with consortium members, has informed the identification of key aspects that users seek to evaluate. These include the **reliability analysis of information sources**, which entails the development of a **robust framework** for the **automated assessment** of an information source's **veracity and credibility**. Moreover, tracking **infodemic trends**, particularly in relation to **topics of public interest** that may be targeted by **malicious actors**, has been highlighted as a priority. Based on what the end-user partners decided in **WP2 and WP4**, the topics that will be studied under the AI4TRUST project will be disinformation content on the topics of **climate change, health, and migrants**. In this regard, the ability to **identify and analyse patterns in the spread of mis/disinformation** is considered essential, as it would enable **stakeholders** to better prepare for and mitigate the impact of **false information**.

Additionally, the implementation of **evidence-based textual inoculation strategies** is regarded as relevant. This entails ensuring that **debunked claims** are accompanied by **contextual information** and **fact-based counterarguments**, thereby fostering **resilience** against mis/disinformation tactics. Finally, the **integration of fact-checking activities** into the **daily journalistic workflow** is deemed highly desirable. This involves the provision of **tools** that seamlessly align with **journalistic processes**, facilitating the **real-time verification** of facts.

Table 1 outlines the existing **gaps,** and the desirable **improvements** associated with the aforementioned aspects.

Property	Gaps	Desirable Improvements
Assessing the Credibility and Reliability of Information Sources	• **Manual comparison of sources**, relying solely on expert input. • **Dependence on verified experts**, limiting broader accessibility and scalability. • **Inability to detect user bias**, affecting the objectivity of assessments. • **Lack of systematic verification** and credibility assessment of sources. • **Insufficient capacity** for continuous monitoring of emerging trends. • **Limited detection** of diverse disinformation signals.	• **Automated reliability assessment** of information sources. • **Establishment of** benchmarks to determine source credibility. • **Automation of editorial standards** to ensure consistency and accuracy. • **Automated rating of webpages** containing potentially misleading claims. • **Integration of fact-checking organisations' datasets** for enhanced verification. • **Automated evaluation of authenticity** for sources originating from social media. • **Access to a comprehensive database** of debunked sources. • **Aggregation of trending topics** related to specific issues. • **Granular labelling and categorisation** of information sources.
Tracking Infodemic Trends in Key Public Interest Issues	• **AI-driven identification** of common narratives. • **Expansion of metrics**, beyond current indicators such as "frequently forwarded" and frequency reports. • **Broader integration of data sources** to enhance analysis. • **Development of an automated system**, as no such solution currently exists. • **Capability to monitor multiple channels simultaneously**, reducing reliance on manual verification.	• **Integration of multiple data sources** for comprehensive analysis. • **Quantification of infodemic risk levels** to assess the severity of mis/disinformation. • **Development of an automated system** for efficient data processing and analysis. • **Incorporation of social media monitoring** to track real-time trends. • **Integration of signals from community input** to enhance detection and response.
Incorporating Fact-Checking into Daily Journalistic Practices	• **Evaluating reliability is time-consuming**, requiring significant resources. • **Journalists often lack the time** necessary to conduct thorough fact-checking.	• **Detection of deepfakes and human-generated mis/disinformation**, along with reliability assessment. • **Access to automated fact-checking for multimodal content.** • **Support for multiple languages** to ensure broader applicability.

	• **Fact-checking tools are constrained** by financial and resource limitations. • **Debunking is either lacking or performed manually,** with time constraints and limited access to databases.	• **Automated reliability checks** with alerts, access to databases, and streamlined processes.
Evidence-Based Textual Countermeasures	**Manually developed types of verified information**	• **Automated generation of textual snippets** containing evidence and counterarguments to support the claim that certain content is mis/disinformation. • **Enrichment of fact-checker reports** with evidence generated by AI4TRUST tools. • **Monitoring of content and delivery of snippets or audiovisual material** that may be misleading. • **Identification of characteristics of false information and patterns** in mis/disinformation narratives. • **Mapping of networks of actors** responsible for disseminating false content.

Table 1: Pilot Requirements Gathered from Consortium Members

4. Assessing Platform Quality

The assessment of a platform's quality, often referred to as the **evaluation process**, is a multifaceted procedure that considers various aspects of the software, including its **validation**. Software validation is defined as *"the confirmation, through examination and the provision of objective evidence, that software specifications conform to user needs and intended uses, and that the particular requirements implemented through the software can be consistently fulfilled"*[1]. However, when discussing software, we typically refer to the individual components that form part of a broader system or platform. Consequently, the validation process is primarily concerned with **providing evidence that the software requirements have been implemented appropriately**.

The ultimate goal of validation is to obtain evidence that the system can fulfil all the **specified requirements** and deliver the corresponding **functionalities** that align with the

[1] General Principles of Software Validation; Final Guidance for Industry and FDA Staff, January 11, 2002.

users' expectations, as defined during the design phase. This process not only enables the assessment and assurance of the product's quality but also establishes a **feedback and development loop** that continuously refines and improves the final product. Software validation is achieved through the development and application of appropriate **quality models**.

4.1. ISO/IEC 25010

ISO/IEC 25010 is a standard that outlines models for system and software quality requirements and evaluation, commonly referred to as **SQuaRE**. This standard comprises two models: **"Quality in Use"** and **"Product Quality"**, each addressing different aspects of software and systems, with respective characteristics and sub-characteristics. The **ISO/IEC 25010:2011** version[2] has been widely adopted for the evaluation of platforms and systems. However, more recent iterations, such as **ISO/IEC 25010:2023**[3] and **ISO/IEC 25010:2024**[4], have since been produced.

- **Quality in Use**: This model encompasses five characteristics that pertain to the outcomes of a user's interaction with the product (e.g., computer system, platform, etc.) within a defined context.
- **Product Quality**: This model includes eight characteristics that focus on both the non-variable properties of software and the variable properties of the system.

The **characteristics** and **sub-characteristics** of the **Product Quality** model are enumerated in **Figure 1**, as the characteristics of the **Quality in Use** model can generally be assimilated into the **Product Quality** characteristics due to shared themes.

The **ISO/IEC 25010 model** is widely recognised and has had a significant impact on a diverse range of stakeholders within the software and system development industry, including developers, quality assurance teams, and others. As an **ISO standard**, it provides a foundation for **standardisation** and a common glossary for specifying, measuring, and evaluating software product quality, thereby facilitating fair comparisons across a broad spectrum of systems.

[2] https://www.iso.org/standard/35733.html
[3] https://www.iso.org/standard/78176.html
[4] https://www.iso.org/standard/78175.html

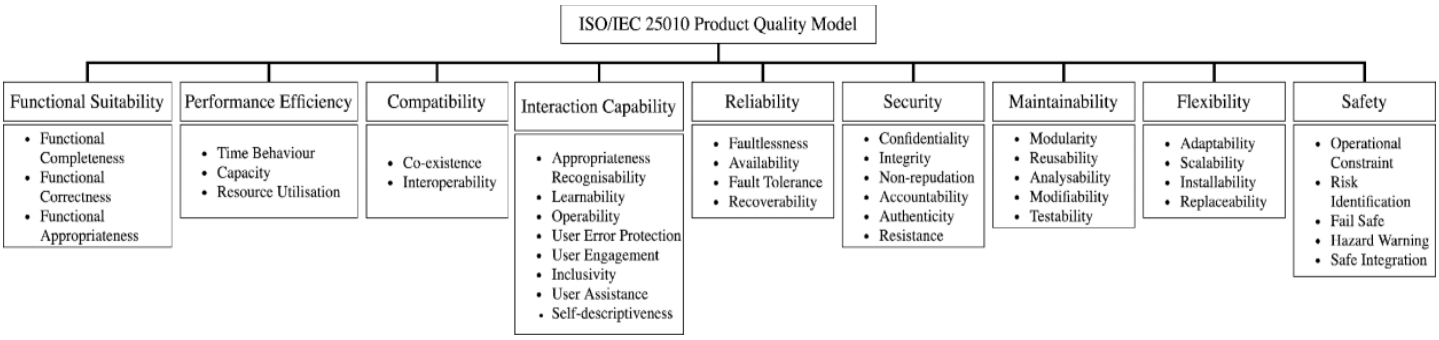

Figure 1: ISO/IEC 25010 Product Model

Furthermore, it ensures **completeness** in the comparison between quality properties and specified requirements. Additionally, the practice of adapting this model to various software evaluation methods is a widely adopted strategy, as its framework is clearly defined and easily adaptable. As a result, **ISO/IEC 25010** has significantly enhanced the ability to conduct thorough assessments of software components, systems, or platforms, leading to outcomes of **higher quality**[5].

