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

AI4TRUST

AI4TRUST

PILOT PLANNING REPORT

D6.1: Pilot planning report

PARTNERS

Document information

Title	D6.1 - Pilot planning report	
Editor	NCSR - D	
Contributors	MALDITA, DEMAGOG, SKYTG24, ELLENIKA, EURACTIV, ADB, EMS, UCAM, UPB, 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 ☐ UPB☐ ☐ SAHER ☒ GDI	☐ SKYTG24 ☐ SKYTG24 ☐ ELLENIKA ☐ EURACTIV ☐ SKYTG24 ☐ ADB ☐ EMD ☐ XXX
Status	☐ Draft ☐ WP Manager accepted ☒ Coordinator accepted	

Requested date of submission	31 August 2023
Requested date for submission	15 September 2023

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

Table of contents

Executive summary

Piloting stands as a central endeavour within the AI4TRUST project. It strives to confront the challenges presented across various contexts, countries and misinformation or disinformation types. The development of a comprehensive piloting plan, tailored to the needs of different stakeholders but also allowing the consolidation of the pilot findings to a set of coherent requirements for the AI4TRUST platform, is thus integral to our project's success.

The present report describes the objectives pursued by the piloting design of the project, the principles under which the pilots are built, and the specific targets and plans for each foreseen pilot.

Firstly, a summary of the current state-of-the-art on specific problems and the challenges still faced is provided for all pilots; the ways the AI4TRUST platform is expected to help overcoming these challenges is analysed; and the resources allowing to properly assess the coverage of these requirements are presented.

Following the specification of the pilots' context, a description of the pilots' organisation in preparatory, execution and validation stages is described, with the activities falling under each stage outlined, in order to ensure that the basic design principles for the pilot plan are met.

Finally, the holistic evaluation methodology that will be applied across pilots - with customisations pertinent to each pilot's distinct characteristics - and will be used to collect, consolidate and analyse the pilot outputs concludes the report.

The presented piloting plan allows responsible partners to timely direct and organise their preparatory actions, while also providing a baseline for technical developments of the project and the preparation of appropriate documentation that will serve the plan prerequisites and assist pilot partners to communicate with the relevant stakeholders.

1. Introduction

The main objective of WP6 is to pilot and validate the AI4TRUST Platform and to plan and perform effective fact-checking and validation activities for media professionals, researchers and policy makers in order to increase their capacity to monitor, detect and record misinformation and disinformation on online social media and traditional media, as well as facilitate the creation and distribution of reliable information. Towards this goal, the specification of a concrete piloting plan accompanied by an appropriate feedback collection and evaluation methodology are essential for assessing the quality and effectiveness of the solutions provided by AI4TRUST.

Consequently, Task T6.1 is responsible for driving all work package activities by setting the context of the piloting sessions, the core objectives of each different session, the feedback collection and communication mechanisms, and ultimately the framework under which the pilots will contribute to the evolution and alignment of the AI4TRUST platform.

As the project targets different stakeholders with different needs and different pain points on their processes, the overall piloting plan entails seven different pilots, each driven and organised by the respective project partner. The present report summarises the activities planned for each pilot, along with their corresponding pilot deployment plan. In more detail, the report is organised as follows:

Section 2 analyses the design of each foreseen pilot, and places it in the relevant context. Namely, for each pilot we describe the current approach to the problem it aims to examine, the difficulties currently faced by the relevant stakeholder group, and the requirements posed by the latter to optimise their work. We proceed to describe how the participants on the pilot will be engaged, informed, and trained on the subject of each pilot based on the conceptualisation of the pilots. Section 3 proceeds to the presentation of the deployment plan for each pilot, based on their objectives and the guidelines for effective piloting to be covered by the deployment plan. Section 4 presents the evaluation methodology for the pilots, setting the main methodological steps to be followed, the feedback needed to be collected, and the relevant KPIs to be measured during and after the pilot execution.

2. Specification of the Pilots' experimentation

The section delves into the essential components that lay the foundation for a comprehensive and successful pilot design. The design builds on the current state-of-the-art, defining a first iteration of desirable requirements for the tools employed in each pilot, whose technical feasibility will be later analysed by the consortium partners. Furthermore, the section discusses an effective stakeholder engagement strategy, and outlines a meticulous event organisation methodology in accordance with the set requirements.

The state-of-the-art analysis serves as the cornerstone of the pilot design, as it provides a thorough understanding of existing technologies, methodologies, and best practices in the communities covered by AI4TRUST. It also serves as the main tool for the identification of potential gaps and opportunities, thus enabling the AI4TRUST consortium partners to build upon the existing knowledge and advance the relevant solutions. Gap and pains analysis directly leads to the requirements elicitation phase. This stage focuses on outlining specific needs and objectives for each pilot. The requirements act as a guiding framework throughout pilot execution, ensuring that the pilot and the observed criteria align with the desired outcomes.

To maximise the pilot's success, an effective stakeholder engagement strategy is vital. By actively involving a sufficient number of representative key stakeholders, each pilot ensures that its perspective reflects the needs of the interested parties and there is a higher likelihood that the directions adopted for tool development within AI4TRUST are realistic and sustainable. Finally, an event organisation and management plan are discussed for each pilot. The events will provide opportunities for the aforementioned stakeholders to collaborate, exchange knowledge and ideas, and gain a deeper understanding of the pilot's objectives and the objectives served by the AI4TRUST tools.

2.1 Maldita Pilot

Maldita.es is a non-profit organisation based in Spain that is dedicated to combating disinformation through a multifaceted approach encompassing journalism, education, technological innovation, research, and policy advocacy. Among its core activities, fact-checking journalism stands out as a primary mission. In the context of the AI4TRUST project, Maldita.es holds 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 AI4TRUST project by facilitating access to datasets gathered through its crowdsourced Disinformation Management System. Maldita.es has chosen to collaborate with AI4TRUST to enhance its technological capacities for monitoring and countering disinformation. By doing so, they aim to further elevate the quality of their journalistic processes while staying at the forefront of the battle against disinformation.

■ 2.1.1 Fact-checking state of the art analysis and best practice search

Maldita.es has established itself as an organisation with a remarkable track record in developing innovative technological solutions customised to meet the specific needs of fact-checkers and journalists. Their expertise extends to the creation of impactful tools such as a semi-automated WhatsApp chatbot and the Disinformation Management System (DMS) database tool. The semi-automated WhatsApp chatbot and the DMS database tool represent the cornerstone of Maldita.es' technological infrastructure. These tools are designed to collect and process instances of disinformation content reported by citizens circulating on the private messaging platform WhatsApp. Their significance lies in enabling real-time monitoring of disinformation on WhatsApp, a critical capability that aids in prioritising newsroom activities and expediting the identification of disinformation. It's worth noting that while these tools provide valuable technological support, the production of verified content remains a hands-on, manual process conducted by journalists. This process adheres to rigorous methodologies and high editorial standards, often involving collaboration with verified experts within the Maldita.es community. This comprehensive approach ensures the delivery of accurate and reliable information to the public.