4.2. The AI4TRUST Platform Quality Model

The **quality of the AI4TRUST Platform** will be defined as the extent to which the platform meets both the defined and implicit needs of its target end-users. The adaptation of the **ISO/IEC 25010** model facilitates the representation of these stakeholder requirements through broad, distinct groups (**characteristics**), each of which is further subdivided into specific classifications (**sub-characteristics**).

Taking into account the nature of the **AI4TRUST Platform**, its constituent components, the target stakeholders, and their potential use cases, we define a set of **characteristics** and **sub-characteristics** that are appropriate for our context. These carefully selected characteristics and sub-characteristics are collectively referred to as the **AI4TRUST Platform Quality Model**. In the following subsections, the **ISO/IEC 25010** standard has been applied to evaluate the **AI4TRUST Platform**. This is achieved by selecting the relevant characteristics and sub-characteristics from the **ISO/IEC 25010** framework and aligning them with the functionalities and use cases that the **AI4TRUST Platform** will support. This adaptation process forms the foundation of the **AI4TRUST Platform Quality Model**. The **AI4TRUST Platform Quality Model** will be consistently employed to assess the platform's quality throughout the piloting phases.

[5] https://link.springer.com/chapter/10.1007/978-3-319-97925-0_42

The **AI4TRUST Platform Quality Model** is illustrated in **Figure 2**. Following the careful selection of relevant characteristics and sub-characteristics, additional evaluation aspects have been incorporated into certain characteristics to better align with the specific context of the **AI4TRUST Platform**. These aspects are described in more detail within each respective characteristic in the following sections. The definitions provided in **italics** and **quotation marks** represent verbatim excerpts from the references cited in the corresponding footnotes.

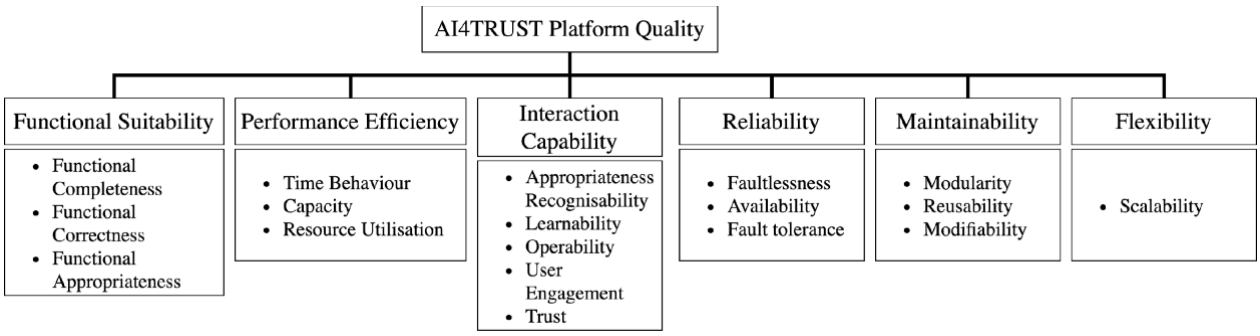

Figure 2: AI4TRUST Platform Quality Model

4.2.1. Functional Suitability

"*This characteristic represents the degree to which a product or system provides functions that meet stated and implied needs when used under specified conditions. This characteristic is composed of the following sub-characteristics:*

- ***Functional completeness**: Degree to which the set of functions covers all the specified tasks and intended users' objectives.*
- ***Functional correctness**: Degree to which a product or system provides accurate results when used by intended users.*
- ***Functional appropriateness**: Degree to which the functions facilitate the accomplishment of specified tasks and objectives."*[6]

4.2.2. Performance Efficiency

"*This characteristic represents the degree to which a product performs its functions within specified time and throughput parameters and is efficient in the use of resources (such as CPU, memory, storage, network devices, energy, materials...) under specified conditions. This characteristic is composed of the following sub-characteristics:*

[6] https://iso25000.com/index.php/en/iso-25000-standards/iso-25010

- ***Time behaviour:*** *Degree to which the response time and throughput rates of a product or system, when performing its functions, meet requirements.*
- ***Resource utilization****: Degree to which the amounts and types of resources used by a product or system, when performing its functions, meet requirements.*
- ***Capacity****: Degree to which the maximum limits of a product or system parameter meet requirements."*[7]

We will adapt the **time behaviour** sub-characteristic by incorporating the time required for a user to complete a single task. Additionally, within the **resource utilisation** characteristic, alongside the technical properties related to the system's performance, the **power consumption** of the Platform will also be taken into account.

4.2.3. Interaction Capability

"*Degree to which a product or system can be interacted with by specified users to exchange information is the user interface to complete specific tasks in a variety of contexts of use. This characteristic is composed of the following sub-characteristics:*

- **Appropriateness recognizability**: *Degree to which users can recognise whether a product or system is appropriate for their needs.*
- ***Learnability:*** *Degree to which the functions of a product or system can be learnt to be used by specified users within a specified amount of time.*
- ***Operability:*** *Degree to which a product or system has attributes that make it easy to operate and control.*
- ***User engagement:*** *Degree to which a user interface presents functions and information in an inviting and motivating manner encouraging continued interaction."*[8]

We adapt this characteristic with an additional sub-characteristic that fit our case of evaluation:

- **Trustworthiness:** The degree to which a user places trust in the overall platform is a critical factor. Trustworthiness, as a concept, was defined within the context of the project in D4.2. This deliverable provides a detailed explanation of how WP4 has defined the trustworthiness of AI systems within the AI4TRUST project, as well as how potential end-users of the AI4TRUST Platform assess the trustworthiness of AI tools in their respective work.

[7] Ibid.

[8] Ibid.

4.2.4. Reliability

"Degree to which a system, product or component performs specified functions under specified conditions for a specified period of time. This characteristic is composed of the following sub-characteristics:

- ***Faultlessness:** Degree to which a system, product or component performs specific functions without fault under normal operation.*
- ***Availability:** Degree to which a system, product or component is operational and accessible when required for use.*
- ***Fault tolerance:** Degree to which a system, product or component operates as intended despite the presence of hardware or software faults."*[9]

4.2.5. Maintainability

"This characteristic represents the degree of effectiveness and efficiency with which a product or system can be modified to improve it, correct it or adapt it to changes in environment, and in requirements. This characteristic is composed of the following sub-characteristics:

- ***Modularity**: Degree to which a system or computer program is composed of discrete components such that a change to one component has minimal impact on other components.*
- ***Reusability**: Degree to which a product can be used as an asset in more than one system, or in building other assets.*
- ***Modifiability**: Degree to which a product or system can be effectively and efficiently modified without introducing defects or degrading existing product quality."*[10]

4.2.6. Flexibility

"Degree to which a product can be adapted to changes in its requirements, contexts of use or system environment. This characteristic is composed of the following sub-characteristics:

- ***Scalability**: Degree to which a product can handle growing or shrinking workloads or to adapt its capacity to handle variability."*[11]

[9] Ibid.

[10] Ibid.

[11] Ibid.

5. Testing the AI4TRUST Platform

This section presents the evaluation aspects of the **AI4TRUST Platform Quality Model**, alongside the corresponding tests designed to assess each quality dimension. It also outlines **key evaluation scenarios** for the Platform, which validate the underlying software components, and includes preliminary **Key Performance Indicators (KPIs)** to measure **performance, reliability, and user experience**. The KPIs were developed in accordance with project-specific requirements, ensuring their alignment with the platform's functional objectives and pilot goals.

Quality characteristic	Test	Description	Preliminary KPIs
Functional Suitability			
Functional Completeness	Observation tests through usage scenarios.	Observations will assess whether the components utilised in a scenario encompass all the functionalities required to complete the relevant tasks.	Coverage of 80% of pilot requirements.
Functional Correctness	Observation tests through usage scenarios.	Observations will evaluate whether the components used in a scenario produce the expected (i.e., correct) results.	Achievement of an accuracy rate exceeding 80% for scenarios[12].
Functional Appropriateness	Observation tests through usage scenarios.	Observations will examine whether the components employed in a scenario are suitable for the specified tasks.	Attainment of an over 80% suitability rate for tasks[13].

[12] Scenario Accuracy: Correctly Executed Scenarios / Total Scenarios Tested.

[13] Task Suitability Rate: Suitable Tasks / Total Tasks Tested.

Performance Efficiency			
Time Behaviour	Performance testing: User interface and Back-end.	Different front-end and back-end tests will focus on response/processing times and throughput rates per request for a single user. Additionally, the time required to complete a single task/scenario will be recorded.	Throughput: At least 100 requests per second[14]. Response: Less than 2 seconds for front-end and back-end[15]. Lead time[16]: Decrease compared to previous methods.
Capacity	Load and stress capacity testing.	Separate tests will evaluate response/processing times and throughput rates per request for multiple concurrent users.	Response: Less than 2 seconds. Throughput: At least 200 requests per second.
Resource Utilisation	Profiling Tools, Performance Counters, and Custom Logging Mechanisms.	Other tests will focus on stress testing the system to gather data on the usage of various resources, including, but not limited to, CPU, memory, disk, network, and power consumption.	Less than 80% utilisation on resources (CPU, Memory, Disk, Network). Estimation of power consumption under idle and under stress.
Interaction Capability			
Appropriateness Recognisability	Observation tests through usage scenarios.	Observations will assess whether the components used in a scenario are appropriate for the requirements gathered from users.	Coverage of 80% of user requirements.

[14] This throughput refers to pages and APIs that do not involve AI models or the use of AI tool APIs, as these components exhibit limitations related to deployment costs, such as GPU utilisation and GPU memory availability.

[15] The same principle that applies to throughput also applies to response.

[16] Lead Time: The amount of time required for a user to complete a single task.

Learnability	Observation tests through usage scenarios.	Observations will evaluate whether the components used in a scenario can be learned effectively within a specified time frame.	Completion of tasks in less than 1 hour on average, without assistance.
Operability	System Usability Scale (SUS).	The SUS questionnaire will determine whether the platform is easy to operate.	System Usability Score (SUS) of over 68[17].
User Engagement	Design evaluation.	The User Interface (UI) will be reviewed through interactive feedback and interaction sessions, focusing on aspects such as navigation, consistency, interaction, and engagement.	Session length exceeding 3 minutes on average.
Trustworthiness	Observation tests through usage scenarios.	Observations and specialised focus groups will assess whether the components align with the factors by which end-users frequently evaluate the trustworthiness of tools, including transparency, explainability, accuracy, data privacy, accountability, and accessibility.	Coverage of the majority of factors identified as important for assessing trustworthiness by end-users[18].
Reliability			
Faultlessness	Longevity testing.	Running the system over extended periods of time to observe performance deterioration, faults, failures, and other issues. Calculating metrics appropriate for measuring reliability. Verifying the correct functioning of the platform with use cases.	Retain scenario accuracy and task suitability after 2 weeks of uptime.
Availability	Longevity testing.	Observations will determine whether the components used in	Retain throughput rates after 2 weeks of uptime.

[17] A score of 68 is considered a good threshold, classifying the platform as "Good," as demonstrated here. For instance, GitLab's initial company target for the System Usability Score (SUS) was 73.

[18] D4.2 of AI4TRUST WP4 outlines the factors for assessing trustworthiness and provides a definition of the trustworthiness of AI systems.

		a scenario are operating consistently.	
Fault tolerance	Fault simulation.	Simulating faults in various components of the platform to observe its availability and performance under fault conditions.	Ensure over 95% platform availability and operational status.
Maintainability			
Modularity	Observation tests through usage scenarios.	Observations will determine whether the components used in a scenario can function independently, such that a change to one component has minimal to no impact on other components.	A change in a particular component should affect the availability of less than 25% of the other components.
Reusability	Observation tests through usage scenarios.	Observations will determine whether the components used in a scenario can be reused in other scenarios or platform applications.	Over 70% of components should be reusable.
Modifiability	Modification simulation.	Simulating modifications to various components of the platform to observe its availability and performance under modifications.	The platform should be available and operational for over 95% of the time. Accuracy and task suitability should be retained during modifications.
Flexibility			
Scalability	Observation tests through usage scenarios,	Observations will assess the platform's ability to operate efficiently under both gradual and sudden changes in workloads, evaluating its capacity to automatically scale resources.	Over 80% of automatic resource adjustments.

Table 2: AI4TRUST Platform Quality Model Tests

5.1. Evaluation Scenarios

This section outlines the **preliminary evaluation scenarios** that will be utilised for observation-based testing, while remaining adaptable to accommodate other types of tests, such as those focused on platform performance sub-characteristics. The three scenarios presented here will be consistent across all evaluation phases to ensure the reliability and consistency of the results. These evaluation phases correspond to the **three piloting, testing, and validation sessions of the AI4TRUST Platform** (i.e., three iterations, as stipulated in the amended Grant Agreement - GA). Furthermore, these evaluation scenarios are **directly linked to the pilot requirements** specified earlier (see Section 3), thereby bridging the required needs of the stakeholders with the platform's functionalities at a higher level.

The evaluation will be structured from two distinct perspectives:

- A **macroscopic perspective** of the AI4TRUST Platform, focusing on the platform as a whole.
- A **microscopic perspective** of the AI4TRUST Platform, focusing on the validation tools provided by the platform.

NCSR-D, the lead organisation for Task 6.3 — "Pilot Deployment and Operation" (T6.3), is **responsible for the organisation and coordination of the evaluation rounds** within the AI4TRUST project. This will be achieved through the concept of **interactive workshops**, which will accommodate multiple participants. The aim of this approach is to actively **engage participants** through plenary discussions and collaboration within the group, as well as to offer better support during their interactions with the tools. Therefore, the workshops will be designed to be interactive, based on the principles of adult learning, where participants share their thoughts and experiences on issues of concern, positioning them as the primary actors in the process.

In addition, one of the primary goals of the pilot workshops is the **co-design process**, where participants, as end-users of the AI4TRUST Platform, play a pivotal role in the decision-making process. As a result, discussions will often be driven by their inputs. The main objective of these workshops, however, will be to focus on the **performance evaluation of both the AI4TRUST Platform and the AI4TRUST tools.**

5.1.1. Analyse News Items Using the AI tools of AI4TRUST Platform

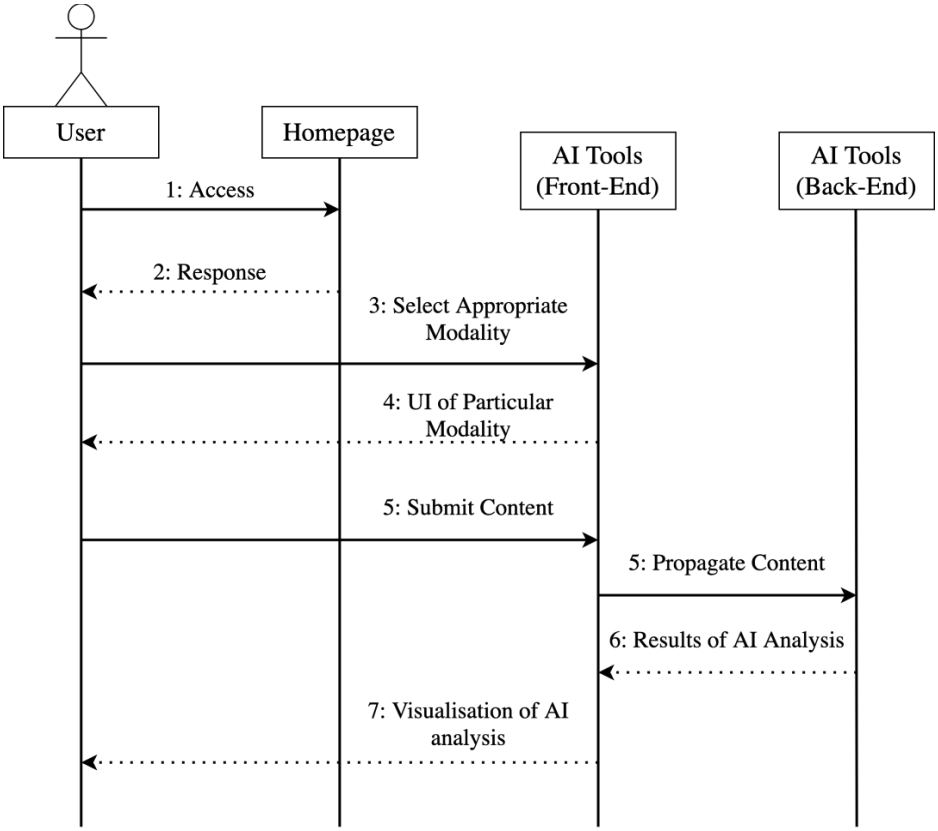

Figure 3: Evaluation Scenario1 - Analyse news items using the AI tools of AI4TRUST Platform

In the **first evaluation scenario**, which corresponds to the **first pilot phase**, it is assumed that the user initially accesses the **user interface** of the AI4TRUST Platform in order to **provide feedback on the platform as a whole**, including aspects such as **navigation, interaction, and usability**. To ensure consistency with the relevant literature and to establish a comparative set of questions, the **platform evaluation tool (questionnaire)** will incorporate questions from the **System Usability Scale (SUS)**[19] and the **Effectiveness, Efficiency, Satisfaction, and Learnability (EESL)**[20] frameworks.

Subsequently, the user must be able to identify the **AI4TRUST toolsuite**, a suite of tools designed to analyse and report on the veracity of multimodal contents. Based on the nature of the content, the user should be able to locate the tools relevant to the modality of the content (from a multi-modal toolkit) and apply the appropriate tools to analyse the content, review, and assess the results. This scenario is depicted in **Figure 3**.