■ 2.1.2 High level requirements of MALDITA Pilot

AI4TRUST service	Relevant process and current gaps	Desirable improvement
Reliability analysis of information sources	• Evaluation of source reliability is carried out manually by specialised journalists in accordance with rigorous editorial and methodological standards.	• Automated determination of the reliability of a source by means of its localization in the social networking structure of the information ecosystem.

	• Manual journalistic process of contrasting sources to analyse their reliability. • Typically relies on the inputs of verified experts from the Maldita.es community. • In Maldita.es' monitoring of information, the organisation relies on tips sent by citizens, but is unable to collect much information about these users so is unable to identify biases among these users as sources.	• Provision of an important benchmark as to whether a certain source is reliable to report from. • Automation process that adheres to the highest journalistic and editorial standards.
Infodemic trends for specific public interest issues	• Maldita.es currently counts on its own DMS to monitor disinformation campaigns, trends, and narratives circulating primarily on WhatsApp based on reports from citizens. • This system has a novel "narratives dashboard feature" that employs AI tools to automatically identify common narratives within content tagged as disinformation in the DMS. • The only quantitative measure for infodemic risk currently in use in the DMS is the Frequently Forwarded feature of WhatsApp and the number of times a content is sent to Maldita.es. • The DMS is state-of-the-art within the fact-checking community, and the tool is used by several fact-checking organisations globally. • However, the DMS tool does not comprehensively draw on	• Monitoring of infodemic trends by drawing on data from multiple digital information sources at the same time. • Quantification of the level of infodemic risk that is associated with a given topic in a given country in a reference time window, so that media can track what are the public interest issues which are likely to be affected by the waves of disinformation in any given moment (the infodemic 'temperature' of a given topic).

	diverse information sources beyond WhatsApp.	

Evidence-based textual inoculation	• The principal types of verified information that Maldita.es fact-checkers produce include:
	  ○ Fact-check articles that debunk hoaxes and disinformation.
	  ○ Explanatory articles and investigations that provide information about issues of public interest.
	  ○ Media literacy materials that employ pre-bunking techniques to equip citizens with critical thinking skills to counter mis-/disinformation.
	• These are developed manually by journalists and media literacy education experts. In the development of more extensive materials, semi-automated investigations extract disinformation narrative data in order to inform the development of content, e.g. media literacy materials to counter prominent climate disinformation narratives.
		• Provision of automated textual snippets containing evidence and counter-arguments supporting the claim that the specific content under inspection is a piece of disinformation.
		• Enrichment of fact-checkers' reports with additional evidence by using advanced NLP tools, to provide extensive textual material as a recommendation that can be used for information and inoculation activities by media journalists.

■ 2.1.3 Stakeholder engagement of media practitioners, fact-checkers, and policy makers

The stakeholders that are going to be involved in the MALDITA pilot will include in-house fact-checker journalists, editors, media literacy practitioners, and engineers, in order to define the AI4TRUST platform pilot according to the real needs and requirements of different kind of stakeholders within the organisation; specifically, Maldita's pilot 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 organisation.
- At least 2 disinformation and public policy experts.

2.1.4 Organisation of participatory workshops and training to plan the testing and validation activities

To prepare the testing and validation activities, we will conduct participatory workshops and training sessions involving the key stakeholders mentioned above to familiarise them with the AI4TRUST platform, its functionalities, and potential improvements to address current gaps. These workshops will also deliver introductory materials to guide stakeholders during the subsequent piloting process. The involvement of different experts from diverse backgrounds will foster collaboration, understanding, and alignment of goals. Preparation for the workshops will involve:

1. Workshop Objective Definition: Clearly define the objective of the workshops, which is to familiarise all stakeholders with the AI-platform, its functionalities, and its potential impact on fact-checking processes.
2. Guiding materials: Development of introductory materials and guide for stakeholders to deliver during the training workshops. The material will rely on cross-pilot guides developed in the context of the platform, and will augment them with content applicable to the specific objectives.
3. Training on AI4TRUST Platform: Begin the workshops with a comprehensive training session led by the developers of the platform. This training should cover:
    a. The technical aspects of the AI4TRUST-platform, its algorithms etc.
    b. The functionality of the source reliability analysis and its practical use application.
    c. The functionality of the infodemic trend monitoring and its practical use application.
    d. The functionality of the evidence-based textual inoculation feature and its practical use application.
4. Use Case Scenarios: Facilitate interactive sessions where fact-checking journalists and disinformation experts present real-world use case scenarios. Discuss how the AI4TRUST-platform can be integrated into their fact-checking processes and how it can help to identify unreliable sources and infodemic trends.

5. Hands-on Practice: Provide hands-on practice opportunities for all participants to use the AI4TRUST platform themselves. Allow fact-checking journalists and media literacy experts to explore the evidence-based textual inoculation feature and experience how it can enhance their fact-checking reports.
6. Feedback: Open discussions and brainstorming sessions to gather feedback, suggestions, and concerns from all stakeholders. Pay special attention to input from media literacy practitioners, as they can provide valuable insights on how to present the AI-generated information to the public effectively.
7. Testing Plan Formulation: Collaboratively develop a detailed plan for the pilot testing phase, outlining specific testing scenarios, data collection methods, and evaluation criteria. Allocate responsibilities among the stakeholders to ensure smooth execution during the testing phase.
8. Validation Criteria: Define clear validation criteria and success metrics for the platform during the pilot testing. Ensure that these criteria align with the overall objectives of the organisation and the expectations of the stakeholders.
9. Ethical Considerations: Dedicate a session to discuss ethical considerations related to the use of AI in fact-checking, ensuring responsible and unbiased AI implementation.

2.2. DEMAGOG Pilot

The Demagog Association is a non-profit organisation that fights disinformation through fact-checking, debunking of fake news, media literacy and education, technology development, research, and policy action. The organisation joined the project as it seeks to improve the quality of public debate, but also to help reduce misinformation and create new tools to automate the process.

■ 2.2.1 High level requirements of DEMAGOG Pilot

Fact-checking journalism and debunking fake news is a primary activity of the organisation. Like other fact-checking organisations, Demagog represents a key target end-user of the AI4TRUST platform: fact-checkers, journalists, and other media professionals. Demagog also provides its data resources to debunk fake news, which will be used to develop AI4TRUST.