[19] J. Brooke, "SUS: A Retrospective" Journal of Usability Studies, vol. 8, no. 2, pp. 29–40, 2013.
[20] J. Jeng, "Usability Assessment of Academic Digital Libraries: Effectiveness, Efficiency, Satisfaction, and Learnability," Libri, vol. 55, no. 2–3, 2005. [Online]. Available: https://doi.org/10.1515/LIBR.2005.96.

The aim of this scenario is to **gather feedback from participants** on key evaluation axes, including u**sability, functionality, user experience, and their general attitude towards using the AI tools** provided by the AI4TRUST Platform as part of their daily workflows. Throughout this interactive journey, we expect participants to reflect on the ease and friendliness with which they were able to find and use the AI tools within the platform **(user experience)**, report any unexpected errors or service denials **(functionality)**, as well as share their opinions on the tools' **usability** within their daily workflows and their attitude towards using them.

As part of the evaluation process, a **questionnaire** will be developed to collect evaluation data, incorporating **questions from both the SUS and EESL frameworks**. Additionally, an **open discussion session** will be held, where participants will be invited to share their experiences and provide further insights.

5.1.2. AI4TRUST Platform: Analytics and Dashboards

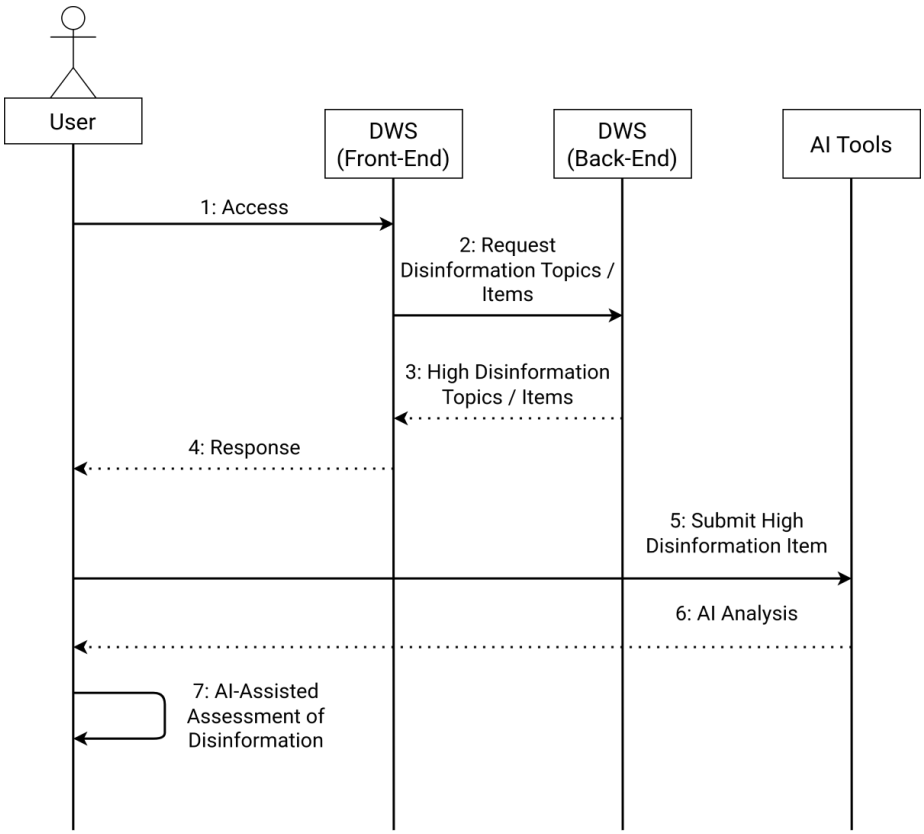

Figure 4: Evaluation Scenario 2 - Disinformation Warning System (DWS) Module

In the **second evaluation scenario**, which corresponds to the **second pilot phase**, the user is expected to explore the AI4TRUST Platform and **perform tasks related to mis/disinformation monitoring and debunking**. The evaluation tasks will involve: **(a)**

identifying the key elements of the AI4TRUST Platform modules and understanding the information they provide (for instance, identifying the Disinformation Warning System (DWS), interpreting its results, and navigating through its contents); and **(b) assessing the quality of each module** using the AI4TRUST quality model. Special attention will be given to the DWS. The DWS is expected to deliver results concerning disinformation items or topics based on automated analysis, which should subsequently be verified and further analysed by fact-checkers. This automated analysis relies on the examination of various data streams (such as social media, traditional media, and other relevant sources). The user must be able to select specific disinformation items or topics and conduct a deeper analysis using the available tools within the AI4TRUST Platform.

This scenario will evaluate all the modules included in the AI4TRUST Platform (e.g., dashboards with aggregated statistical data, fact-checker debunks, outputs from the DWS, etc.), as outlined in D5.4. Figure 4 illustrates the user's interaction with the DWS module as an example, with other modules following a similar interaction pattern. In a manner similar to the previous scenario, this evaluation aims to **gather feedback from participants** regarding key modules of the AI4TRUST Platform. Throughout this interactive process, we expect participants to reflect on the ease with which they can find and use specific modules within the Platform, assess whether the modules function as intended, and report any unexpected errors or service denials. An **evaluation questionnaire** will be used here as well, accompanied by an **open discu**ssion session to capture further insights from the end-users.

5.1.3. Curation of the AI4TRUST Toolkit

In **the third evaluation scenario**, which corresponds to the **third pilot phase**, the user receives analysis results from various tools of the AI4TRUST Platform, which are ranked according to the confidence levels assigned by the different AI models and classifiers. The user must be able to easily review and revise this information, providing feedback and corrections that will be further utilised by the AI4TRUST Platform to enhance the quality of the results produced by the integrated tools. This scenario focuses on **assessing the quality and usability of the provided interfaces**.

In contrast to the other scenarios, this particular evaluation scenario is primarily centred on **usability and user experience**. During this scenario, participants are required to critically evaluate and validate the information presented to them. We aim to assess how closely the automatic analysis provided by the AI tools aligns with the manual assessment performed by the users. Through this process, participants will be able to offer valuable

feedback, which will be collected via a **questionnaire**, to suggest improvements for the quality of the AI tools or other Platform modules. The key evaluation axes for this scenario are **user experience and functionality**.

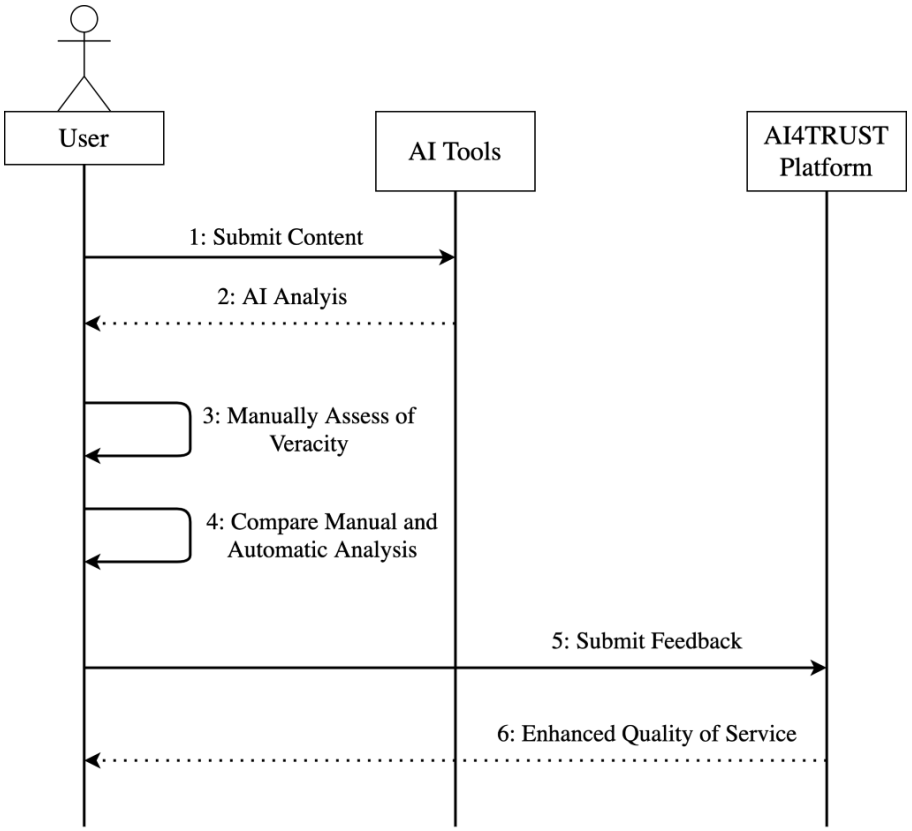

Figure 5: Evaluation Scenario 3 - Curation of the AI4TRUST Toolkit

5.1.4. Connection to Pilot Requirements

There is a **direct connection between the evaluation scenarios and the pilot requirements** (Section 3). These requirements encompass a range of tasks, such as the automated assessment of information source reliability, the establishment of benchmarks for determining source credibility, and the automation of editorial standards to ensure accuracy and consistency. Additionally, the Platform is expected to automate the rating of webpages containing potentially misleading claims, integrate datasets from fact-checking organisations for enhanced verification, and evaluate the authenticity of sources from social media platforms. Other key requirements include providing access to a comprehensive database of debunked sources, aggregating trending topics related to specific issues, and offering granular labelling and categorisation of information sources.

The integration of multiple data sources for comprehensive analysis is another critical aspect, along with quantifying infodemic risk levels to assess the severity of misinformation

and disinformation. Moreover, the Platform needs to support efficient data processing and analysis, incorporate social media monitoring to track real-time trends, and integrate signals from community input to enhance detection and response. End-users also expect the platform to detect deepfakes and human-generated misinformation, along with assessing the reliability of these sources. The ability to access automated fact-checking for multimodal content, support multiple languages, and conduct automated reliability checks with alerts and streamlined processes are also vital. Furthermore, users will require the Platform to generate textual snippets containing evidence and counterarguments to support claims of disinformation, enrich fact-checker reports with AI-generated evidence, and monitor content for potentially misleading material.

Scenario 1 addresses the reliability analysis of information sources, as it facilitates the AI-assisted analysis of news items through the use of the AI tools available on the AI4TRUST Platform. Additionally, certain tools may support fact-checkers by providing counterarguments for their reports. **Scenario 2** aligns with the infodemic trends requirement, enabling interaction with the AI4TRUST Platform analytics, particularly the Disinformation Warning System (DWS) module. The integration of fact-checking activities into the daily journalistic process is covered across all three scenarios, as they collectively enable the seamless and efficient AI-assisted assessment of news items and/or disinformation topics. **Scenario 3** is closely aligned with the requirements of the AI4TRUST end-users, who have specific needs related to the evaluation and verification of information. This scenario will focus on assessing the user experience and functionality of the AI4TRUST Platform, evaluating how well it meets these requirements. It will examine how users interact with the Platform, whether it supports their needs effectively, and how efficiently it enables them to carry out the tasks related to disinformation identification, analysis, and debunking. Ultimately, this evaluation will help ensure that the Platform is not only functional but also intuitive and reliable, supporting end-users in their daily work of verifying and countering mis/disinformation.

Evaluation Scenarios	Pilot Requirements
Scenario 1	• Analysis of the reliability of information sources. • Evidence-based textual inoculation strategies. • Integration of fact-checking activities into the daily journalistic workflow.
Scenario 2	• Analysis of infodemic trends related to specific public interest issues.

	• Integration of fact-checking activities within the daily journalistic workflow.
Scenario 3	• Integration of fact-checking activities within the daily journalistic workflow.

Table 3: Connecting Use Scenarios to Pilot Requirements

In conclusion, **the evaluation scenarios outlined above comprehensively address all the pilot requirements** gathered from the AI4TRUST end-users during the pilot preparatory phase (see Section 3 above on collection of pilot requirements, and Section 6 below).

6. Other Activities Relevant to Evaluation

In addition to the three phases of testing, validation, and evaluation of the AI4TRUST Platform, we have outlined several relevant activities for the piloting sessions. These activities are structured around three distinct phases: **preparation, pilot execution, and post-pilot phases**.

6.1. Preparatory Activities

This section outlines the preparatory activities to be carried out **prior to the execution of the piloting sessions** (see Sub-section 6.2), which will ensure a seamless and efficient execution phase, involving both end-users who are part of the AI4TRUST consortium (i.e., fact-checkers and journalists) and additional key stakeholders where applicable (e.g., within the third piloting phase).

Schedule and Duration: The schedule and duration for each testing session will be determined, taking into account the availability of participants and allocating sufficient time for each activity. This includes providing breaks and accounting for potential technical issues. The schedule will also ensure that it accommodates participants from different time zones, if necessary.

Participant Invitation: Relevant stakeholders will be invited to participate in the testing sessions. They will be informed of their involvement and responsibilities to ensure the smooth execution of the piloting session. Coordination of internal and external planning is

crucial to align end-users and key stakeholder timetables, workloads, and schedules for their active participation and effective feedback collection.

Setup of Communication Channels: A designated point of contact will be established within each organisation involved in the piloting session to facilitate direct communication with end-users and key stakeholders throughout the session. Regular meetings will be scheduled with all points of contact in the lead-up to and during the execution phase to monitor progress and ensure clear communication. Organisation emails will also serve as an additional communication tool for sharing written and digital information and materials.

Preparation of Briefing Content: Detailed briefing documents and instructional materials will be prepared to help participants prepare for the validation of the platform. These materials will act as an offline resource to familiarise stakeholders with the Platform, the evaluation process, and their respective objectives. The content will clearly communicate the purpose of the testing sessions, the specific tasks and use scenarios to be executed, and the expected outcomes.

Facilitation and Moderation: A qualified facilitator or moderator will be assigned to guide participants through the testing sessions. This facilitator should be experienced in online testing methodologies, ensuring a smooth and productive experience for all participants. Each involved organisation should designate a facilitator/moderator/coordinator and select an internal team of qualified users based on their expertise in relevant aspects of the platform, such as disinformation signals, language proficiency, or expertise in a specific topic of interest.

Setup of Testing Environment: The technical setup will be ensured to provide participants with seamless access to the platform and enable them to execute the use scenarios. Functionality tests will be conducted in advance to verify that all end-users and key stakeholders involved in the testing sessions can access the AI4TRUST Platform without any issues or interruptions.

6.1.1. Stakeholder Training / Participatory Workshops

To prepare for the testing and validation activities, we will conduct **participatory workshops and training sessions** involving the end-users who are part of the AI4TRUST consortium (i.e., fact-checkers and journalists) and additional key stakeholders where applicable (e.g., policy-makers within the third phase piloting), as previously mentioned, to familiarise them with the AI4TRUST Platform, its functionalities, and potential improvements to address any existing gaps. These workshops will also provide

introductory materials to guide stakeholders throughout the subsequent piloting process. By **engaging experts from diverse backgrounds**, we aim to foster collaboration, mutual understanding, and alignment of goals. The preparation for the workshops will include the following steps:

Workshop Objective Definition: Clearly defining the objective of the workshops is critical. The primary aim is to familiarise all end-users and key stakeholders with the AI4TRUST Platform, its functionalities, and its potential impact. Additionally, the workshops will present the objectives of the testing sessions and the success metrics for the Platform during the piloting phase.

Briefing Materials: Developed briefing materials will be presented to introduce the end-users and key stakeholders to the context of the Platform and its functionalities. These materials will also outline the appropriate use scenarios and objectives of the testing sessions. The briefing content will act as an offline resource, complementing the online training on the platform.

Use Scenarios: The workshop will include a presentation of the use scenarios designed to inform participants about the testing activities. At the same time, these scenarios will serve as a discussion point for exploring how the platform can be integrated into their daily workflows and activities.

Training on the AI4TRUST Platform: A comprehensive training session will be organised to help end-users and key stakeholders familiarise themselves with the AI4TRUST Platform and its use for the relevant use scenarios. This training will cover the Platform's technical aspects and will provide oral instructions to serve as a tutorial for understanding the functionalities involved in the use scenarios.

Hands-on Practice: Participants will be given hands-on practice opportunities to directly engage with the AI4TRUST Platform. This will allow end-users and key stakeholders to explore the Platform's various features and functionalities, ensuring they gain practical experience.

Feedback: Open discussions and brainstorming sessions will be organised to collect feedback, suggestions, and concerns from end-users and key stakeholders. This will help refine the Platform and its functionalities based on the experiences and needs of the participants.

Ethical Considerations: A dedicated session will focus on the ethical considerations surrounding the use of AI in the end-users and stakeholders' activities. This will ensure the responsible and unbiased implementation of AI, addressing concerns such as data privacy, fairness, and transparency.

6.2. Pilot Execution Activities

This section outlines the **activities planned for the pilot execution phase** that will pave the way for post-pilot and community-building efforts (see Sub-section 6.3). These activities are critical to ensuring the success of the pilot and facilitating ongoing improvements and stakeholder engagement after the testing phase.

Documentation and Observation: A robust system for documenting and observing the testing sessions will be established. This may involve capturing video recordings, taking detailed notes, and utilising screen-sharing and remote observation tools. It is essential to ensure that all necessary data is collected in a thorough, accurate, privacy-preserving, and secure manner, providing a comprehensive record of the testing sessions for further analysis and future reference.