AI4TRUST service	Relevant process and current gaps	Desirable improvement
Reliability analysis of information sources	• Evaluation of source reliability is carried out manually by fact-checkers in accordance with rigorous standards imposed by international fact-checking agencies: IFCN and EFSCN. • Process of contrasting sources and cross-check method for maximum reliability • Verifying the credibility of sources - access to verified and reliable sources and information makes the process easier and faster, something that is lacking in many newsrooms due to the speed of publication.	• Ensure appropriate, consistent and uniform standards for source and information verification. • Process automation based on methods used by independent fact-checking agencies
Infodemic trends for specific public interest issues	• Currently, there is no consistent system for monitoring the threat of infodemics either locally or globally. False information is found manually using social media monitoring tools in the media scan process. A system of alerts and mapping of the growth of false content by sector and medium would accelerate the process of content verification, which is crucial in the case of fact-checking. This action would also allow more data to be obtained on the characteristics of disinformation and its variation across media.	• Automate the system to recognize the growth of fraudulent and false content and report the risk in specific information areas and the risk of scams in social media.
Evidence-based textual inoculation	• The principal types of verified information that Demagog fact-checkers and analysts produce include: ○ Fact-check articles that verify political statements.	• Semi-automated content monitoring in the infosphere and delivery of text snippets and audio-visual material that could potentially mislead the viewer.

	○ Analyses debunking fake news and hoaxes. ○ Explanatory analyses of the current state of research ○ Articles and educational materials aimed at improving knowledge and digital competence in critical thinking, pre-bunking, media literacy, recognizing manipulated content and scams. ● As with the other fact-checking agencies, the process is manual. Implementing automation can help identify suspicious content more quickly and recognize the characteristics of mis- and disinformation, identifying narratives, plots, and the networks disseminating them.	● Recognizing the distinctive characteristics of fake news and finding patterns that may represent particular disinformation narratives. ● Locating networks of actors disseminating false content in a similar time and information sector.

■ 2.2.2 Stakeholder engagement of media practitioners, fact-checkers and policy makers

The pilot that will address the needs and requirements of stakeholders and the characteristics of the platform will include:

- 5 Fact-checkers (languages: Polish, English).
- 1 Project Coordinator.

■ 2.2.3 Fact-checking state of the art analysis and best practice search

Demagog Association has nine years of experience in information verification and is the first Polish fact-checking organisation. Demagog is a member of the International Fact-Checking Network and the European Fact-Checking Standards Network. In addition, it is a partner of Meta in the Third-Party Fact-Checking Program.

In its practice, the organisation verifies the statements of politicians, debunks fake news and conspiracies, prepares pre-bunking materials and writes analyses summarising the state of the art of technology and the process of its regulation. In addition to its educational

activities in media literacy and critical thinking, Demagog is also involved in the development of technological tools — building an educational platform and a climate chatbot (which automates the verification of climate change information).

The processes of media scanning, information verification, evaluation, and analysis writing are carried out manually by trained analysts, complying with rigorous methodological and editorial standards.

■ 2.2.4 Organisation of participatory workshops and training to plan the testing and validation activities

- Introduction: AI4TRUST Platform will be introduced to Demagog's fact-checkers and disinformation experts.
- Objectives: gathering user's feedback, testing usability and reliability of the tool's results in accordance with requirements.
- Testing Methods: usability testing in their environment with interactive sessions and focus on methods of using the tool in daily practice and language differences.
- Evaluation and feed-back: Questionnaires (written, interviews or combined form) made in conjunction with project partners to determine clear validation criteria and success metrics for the platform during the pilot testing; Open discussion session to validate the experience of media users and practitioners.

2.3. SKYTG24 Pilot

Sky TG24 is part of Sky Italia. It operates under the Sky Group, which is Europe's leading entertainment provider with 23 million subscribers. Sky Group is a division of Comcast NBCUniversal.

Sky TG24 provides daily coverage to the country, delivering over 7,000 hours of live content each year. Within this extensive coverage, 3,500 hours are dedicated to in-depth analysis, ensuring comprehensive reporting on the most important Italian and international news stories.

Sky TG24 adopts a modern and innovative narrative, employing diverse formats and languages, on air as on the website. This allows for robust interaction with the public, fostering active participation from users.

The network excels in its ability to deliver breaking news promptly and effectively and to tackle significant issues of our time, which have a tangible impact on people's lives, through in-depth news, investigative reports, reportage, and interviews.
In the past years, SkyTG24 carried out investigations[1,2,3,4,5] and provided insights into the role of bots in social media, revealed the existence of networks of Italian sites specialised in spreading fake news and highlighted the tendency of social media algorithms to suggest increasingly extreme and polarising content.
SkyTG24 pays great attention to the issues of technological innovation and artificial intelligence. The use of AI, the potential benefits and possible risks are an important topic of the journalistic activity.
Among the Italian brands included in the survey of Digital News Report 2023,[6] SkyTG24's trust scores it's one of the best (71%).[7]

■ 2.3.1 Fact-checking state of the art analysis and best practice search

Sky TG24 is also characterised by TV programs and daily digital data journalism and literacy activities (e.g. "Numeri"), with specific attention to economic, political and social issues. These activities aim to counter the spread of misinformation and disinformation.

Other examples of these journalistic activities are the TV/digital formats of scientific dissemination and fight against mis/disinformation, such as "Pillole di vaccino" (related to the COVID-19 vaccination campaigns) and "Impact" (a focus on climate change).

Fact-checking activities such as debunking is carried out only when assessed relevant for public opinion. In the past years, Sky TG24 has run specific fact-checking activities on news related to the COVID-19 pandemic (exposing some false information circulating on Italian social media), war in Ukraine and other current affairs news events.

Specialised agencies are sometimes involved in verifying/debunking information (e.g. Storyful and F5). In some cases, tools such as Google Search Imagine, Tineye, CrowdTangle, Whois are used by journalists.

■ 2.3.2 High level requirements of SKYTG24 Pilot

AI4TRUST Service	Relevant process and current gaps	Desirable improvement
Reliability analysis of information sources	Evaluation of source reliability is part of journalistic activity. Generally, it is carried out manually, in certain instances leveraging social media listening tools such as Crowdtangle (social/digital newsroom) or with the support of specialised agencies such as for example Storyful and, in Italy, F5.	• To learn and experiment new solutions that can improve journalism. • Automated/Assisted assessment of the authenticity of social media sources (for example social media profiles, groups, etc) • Access to a database of content already debunked by other news/fact-checking organisations.
Infodemic trends for specific public interest issues	Sky TG24 offers a beat-coverage of issues in the global agenda. We use tools like Crowdtangle, X (formerly known as Twitter), Google Trends to intercept trending topics when/before they get viral. Nonetheless, Sky TG24 doesn't always have the capacity to properly monitor emerging trends, as there is an abundance of content.	• A dashboard aggregating trending topics on specific public interest issues (for example, viral video, tweets, pictures, also from other countries on topics such as public health, war in Ukraine, climate crisis). The platform could be useful for informing daily coverage (both on-air and online), to identify infodemic trends in advance. • It is important to be very careful how sources are categorised, avoiding overly simplistic labels (true/fake source) that do not capture the complexity of the information ecosystem. For example: a news source may produce unreliable/ biased content on some topics and be very authoritative on other topics.

Integration of fact-checking activities in the journalistic process	Quite often, journalists lack time to properly evaluate the reliability of a video or a photo, for example.	• Provision of automated/ assisted assessment to determine whether an image, video, audio or text is human-created or machine made. • Provision of automated/ assisted assessment of the determination of the reliability of text/video/photo/audio (both Ai generated and user generated content) in order to reduce the time required to journalists to assess a content

■ 2.3.3 Stakeholder engagement of media practitioners, fact-checkers and policy makers

The Stakeholders that Sky TG24 will involve in the pilot project journalists:

- 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, TBC.

■ 2.3.4 Organisation of participatory workshops and training to plan the testing and validation activities

In preparation for the test and validate the AI4TRUST Platform through its usage in real scenarios:

- Introduction: The Platform will be introduced to SkyTG24's journalists, to familiarise them with the AI4TRUST platform and its functionality. Definition of the goal of the pilot project and define clear validation criteria.
- Objectives: usability testing, evaluation and validation criteria, user's feedback.
- Testing Methods: usability testing in in real scenarios, interactive sessions and focus on how to use the tool in daily journalistic activities.

- Feed-back: Open discussion sessions to gather suggestions, opinions, directions, also about ethical considerations about AI and its use in the fight against mis-disinformation. Questionnaires made in conjunctions with project partners.

2.4. ELLINIKA Pilot

Ellinika Hoaxes (EH) is a Greek non-profit fact-checking organisation, the first in Greece to coordinate separate related initiatives and get certified by the IFCN. EH is solely focused on fact-checking misinformation and disinformation, with no news division. EH has provided a full dataset since 2019 of its articles, fact-checked content, and related metadata to AI4TRUST, and aims to become an end user of the platform.

■ 2.4.1 Fact-checking state of the art analysis and best practice search

At EH, the team uses Workplace by Meta, which is essentially a Facebook clone for private teams. It features both Chats (Slack-like channels) and Groups (which contain posts with individual comments and replies to them). These Groups are especially useful for collecting content that should be fact-checked, and aggregating relevant claims and data. Ideally the AI4TRUST platform would be able to integrate the information from this platform in the future.[8]

To collect claims for fact-checking, we rely on a combination of channels: our own everyday manual searches, community messages (mostly through Facebook Messenger but also through an email form or direct emails) and a Meta provided platform (which is not accessible to entities outside the Meta TPFC program). Individual fact-checkers take on manual searches continuously, but we also have a dedicated comms specialist who reviews the channels and collects similar content in Workplace Posts.

Next, for the fact-checking itself, the first step is analysis of content (e.g. text, image, video), isolating specific claims, grouping the similar ones, and collecting the relevant data available to support or refute each group of claims.

Techniques for collecting the fact-checking data can vary widely depending on the specific type of claim and the fact-checker's abilities. A good introduction to our techniques has been published in short tutorial videos[9] by the AFP. For example, to check a claim about a depicted incident, the fact checker looks for the oldest source of the image on the internet and tries to place it to its specific context; geolocation or personal contact may be needed.

However, a scientific claim will be checked very differently, with an analysis of up-to-date guidelines by authorities and renowned organisations, as well as reviews of primary research, which requires relevant knowledge.

Regarding the tools used, the team employs a variety of tools tailored to the specific requirements that may arise during EH's operations. While you can find tutorials for fundamental tools such as Google Search, Reverse Image Search, Maps, and Translate in the AFP resources, there's also an array of advanced tools available. A recent compilation of these advanced tools can be found in this presentation[10] given at the latest GlobalFact, an international fact-checking conference.

A notable recent development is the new platform of Google Fact Check Tools[11] which lists the history of searched images (see the presentation linked above). Integrating AI4TRUST with the Fact Check suite would be particularly beneficial for the platform. The latest category of tools used by EH are recently developed AI tools.

Although caution is still required due to their limitations, a few high-level tools are already proving to be very useful with careful application. ChatGPT Plus with GPT-4 can combine a quite thorough understanding of different fields to guide the checker on their search; its more recent additions (API plugins) allow for semantic search in YouTube transcriptions and PDF text, as well as (Code Interpreter) data analysis and visualisation. Also, Bing Chat in Creative mode (which again uses GPT-4) can search the internet live in many languages and offer fast appraisals of arguments; currently this search can't utilise the full background knowledge of GPT-4, but it aims to be up to date, beyond the cutoff training date.

LLMs can be challenging and expensive to deploy successfully, but there have been major strides in recent months with open-source products,[12] such as Llama 2 by Meta, which can be fine-tuned for specific uses.

■ 2.4.2 High level requirements of Pilot

Firstly, EH will provide insights on how they wish the platform could work as a final product, echoing EH's comments in the User Requirement Form. This will state EH's thought process more clearly than the table form, which is provided below.

The description below serves as an idealised concept and not for a pilot. Perhaps it would also not be fully technically feasible with current constraints, but it could provide ideas for certain features.

Main page: list of currently viral and/or impactful stories on my selected region. EH can change the criteria for ranking and theme/ geographical region covered (e.g. focus more on

virality or hate speech, and focus on Greece or EU). Each element of the list is a synthesis of related stories, like how Google News groups them.

Each story has metrics for each ranking criteria, with details available (e.g. how virality has changed during the last hours/days).

Clicking on each story, EH can see all the sources promoting it, but the original is pointed out. Then analysts get the summary of the story, and separately, its main claims, and if applicable, who exactly made the claim. An estimate of how trustworthy each claim is provided, as well as a summary rating/debunking it.

Selecting each claim, the source promoting the claim opens, the places of the article making the claim are highlighted, and clicking on them, analysts can read the rating with the sources the debunking is based on. In this page, staff members can quickly see and click on all the claims/conclusions, like how comments work in Google Docs.