Data Analysis and Interpretation: A clear plan for analysing and interpreting the collected data will be developed. This plan will identify key metrics, emerging themes, and patterns that need to be considered during the analysis. By focusing on these critical aspects, the pilot leaders will be able to derive valuable insights and actionable recommendations from the piloting sessions that can guide future iterations of the Platform.

Reporting and Action Steps: A comprehensive report will be prepared summarising the findings, insights, and recommendations derived from the piloting sessions. This report will clearly outline the action steps that need to be taken based on the results. These steps may include adjustments or improvements to the tested product or service, ensuring that the platform evolves to meet the needs and expectations of its users.

Feedback Collection Mechanisms: Effective feedback collection mechanisms will be implemented to gather valuable insights from participants. Encouraging participants to provide detailed and constructive feedback will play a critical role in the continuous improvement of the Platform. This feedback will not only inform the testing process but will also contribute to refining the platform’s functionalities, ensuring it remains aligned with user needs and expectations.

These activities, when carefully executed, will ensure that **the pilot phase is documented thoroughly, analysed effectively, and followed by actionable improvements**, all while fostering community engagement for future development.

6.3. Post-Pilot Activities / Community Building

This section outlines the activities involved in the **final phase of the piloting process**, focusing on acknowledging contributions, maintaining ongoing communication, ensuring privacy, and fostering continuous engagement with the community.

Acknowledgment and Recognition: Recognising and appreciating the contributions of community members who actively participate in the online testing is essential. This can be achieved through various forms of acknowledgment such as public recognition, certificates of participation, or offering incentives. This recognition serves to express gratitude for their time, effort, and valuable input, which helps build positive relationships and encourages further involvement.

Regular Updates and Progress Reports: It is vital to keep the community informed about the progress of the pilot and the AI4TRUST services. Providing regular updates on key milestones, the insights gained from community feedback, and explaining how their input has influenced the development of the project ensures transparency and trust. This practice helps maintain an engaged community by keeping participants in the loop and showing the tangible impact of their contributions.

Privacy and Data Protection: Ensuring the privacy and data protection of community members participating in online testing is a priority. It is essential to communicate clearly how participants' personal information will be handled, stored, and used, adhering to relevant privacy regulations. Transparency in data protection practices reassures community members that their privacy is respected and safeguarded throughout the piloting process.

Continuous Engagement: To sustain engagement beyond the piloting sessions, it is important to maintain a community where participants can stay connected, share their experiences, and continue providing valuable input for future developments. This continuous engagement helps to create a loyal and active community, allowing for ongoing collaboration and feedback as the Platform evolves. Maintaining this continuous engagement is crucial not only for fostering a loyal and active community but also for ensuring the long-term sustainability of the AI4TRUST project. By keeping the lines of

communication open, the Platform and its tools can be continuously improved based on real-world feedback, ensuring their relevance, effectiveness, and broad adoption as the AI4TRUST Platform evolves and is increasingly exploited across diverse domains.

These activities ensure that the final phase of the piloting process not only acknowledges and appreciates participants but also maintains a **strong connection with the community**, fosters **transparency**, and supports **long-term engagement**.

7. Piloting Plan - Timeline

This section delineates a **comprehensive Piloting Plan**, providing a detailed **roadmap** of the pilot activities and the specific **timelines** associated with each distinct phase of the pilot. The primary objective is to offer a clear and structured overview of the **piloting process**, ensuring that end-users and key stakeholders are adequately briefed on the progression of the piloting plan, as well as the achievement of the relevant project milestones. In addition, this section discusses an **effective end-user and stakeholder engagement strategy**, which is tailored to two distinct phases of participant involvement. To optimise the success of the pilot, an effective stakeholder engagement strategy is paramount. Actively involving a sufficient number of representative end-users and key stakeholders ensures that the needs and perspectives of the interested parties are accurately reflected, thereby increasing the likelihood that the directions adopted for Platform and tools development within AI4TRUST are both **realistic and sustainable**.

7.1. Stakeholder Engagement – Testing Participation

This section outlines the **end-users and stakeholder engagement efforts** and highlights the participation of organisations across the first, second, and third piloting phases. It provides a comprehensive overview of the involved entities and their backgrounds, as well as the human resources they have contributed to the piloting sessions.

Engaging end-users and stakeholders is a continuous process that evolves throughout the lifecycle of the piloting phases and, consequently, the project itself. Ongoing efforts to engage end-users and stakeholders enrich the various stages of the piloting sessions. Therefore, following the completion of the first phase of the evaluation, it is crucial to maintain communication with the community. This ensures that they remain informed about the progress of AI4TRUST (including functional updates) and are provided with

continued access to interact with the Platform and its tools. This approach keeps pilot participants involved and motivated in anticipation of subsequent evaluation phases, fostering a sense that they are actively participating in a **co-creation process** and that their feedback holds significant value.

Below are some suggested activities that could contribute to this goal and help sustain the community's interest:

- **Publishing relevant announcements** on the social media platforms of the partners and the project, expressing gratitude to the participants.
- **Scheduling online meetings** among involved partners, during which the main findings of the evaluation will be presented.
- **Sending updates to the community** as the **AI4TRUST Platform** and its tools undergo functional improvements, encouraging members to continue using the tools and provide feedback and suggestions for further enhancements.
- **Sending out invitations** to participate in subsequent piloting rounds.
- **Ensuring participants** that their personal data will be managed with respect and confidentiality, in compliance with applicable legislation, and using the appropriate security measures.

7.1.1. Phase 1

Media professionals (i.e., fact-checkers and journalists) that are part of the project's consortium will serve as the first stakeholder groups involved in the piloting activities of the AI4TRUST Platform. They will share their insights and provide staff members to offer feedback regarding the Platform's features, tools, and modules, with the goal of refining the platform. The organisations involved during this phase of the piloting sessions are: **MALDITA**, **DEMAGOG**, **SKY TG24**, **ELLINIKA**, **EURACTIV**, **ADB**, and **EMS**. Below, the estimated number of participants involved in the pilot workshops per partner is provided based on their current capacity in the field; these numbers have been confirmed by each respective partner. However, partners will also seek to invite external participants to increase the number of attendees. Deliverable **D6.2** of WP6 provides a more precise indication of the number of participants and the professional titles of each individual.

MALDITA

Maldita.es is a non-profit organisation based in Spain dedicated to combating disinformation through a multifaceted approach, encompassing journalism, education, technological innovation, research, and policy advocacy. Fact-checking journalism is one of its core activities. In the context of the AI4TRUST project, Maldita.es plays a pivotal role as both a key end-user and a valuable contributor to the AI4TRUST Platform. As end-users, they represent an essential target audience, including fact-checkers, journalists, and other media professionals who will benefit from the platform's capabilities. Additionally, Maldita.es will actively contribute to the project by facilitating access to datasets gathered through its crowdsourced Disinformation Management System. By collaborating with AI4TRUST, Maldita.es aims to enhance its technological capacities for monitoring and countering disinformation, thereby improving the quality of its journalistic processes and reinforcing its position in the fight against disinformation. Maldita will involve in-house stakeholders, such as fact-checking journalists, editors, and media literacy practitioners, to align the platform with the real needs and requirements of users active in the fact-checking domain. Specifically, they will involve:

- At least 7 fact-checking journalists, primarily from Maldito Bulo (the hoax and fact-checking unit);
- 1 editor who oversees newsroom coordination;
- At least 3 media literacy practitioners;
- At least 1 computer engineer responsible for technological development and implementation within the organization;
- At least 2 disinformation and public policy experts.

DEMAGOG

The Demagog Association is a non-profit organization from Poland focused on combating disinformation through fact-checking, debunking false information, media literacy and education, technology development, research, and policy action. The organisation joined the AI4TRUST project to improve the quality of public debate, reduce misinformation, and create new tools to automate the process, while remaining at the forefront of the battle against disinformation. Demagog will involve the following stakeholders to align their needs and requirements with the features of the AI4TRUST Platform:

- 5 fact-checkers (languages: Polish, English);
- 1 project coordinator.

SKYTG24

SkyTG24 is part of Sky Italia and operates under the Sky Group, Europe's leading entertainment provider, with 23 million subscribers. It provides daily coverage in Italy, delivering over 7,000 hours of live content each year. Sky TG24 excels in its ability to deliver breaking news promptly and effectively, tackling significant issues through in-depth reports, investigative journalism, and interviews. Sky TG24 will involve the following participants:

- 1 pilot project coordinator;
- 1 video producer;
- 1 social media manager;
- 1 media literacy practitioner;
- 4 journalists (freelance, TV, and digital newsroom);
- 1 digital product manager.