From this page, analysts can also get a list of who are the most pertinent cooperating experts and can message them on the same platform. They are able to provide a full text response but also make comments on the page like Google Docs.(For the communication with experts, one practical model has been developed[13] by Science Feedback. Here's an article example[14])

This platform could also offer a separate "forensics" mode, more manual, where analysts can apply a semi-automated fact-check process with a chatbot, and I could use tools on specific images or videos for analysis, geolocation, and reverse search.

Next, the table below indicates the top areas of possible improvement in the context of the EH experience.

AI4TRUST service	Relevant process and current gaps	Desirable improvement
Infodemic trends for specific public interest issues	As stated, currently many channels are needed to reliably detect misinformation trends, and despite quite thorough approaches, some content may still be missed or be detected after they have gone viral for days.	• A new platform could continuously monitor social media and webpages with a history of untrustworthy content to detect such signals as early as possible. • Another idea is to implement a browser extension and/or smartphone app which will be able to utilise signals by the community.
Archiving of sources	Link Rot[15] is an important problem in internet sourcing, and especially fact-checking that aims to stay on the record. We aim to archive all our sources, especially suspect ones, so that their original text can stay accessible. This too is time consuming. Currently there are 3 main services: the Wayback Machine (WM, which can be clunky), Archive.today (AT, which is handy but opaque[16]) and Perma.cc (which allows for easy batch archiving but requires a subscription).	A new platform could allow for mass selection of sources providing a claim, as well as optionally the links in their own text. Then, it could check if there are already archives of them, and if yes, verify that they work (especially WM returns errors frequently), otherwise archive them in WM, AT or both. Lastly, check when the archive links are ready and provide them for easy copy on click.
Reliability analysis of information sources	Currently, the fact-checker has to manually check or remember which pages have an untrustworthy track record and has to manually rate the sources cited in articles.	• A new platform could automatically rate the trustworthiness of pages presenting suspect claims, as well the trustworthiness of their sourcing. • MBFC, Wikipedia, and fact-checking organisations host collections of such records, but they need to be integrated together and work in a practical application.

Claim collection and grouping	Again, the fact-checker has to manually do it, and it can be time consuming for large poorly written articles and especially videos.	A new platform could do that automatically. It would be especially useful if video transcription to text could also be automatic, with correspondence between text and video selections (e.g. how HappyScribe[17] does it).
Claim appraisal	This is the most difficult part of fact-checking. Requires multiple searches, domain specific knowledge, focus, extensive cross-referencing between lots of sources. It commonly runs recursively, where one search turns up info that needs its own separate searches.	• An ideal new platform would be able to offer both wide-background appraisals of arguments in the style of ChatGPT 4 and also summaries of up-to-date internet searches in the style of Bing Chat Creative. The second case is what's more directly needed to the average current affairs fact-check, but a Microsoft executive has stated that it would be "too pricey" to offer through API. And even just Bing search has had a large increase in pricing. • Another challenging part is sourcing, as the fact-checker has to be able to verify everything the AI claims; unfortunately, for now, even state of the art LLMs can make up plausible sources or provide unrelated URLs as citations, even when the explanations are correct.

2.4.3 Stakeholder engagement of media practitioners, fact-checkers, and policy makers

The stakeholders all belong to the EH team, and consist of at least 6 fact-checkers, 1 team editor and 1 AI4TRUST project manager.

2.4.4 Organisation of participatory workshops and training to plan the testing and validation activities

Due to the relatively small size of our team, the process of the workshops will be simple. They will be conducted locally in our main office. First, a presentation will introduce the team to the tool, its functions and its advantages. Then, the editor and project manager will show a few practical examples and help the team members familiarise with the tool one by one. Next, the team members will explore the tool further on their own, try to apply it to everyday work and integrate it to their current processes. The editor and manager will note any questions, difficulties, possible improvements, and address live what they can. Afterwards, the team members will fill questionnaires regarding their experience and the manager and editor will create a report.

2.5. EURACTIV Pilot

EURACTIV, an independent pan-European media network, was founded in Brussels in 1999 and since then has become a well-respected source of wide-ranging, unbiased information on EU affairs. Specialised in a range of EU policy areas including Energy & Environment, Economy & Jobs, Politics, Digital, Agrifood, Global Europe, Health, and Transport, EURACTIV sparks and nourishes policy debates among stakeholders, including government, business, and civil society.

EURACTIV fact-checking activities have been enhanced thanks to the TRUE INFO project, where a considerable amount of time has been devoted to the creation of verifiable and trustworthy content related to the Russian invasion of Ukraine. To analyse the flow of content across platforms, languages and websites, as for identifying the accounts involved in spreading news, different tools were employed. These relied on a wide array of technology types (e.g. crowdsourcing, source labelling, metadata, blockchain, machine

learning) and were used to detect false claims in the forms of text, image, video or audio, facilitating the work of EURACTIV journalists.

Moreover, EURACTIV joined the Trust Project in 2023. This is a global network of news organisations who adhere to a set of eight "Trust Indicators" that represent a gold standard for trustworthiness and transparency in media. The project works with technology platforms to affirm and amplify journalism's commitment to transparency, accuracy, inclusion and fairness in order to help the public make informed news choices.

■ 2.5.1 High level requirements of EURACTIV Pilot

AI4TRUST service	Relevant Processes and Current Gaps	Desirable Improvement
Integrating fact-checking work in the daily journalistic process.	• Time constraints: journalists may lack time to engage in fact-checking activities because of the continuous flow of news they have to report on. • Staff and budget constraints: employing fact checking tools often come with monetary and resource constraints, especially because staff members need to be constantly updated and trained on the continuous emergence of novel AI tools and ever-changing tech landscape.	EURACTIV journalists will have better access to fact-checked multimodal content through automatization, which will be useful to provide trustworthy news in less time.
Managing fact-checking across multiple languages and macro/micro topics	**Fact Checking tools being language specific:** tools being constrained by designated keywords, limiting the processes to a primary language.	Fact-checking to be done in multiple languages - determining the effectiveness of AI tools. Mechanisms for enriching the keyword pool to be included in the platform.
Accuracy of tool in written media	**Variety of topics having different forms of**	Reliability of fact-checking across varied topics, from climate and

	disinformation: the distinction between mis and disinformation is important between different topics.	energy to European politics.

■ 2.5.2 Stakeholder engagement of media practitioners, fact-checkers and policy makers

The stakeholders that EURACTIV will involve in the pilot-project are the following:

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

■ 2.5.3 Fact-checking state of the art analysis and best practice search

Europe's media and information sector has fully entered the 'multi-modal dimension'. The media sector is undergoing continuous innovations that occur at a pace never seen before. Besides tackling established challenges like ownership concentration and declining trust in the media, Europe is currently driving decisions of huge societal import: e.g. how to regulate AI, whose speed of employment has been exponential. The exploratory phase is coming to an end, and we are already grappling with the impact of AI technologies that could produce actual value, both editorially and for news businesses.

As with every major technological innovation, the discussion verges on opportunities vs. risks. On the one hand, numerous opportunities, namely tools such as ChatGPT, Midjourney and DALL-E to automate text writing and image creation, are already embedded in the media sector, with the possibility for journalists and fact-checkers to verify digital content in an advanced way and thus offer higher quality information. On the other hand, there are many risks involved. From copyright infringement to the lack of transparency, through the creation of new monopolies and the impact on the labour market, to the issue that most concerns experts today: the possibility of mass creation of manipulative content – of crucial importance for the future of the media.

■ 2.5.4 Organisation of participatory workshops and training to plan the testing and validation activities

Using EURACTIV's already established database of journalists in multiple languages, AI4TRUST services and tools will be shared with the media target group to determine the effectiveness of their ability to combat disinformation in written media.

- Introduction: The AI4TRUST tool will be introduced to EURACTIV's journalists and fact-checkers.
- Selection of the journalists and fact-checkers done through internal channels to ensure a varied sample of testers.
- Testing method and practice: The usability of the tool will be tested across different languages and topics. The journalists will be given the opportunity to use the tool in their everyday life in content creation to determine its realistic practicality.
- Evaluation: Questionnaires will be created based on the primary concerns of AI4TRUST and pilot cases - considering both quantitative and qualitative issues. Questionnaires will be shared with the journalists' post-practice to understand the advantages and disadvantages of the tool's efficacy.

2.6. ADB Pilot

Association Digital Bridge (ADB) operating EURACTIV in Romania is organised as a media NGO working to improve quality journalism in a country with a heavy consumption of internet (7 hours/day) and television (over 3 hours/day). Television remains the main source for news (90%, the highest in Europe - Eurobarometer 2022[18]), and videos have significant impact in the online environment. Media are extensively influenced by political parties, through financial means, and disinformation is widely spread (reports by media[19]).

Within this challenging context, ADB, operating under EURACTIV Romania, specialises in the realm of high-quality journalism, particularly focusing on public policies. Fact-checking in relation to public policies is extremely important, and ADB launched in 2019 *Facts, not Fake*,[20] a section on which in-house journalists and freelancers from Romania and Republica Moldova focused on dismantling major disinformation narratives identified in both countries. The fact-checking was done through specific journalistic techniques (including research, on the ground documentation). ADB is highly interested in upgrading this initiative by utilising AI tools for Facts, not Fake. From a specialized platform like euractiv.ro, these techniques can be disseminated and expanded to engage other teams of journalists, thereby amplifying their influence. The objective is to apply debunking techniques, defined as the process of dismantling false or distorted information, not only

after it has been published but also during the journalistic research phase, prior to article publication.

ADB cooperated with freelance reporters and started to develop partnership with fact-checking organisations such as Funky Citizen (Bucharest), Misreport (Cluj), Stop Fals (Chisinau, Republic of Moldova), and has created guides on tools and databases that can be used to tackle disinformation. Our primary objective is to enhance journalists' capacity to use automated tools in the process of fact-checking (either it is before publication - in the routine journalistic process or in the post-publication process - debunking). Currently, journalists don't have fact-checking units in their newsrooms or they remain rare and few journalists are trained in using existing AI tools.. Also, there is sporadic or almost non-existent cooperation between fact-checkers and media organisations.

ADB has a small team of professional journalists in Bucharest and in Republica Moldova and has partnerships with key media organisations like HotNews.ro and over 30-member PressHub, activating at the local level, across the country. ADB launched its section *Facts, not Fake*[20] in 2019, which was extended to the PressHub network; however, the work was done almost entirely manually, based on specific journalistic research techniques. Through the current pilot project, automated tools to support the monitoring of the widely spread content and the fact-checking process will make it more efficient and potentially able to multiplicate it.

■ 2.6.1 High level requirements of ADB Pilot

Facts, not Fake AI unit - The team reporting for Facts, not Fake will be trained in using AI tools for monitoring disinformation linked to public policies, most of it in video/audio content.

AI4TRUST service	Relevant processes and current gaps	Desirable Improvement
Integrating fact-checkers' work in the journalistic process. Journalists' access to debunked content which can be utilised in two ways: • to prevent the continued dissemination of disinformation narratives; • to bring the debunked content to the forefront of the media	In Romania, the establishment of dedicated debunking units within major newsrooms is exceedingly rare. ADB seeks to pioneer the creation of a pilot unit that seamlessly integrates the work of professional fact-checkers into the daily routines of journalists. The current fact-checking process is	To enhance and streamline the debunking process, ADB envisions the implementation of several automated features, including: • Automated Scanning: This involves the automated scanning of fact-checkers' public databases in various languages. The aim is to identify major disinformation narratives that can be effectively

agenda, through agenda-setting techniques.	largely manual and sporadic. It involves: ● Manual Verification: Fact-checkers manually check various sources, including fact-checking websites. ● Time Constraints: Journalists face significant time constraints in engaging with debunking activities, as they are primarily focused on real-time reporting. ● Limited Access: There is a notable absence of rapid access to comprehensive fact-checking databases. This lack of accessibility hampers the swift identification of similar falsehoods, whether archived or in different languages, that have been previously debunked.	dismantled through journalistic pieces, making this valuable information accessible to a larger audience. ● Automated Alerts: Journalists will benefit from automated alerts regarding debunked content that is of particular relevance to their areas of expertise. This feature ensures that journalists are promptly informed about disinformation narratives within their domain. ● Improved Access: The implementation of automation will significantly enhance access to debunked content, streamlining the process of identifying and utilizing valuable information in the fight against disinformation.
Audio/video monitoring of key debunked content	● Sporadic due to human resources involved. ● Difficult to implement in real time (live debunking) ● limited access of journalists to substantial audio/video Transcriptions and Translation Integration	Automated transcriptions of key audiovisual content (i.e. debates on public policies), when needed, potentially integrated with automated translation
Explanatory pieces focused on fact-checking process	Explanatory pieces have focused mostly on the credibility of the sources and of the piece of information, and less on the fact-checking and debunking process as such. It is important to show to larger audiences what tools to use to try to be themselves fact-checkers.	Automated identification of relevant explanatory pieces of the debunking process.