ELLINIKA

Ellinika Hoaxes (EH) is a Greek non-profit fact-checking organisation, the first in Greece to coordinate related initiatives and become certified by the IFCN. EH focuses solely on fact-checking misinformation and disinformation. It has provided a full dataset of its fact-checked articles and related metadata to AI4TRUST and aims to become an end-user of the platform. Ellinika Hoaxes will involve the following participants:

- At least 5 fact-checkers;
- 1 team editor;
- 1 project manager.

EURACTIV

EURACTIV is an independent pan-European media network, founded in Brussels in 1999. It has become a well-respected source of unbiased information on EU affairs and sparks policy debates among stakeholders. EURACTIV's fact-checking activities have been enhanced through the TRUE INFO project, where significant time has been devoted to creating verifiable content related to the Russian invasion of Ukraine. The organisation has used various tools to analyse content flows and detect false claims, supporting its journalists' work. EURACTIV will involve the following participants:

- 3 in-house freelance journalists;
- 1 fact-checking project coordinator in Brussels;
- Policy makers reached through the project's Final Forum.

ADB

Association Digital Bridge (ADB), operating under EURACTIV in Romania, is a media NGO working to improve quality journalism in a country heavily influenced by political parties and widespread disinformation. ADB specialises in high-quality journalism, particularly focusing on public policies. ADB is highly interested in upgrading its "Facts, not Fake" initiative by utilising AI tools for fact-checking, both pre- and post-publication. ADB will involve the following participants:

- 2 journalists in Romania;
- 1 media expert (academic, specialised in media policies);
- 1 technical specialist;
- 1 AI fact-checking tools coordinator.

EMS

Europejskie Media SP ZOO (EMS), operating under EURACTIV in Poland, focuses on providing reliable and unbiased information to its readers. EMS has been extensively involved in projects dedicated to combating disinformation. As part of the Media Against Disinformation project, EMS focused on detecting and combating disinformation targeting the EU, its Member States, and the general public. EMS will involve the following participants:

- At least 4 journalists;
- 1 fact-checking specialist;
- At least 1 podcast producer.

7.1.2. Phase 2

During the **second phase** of end-user and stakeholder engagement, a campaign will be developed to target additional participants for future piloting sessions. In addition to media professionals, the campaign will also aim to involve other key stakeholder groups, such as **policymakers** and **researchers**. Furthermore, the consortium members will contribute by identifying a list of key stakeholders who are capable of providing additional users, primarily for piloting the third version of the Platform (although additional stakeholders may also be invited during the first and second iterations by AI4TRUST end-users). This process will involve careful consideration of the **end-users and stakeholders' relevance** and the corresponding impact on the project's objectives.

The approach to engaging stakeholders will involve **direct invitations**, which will be incrementally promoted through communications from members of the consortium. This

strategy ensures **clear communication**, while also fostering a sense of **collaboration** and **partnership** between the project and its stakeholders. Involving stakeholders from diverse backgrounds and perspectives is highly valued, as it facilitates the collection of **varied feedback** and ensures that the further development of the platform will effectively meet the needs of its users.

Finally, the inclusion of **AI4TRUST** among the sister projects of **Cluster 4 of Horizon Europe**, within the framework of the **AI against Disinformation Cluster**, provides a unique opportunity to expand networking and engagement with a broader and more diverse range of **stakeholders**. This not only enhances participation in the **piloting** phases and ensures the **sustainability** of the **AI4TRUST Platform** but also strengthens the **dissemination** and **communication** efforts of the project, ensuring a wider and more lasting impact.

7.2. Roadmap of Pilot Activities

This section outlines the **roadmap** of the pilot activities, while simultaneously presenting a **timeline** for each distinct stage of the pilot. It serves as a **comprehensive overview** of the **piloting process**, ensuring a structured and clear understanding of the sequence of events and milestones to be achieved. A specific **focus** is placed on **Piloting Phase 1** (see Sub-section 7.2.1), which concludes with the deliverable **D6.2** of **WP6** (Figure 6). The final paragraph of this section (see Sub-section 7.2.2) will delineate the second and third phases of piloting, providing a clear outlook on the subsequent steps of the process.

7.2.1. Piloting Phase 1

As the previous sections comprehensively outlines these actions (see Sections 5 and 6), only a portion of them is represented in **Figure 6**. **Green blocks** denote **preparatory activities**, **yellow blocks** represent **execution phase activities**, while **red blocks** signify **post-pilot phase activities**. The **timeline** is as follows: commencing with the initiation of the **preparation phase**, the first course of action is to assign respective **pilot points of contact** to establish direct channels of communication. Following this, the **preparation of briefing material** will take place, with its finalisation completed just before the delivery of the first version of the **AI4TRUST Platform**. After the release, the briefing material will be adjusted to include further details on specific components of the Platform. Subsequently, **preparatory workshops** will be initiated. The **first pilot session** will begin after the completion of the preparatory workshops, marking the commencement of the **pilot execution phase**. This phase aims to achieve the relevant **project milestones**. The pilot workshops will predominantly be conducted in **physical form**, with participants gathered

at the host organisation. Where not possible, they are made in a hybrid format. **Moderators from NCSR-D** will facilitate the interactive process and guide the participants seamlessly through the evaluation scenarios. The workshops will span a few hours and collect both **qualitative** and **quantitative feedback** via **questionnaires** and **moderator reports**. The **pilot execution phase** will conclude, transitioning into the **post-pilot phase** with the preparation for the **1st pilot evaluation report (D6.2)**.

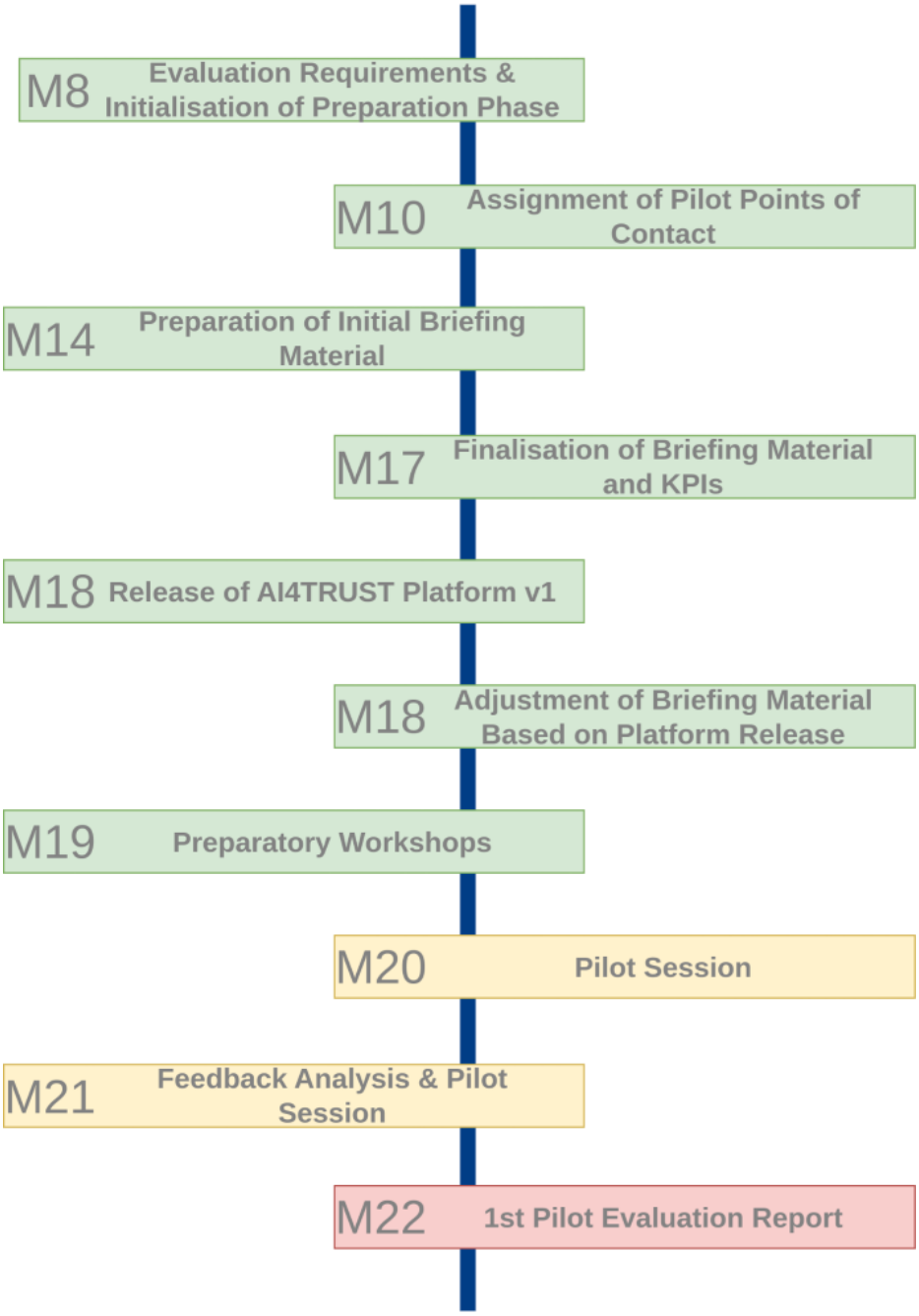

Figure 6: Timeline of Piloting Phase 1

7.2.2. Subsequent Piloting Phases

As we progress into subsequent piloting phases, **the approach will be refined** based on insights and outcomes from Phase 1. The structure established in **the initial phase will remain intact**, ensuring continuity and consistency, including the categorisation of activities into three distinct groups, each designed to address specific phases of the pilot session. Furthermore, **the recruitment and engagement strategy will be reinforced** to involve a broader spectrum of users, ensuring robustness and covering a wide range of evaluation scenarios. The timeline and users involved will be adjusted in future iterations. A **standardised evaluation plan** that can be applied to all subsequent piloting phases is described in D6.2.