2.6.2 Stakeholder engagement of media practitioners, fact-checkers and policy makers

The stakeholders that ADB will involve in the pilot-project will consist of in-house or freelance journalists and partnering fact-checkers, as follows:

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

2.6.3 Fact-checking state of the art analysis and best practice search

There is an acknowledged gap between fact-checked content and journalistic activities in Romania. Journalists don't have time and are not trained in debunking (correcting and debiasing disinformation after a certain item is published) - studies show (Newsreel, 2022[21]). The act of debunking is either limited or virtually absent within most Romanian newsrooms. This stems from the absence of specialised teams and the rapid pace inherent in journalistic work. During a conference in Bucharest, in June 2023, chief editors and seasoned journalists from the three major television networks acknowledged the absence of debunking units in their respective newsrooms. On rare occasions, debunking occurs primarily through journalistic means, occasionally involving the citation of fact-checkers.

ABD is a non-governmental organization (NGO) operating the specialised website EURACTIV in Romania, with a focus on public policies; with a small team, it has the flexibility to pilot a potential fact-checking alert unit in its newsroom.

Since 2019 ADB's journalists have piloted *Facts, Not Fake*[20] section exposing major disinformation narratives identified in Romania and Republic of Moldova and publishing investigative stories and reports on-the-ground (reportages) or/and interviews with experts in an effort to dismantle fake. For this section, ADB has worked with freelance reporters and currently works at developing key partnership with fact-checking organisations such as Funky Citizens (based in Bucharest, Romania), Misreport (based in Cluj, Romania) and with Stop Fals (Association of Independent Press, based in Chisinau, Republic of Moldova). Also, ADB's journalists closely followed platforms tackling disinformation such as EUvsDisinfo.

Our primary objective is to enhance the organisation's ability to identify and dismantle prevalent instances of fake news in order to set up a pilot project of testing AI tools for fact-checking in a newsroom environment. Currently, this is achieved through manual monitoring and fact-checking, primarily employing journalistic methods. Additionally, our

journalists would benefit from further training in effectively utilising existing AI tools to bolster their efficiency with a multiplying effect, as ADB is a respected outlet in terms of qualitative journalism and media ethics, with key partnership with key media and fact-checking organisations.

Quality journalism mandates the verification of information prior to publication, whereas debunking involves confirming the accuracy of information already disseminated. Importing the debunking techniques in the day-to-day journalistic routine will increase the quality of fact-checking in daily mass reporting.

The process of fact-checking is often intertwined with journalistic research and is perceived as an integral component of it although there are clear procedures followed by fact-checkers to which Romanian journalists aren't always trained to use. The second main objective will be to pilot the automated tools as potential bridge between the fact-checkers community (mostly done by NGOs) and journalistic professionals, as those communities are perceived as working separately in Romania.

■ 2.6.4 Organisation of participatory workshops and training to plan the testing and validation activities

- Briefing workshop of the piloting unit to test the AI4TRUST platform
- Objectives: The primary objectives of this workshop are to collect valuable user feedback and conduct usability testing of the AI4TRUST platform.
- Testing Methods: Usability testing will be performed in the piloting unit's own working environment. This testing will be followed by feedback, which can be provided in written form or through interviews, or a combination of both.
- Participants: Members of the piloting team will actively participate in the usability test. They will use written questionnaires, developed in collaboration with project partners, to provide feedback. In cases where more detailed documentation is required, interviews may be conducted.
- Timeline: The planning of this workshop will align with the project timeline. Access credentials and necessary resources will be provided to facilitate the testing process.
- Communication of Results: The conclusions drawn from the usability testing will be communicated to the relevant parties through one of the indicated methods—questionnaires, interviews, or a combination of both. This will ensure that the insights gathered during the workshop contribute effectively to the ongoing development of the AI4TRUST platform.

2.7. EMS Pilot

Europejskie Media SP ZOO (EMS) operating EURACTIV in Poland is a media organisation, focusing mainly on providing its readers with reliable and unbiased information, including expert opinions and commentary. Striving to build resilience against disinformation, EMS is also extensively involved in fact-checking activities.

In recent years, the ESM has been involved in numerous projects dedicated to the fight against disinformation. As part of our Media Against Disinformation project, EMS focused on detecting and combating disinformation that targets the European Union, its Member States, politicians, and the general public, while providing reliable information. Our activities were aimed at both politicians and decision-makers at the European level (to explain the effects of their actions) and to action experts (to offer non-political reporting based on data and expertise). As a result, we focused on creating more durable and structured tools to ensure media literacy in Poland, increasing our team's fact-checking competence, working out new ways to deliver verified content and spreading the truth and disproving history beyond our current readership. As part of the project, we created a toolkit on various available tools, how to approach different pieces of information, how to consume information critically and consciously, and also provide advice to empower citizens in their daily news. Our journalists and fact-checkers will check their usefulness first personally. We also create a series of articles (Fake of The Month) debunking false information appearing in the public sphere, as well as a series of podcasts - conversations with experts on disinformation.

Within the Checks4Media East project, our journalists focused on turning fact-checking into a sustainable media activity in Eastern Europe as well as in terms of business model beyond short-term journalistic work, using experience and good practices from Europe and abroad, as well as using proven Central European networks, the project will explore possible models and apply experiments in a sandbox environment.

As part of our activities in the Immunion project, we focused on combating misinformation and disinformation about vaccines through a series of articles and interviews with experts. Despite the relatively small size of the team (less than 15 people) which operates all over Poland, all EMS journalists have thorough training and engage in fact-checking activities.

■ 2.7.1 High level requirements of EMS Pilot

AI4TRUST Service	Relevant Processes and Current Gaps	Desirable Improvement
A tool enabling seamless integration of daily journalism with real-time fact-checking capabilities.	Enhanced devices for assessing the accuracy of online information.	Providing new, intelligent solutions.

■ 2.7.2 Stakeholder engagement of media practitioners, fact-checkers and policy makers

The stakeholders that EMS will involve in the pilot-project will consist of journalists and fact-checkers:

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

■ 2.7.3 Fact-checking state of the art analysis and best practice search

In Poland, fact-checking and countering disinformation present a complex landscape characterised by both challenges and opportunities. The country has seen a rise in the dissemination of false information, particularly in the digital realm, where social media platforms are often used to spread misleading content. According to a study conducted by the Reuters Institute for the Study of Journalism,[22] approximately 44% of internet users in Poland encounter news that they believe to be false at least once a week.

A major challenge lies is the polarisation of media consumption and the prevalence of echo chambers, which reinforce existing beliefs and hinder the acceptance of accurate information. This phenomenon is evident in the context of political, social, and cultural debates.

■ 2.7.4 Organisation of participatory workshops and training to plan the testing and validation activities

Methodology for Evaluation: Commencing with usability assessments, succeeded by a synthesis of feedback (captured via a feedback form) coupled with a collaborative dialogue within focus-group sessions.

Enlistment of Participants: EMS will extend invitations to journalists, fact-checkers, and media experts, encouraging their involvement in this focused group.