With specific regard to the **second and third piloting phases**, as outlined in **Section 5**, these will build upon the foundational structure and insights derived from the Piloting Phase 1 (see Sub-section 7.2.1). In the second evaluation scenario, which corresponds to the **second pilot phase**, the user is expected to explore the **AI4TRUST Platform** and perform tasks related to **mis/disinformation monitoring** and **debunking**. The evaluation tasks will involve: (a) identifying the key elements of the **AI4TRUST Platform modules** and understanding the information they provide (for instance, identifying the **Disinformation Warning System (DWS)**, interpreting its results, and navigating through its contents); and (b) assessing the quality of each module using the **AI4TRUST Quality Model**. Therefore, this scenario will evaluate all the modules included in the **AI4TRUST Platform** (e.g., **dashboards** with aggregated statistical data, **fact-checker debunks**, outputs from the **DWS**, etc.), as outlined in **D5.4 and D5.6 of WP5**. This evaluation aims to gather feedback from participants regarding the key modules of the **AI4TRUST Platform**. Throughout this interactive process, we expect participants to reflect on the ease with which they can find and use specific modules within the Platform, assess whether the modules function as intended, and report any unexpected errors or service denials. An evaluation **questionnaire** will be used here as well, accompanied by an **open discussion session** to capture further insights from the end-users.

In the third evaluation scenario, which corresponds to the **third pilot phase**, the user receives analysis results from various tools of the **AI4TRUST Platform (v.2 - D5.6 of WP5)**, which are ranked according to the confidence levels assigned by the different **AI models** and **classifiers**. The user must be able to easily review and revise this information, providing feedback and corrections that will be further utilised by the **AI4TRUST Platform** to enhance the quality of the results produced by the integrated tools. This scenario focuses on assessing the **quality** and **usability** of the provided interfaces.

In contrast to the other scenarios, this particular evaluation scenario is primarily centred on **usability** and **user experience**. During this scenario, participants are required to critically evaluate and validate the information presented to them. We aim to assess how closely the automatic analysis provided by the **AI tools** aligns with the manual assessment performed by the users. Through this process, participants will be able to offer valuable feedback, which will be collected via a **questionnaire**, to suggest improvements for the quality of the **AI tools** or other **Platform modules**.

Roadmap for the Second and Third Piloting Phases of the AI4TRUST Project

The roadmap for the **second** and **third piloting phases** of the AI4TRUST project outlines the key activities and milestones in accordance with the project timeline, incorporating feedback from the first phase and preparing the Platform for further development and refinement.

Second Piloting Phase:

1. **Release of AI4TRUST Platform v2 (D5.6) – 31st March 2025**
    - The **second version of the AI4TRUST Platform** will be released, incorporating improvements and updates based on the first phase of piloting. This version will include several enhancements to the Platform and its tools (including the DWS), reflecting user feedback and identified needs.
2. **Second Piloting Session – April 2025**
    - The **second piloting session** will be initiated, involving **end-users** from partner organisations and invited relevant stakeholders. These users will interact with the updated version of the Platform, engaging in specific tasks designed to assess the new functionalities and performance of the Platform.
    - During this session, special attention will be given to testing the new and updated features, such as the **updated DWS** and other improved tools in the Platform, as well as gathering detailed qualitative and quantitative feedback from participants.
3. **Internal Milestone Report – May 2025**
    - A report summarising the outcomes of the **second piloting session** will be prepared internally, documenting the insights gathered, user feedback, and

any issues encountered. This report will help guide the next steps in the development process and prepare for the integration of further feedback into the Platform.

4. **Feedback Integration and Platform Updates – May / September 2025**
    - Based on the feedback from the second piloting session, the Platform will be updated. The development team (WP3, WP4) and the system integrator (WP5) will incorporate improvements and address any issues identified during the testing. This will include the integration of new functionalities, such as the **Social Network Analysis (SNA) tools** (WP4) and additional updates to the **DWS** (WP3).
5. **Preparation for Third Piloting Phase – August 2025**
    - The pilot leader will prepare for the **third piloting phase**. This phase will test the newly integrated tools and assess their effectiveness in real-world conditions.

Third Piloting Phase:

1. **Start of Third Piloting Phase – Late Summer / Early Autumn 2025**
    - The **third piloting phase** will begin with the deployment of the updated AI4TRUST Platform, now equipped with the new SNA tools (WP4), the revised DWS (WP3), and other updates based on previous feedback.
    - The third piloting phase will focus on a more refined set of evaluation scenarios, incorporating real-time feedback, critical validation of the new tools, and their integration within the Platform.
2. **Piloting Sessions – Late Summer / Autumn 2025**
    - The piloting sessions will involve both the AI4TRUST end-users and a diverse set of stakeholders, further enhancing the breadth and depth of the feedback gathered. These sessions will focus on validating the **user experience** and the performance of the **SNA tools (WP4)** and **updated WP3 tools** under more complex scenarios.
    - The third phase will involve a combination of **interactive workshops**, **feedback gathering**, and **evaluation tasks** to assess the improvements in platform usability, functionality, and accuracy.

3. **Submission of Piloting Sessions Report (D6.3) – 31st October 2025**
    - The results from the **second** and **third piloting sessions** will be compiled into the **Piloting Sessions Report v2** (D6.3), which will include detailed insights from the two iterations, highlighting the progress made, challenges encountered, and future recommendations.
    - This report will serve as the basis for the **final updates** to the Platform and its tools, leading into the final stage of the project.
4. **Post-Piloting Platform Integration and Updates – November 2025 / February 2026**
    - Following the completion of the third piloting phase, the feedback from both sessions will be analysed and used to update the platform and its tools. This will involve further integration and refinement of the AI models, tool functionalities, and the Platform's overall usability. These improvements will feed into the final updates for the Platform.

Final Version of the AI4TRUST Platform Release (D5.7) – 28th February 2026:

- The final version of the **AI4TRUST Platform** will be released, incorporating the lessons learned from the piloting phases, and will serve as the culmination of the project's objectives. The final version will be fully updated with all the integrated features and tools, marking the conclusion of the AI4TRUST project.

This roadmap ensures that the **AI4TRUST Platform** is continuously improved and refined throughout the piloting phases, allowing for the integration of end-users and stakeholders' feedback, the addition of new features, and the achievement of key project milestones, all contributing to the ultimate success of the project.

8. Conclusions

This deliverable presents a **state of the art** in fact-checking, from which the pilot specifications are derived. The **pilot requirements** are extracted and enumerated, drawing directly from the established fact-checking frameworks. An **evaluation framework** known as the **AI4TRUST Platform Quality Model**, based on the widely recognised **ISO/IEC 25010**, is introduced. **Evaluation scenarios** for each of the three piloting phases are developed and presented, each specifically aligned with the respective pilot requirement.

Furthermore, additional activities related to the evaluation are classified into three distinct groups, providing a clear overview of the **evaluation process**. The **piloting plan** and **timeline** are thoroughly detailed, including an enumeration of activities related to **participation recruitment** and **stakeholder engagement**.

This deliverable contextualises the piloting sessions by establishing the **main objective**, the **means** and **resources** required to achieve it. Additionally, it presents the **methodology** to be followed in order to gather meaningful feedback, properly analyse it, and generate **actionable and concrete insights that will inform the further development of the AI4TRUST Platform**.

As the AI4TRUST Platform develops, more detailed assessment criteria will be identified, which could lead to important improvements in how the pilot is conducted, how feedback is collected, and in the platform's modules and tools. Consequently, **the relevant methodologies will likely be re-assessed and calibrated** for subsequent piloting cycles, as outlined in the work plan, while always adhering to the **core principles** and **guidelines** presented herein.