A Span of 30 days will be allocated for testing, with furnished access credentials. The evaluation will be executed within their operational setting. The outcomes will be conveyed through a written survey, harmonised with the undertakings of the other collaborators.

3. Pilots' execution Work Plan for the two piloting sessions

In the context of the presented pilot design framework, the section outlines the guidelines and core actions for implementing the piloting sessions effectively and efficiently.

The work plan of each pilot establishes the structure and framework for carrying out all tasks and activities required for the pilot's realisation. It furthermore specifies involved partners and their roles, key milestones in the execution process, and the factors expected to be monitored during the pilot and contribute to the evaluation process described in Section 4, as well as the means to monitor them.

Additionally, we present for each pilot the nature and timing of actions directed towards engaging the community relevant to the pilot, making the purpose and scope of the pilot clear, and ensuring their interest and involvement.

3.1 Pilots' execution Work Plan guidelines

To ensure that the objectives of the pilot sessions are met, and the users involved provide meaningful feedback and are engaged in the process, the following directives should be pursued by all pilots when defining and timing their individual work plans.

1. Define Objectives: Clearly define the objectives of the online testing sessions. Identify the specific goals you aim to achieve through the testing process, such as gathering user feedback, identifying usability issues, or evaluating the effectiveness of a product or service.

2. Identify Testing Methods: Determine the appropriate testing methods based on your objectives. This could include surveys, usability testing, focus groups, interviews, or remote observation sessions. Choose methods that align with your goals and the resources available.

3. Participant Recruitment: Develop a strategy for participant recruitment. Determine the target audience for the testing sessions and outline how you will reach out to them, whether through online advertising, targeted invitations, or leveraging existing user communities.

4. Schedule and Duration: Determine the schedule and duration for each testing session. Consider the availability of participants and allocate sufficient time for each activity,

allowing for breaks and potential technical issues. Ensure that the schedule accommodates participants from different time zones if necessary.

5. Pre-session Preparation: Provide clear instructions and materials to participants prior to the testing sessions. This may include access credentials, briefing documents, or software installations. Communicate the purpose of the session and any specific tasks participants should be prepared to complete.

6. Test Environment: Set up a reliable and user-friendly test environment for participants. Ensure that any necessary software, tools, or platforms are accessible and properly configured. Test the environment beforehand to identify and resolve any potential technical issues.

7. Facilitation and Moderation: Assign a qualified facilitator or moderator to guide participants through the testing sessions. The facilitator should be experienced in online testing methodologies, capable of ensuring a smooth and productive experience for participants.

8. Documentation and Observation: Establish a system for documenting and observing the testing sessions. This could involve capturing video recordings, taking notes, or utilising screen-sharing and remote observation tools. Ensure that all necessary data is collected accurately and securely.

9. Data Analysis and Interpretation: Develop a plan for analysing and interpreting the collected data. Determine the key metrics, themes, and patterns to look for during the analysis process. This will help derive valuable insights and actionable recommendations from the testing sessions.

10. Reporting and Action Steps: Prepare a comprehensive report summarising the findings, insights, and recommendations from the online testing sessions. Clearly outline the action steps to be taken based on the results, including any necessary adjustments or improvements to the tested product or service.

11. Post-testing Engagement: Communicate the outcomes of the testing sessions to participants and involve them in the feedback loop. Share how their input has influenced decision-making or product improvements, fostering a sense of collaboration and community engagement.

3.2 Pilots' community building plan guidelines

To ensure broad and meaningful participation from the targeted stakeholder groups, the community building plan to be built by each pilot will adhere to specific guidelines and best practices. Namely, the following principles will be considered:

1. Clear Communication Channels: Establish clear and accessible communication channels for community members to ask questions, provide feedback, and seek assistance. This could include dedicated email addresses, online forums, or a designated support system.

2. Transparent Information Sharing: Pilot representatives will provide timely and transparent information about the purpose, goals, and process of the testing sessions. They will also share details about how community members' feedback will be used and any potential impact their involvement may have.

3. Clear Instructions and Expectations: We will clearly outline the instructions, expectations, and timeline of the process, and provide step-by-step guidance on how participants are expected to be involved, what tasks they are expected to perform, and any specific criteria or guidelines they should follow.

4. Feedback Collection Mechanisms: Implement effective feedback collection mechanisms, to gather insights from participants. Encourage participants to provide detailed and constructive feedback that can contribute to the improvement of the testing process.

5. Acknowledgment and Recognition: Recognize and appreciate the contributions of community members who actively participate in the online testing. This can be done through public acknowledgments, certificates of participation, or incentives to show gratitude for their time and effort.

6. Regular Updates and Progress Reports: Keep the community informed about the progress of the pilots and the AI4TRUST services. Provide regular updates on key milestones, insights gained from community feedback, and how their input has influenced the project's development.

7. Privacy and Data Protection: Ensure the privacy and data protection of community members participating in online testing. Clearly communicate how their personal information will be handled, stored, and used, adhering to relevant privacy regulations.

8. Continuous Engagement: Foster ongoing engagement beyond the testing sessions by maintaining a community platform or forum where participants can stay connected, share their experiences, and continue to provide input for future developments.

3.3 MALDITA Pilot execution and community building plan

■ 3.3.1 MALDITA Preparatory phase

The preparatory phase of the Maldita pilot will run for a period of 1 month and involve the following:

1. Communication Channels: A point of contact for the pilot will be established in the organisation as representative of the pilot to communicate directly with stakeholders involved throughout the pilot. In the lead-up to and during the pilot execution, weekly meetings will be held between all stakeholders to monitor implementation of the pilot and ensure clear communication throughout. Organisation emails will be used as an additional line of communication to share written and digital information and materials.
2. Participant Invitation: Involved stakeholders will be informed of their involvement and negotiations carried out to ensure the effective implementation of the pilot that aligns with stakeholder timetables and workloads.
3. Briefing Content Preparation: Prepare detailed briefing documents and instructional materials for participants. Clearly communicate the purpose of the testing sessions, specific tasks, and expected outcomes.
4. Access to Tools/Platforms: Provide participants with access credentials to the AI4TRUST platform and any necessary software or tools. Run functionality tests with stakeholders to ensure they are able to access the platform problem-free.
5. Stakeholder Training Workshops: Delivery of training sessions outlined in 2.1 for participants to familiarise themselves with the platform and online testing methodologies. These trainings will be conducted online with staff in the Maldita newsroom.
6. Definition of pilot objectives: Following the training with stakeholders, objectives and KPIs for the pilot testing will be collaboratively defined and communicated to all involved via the weekly pilot meeting and email.

■ 3.3.2 MALDITA Execution phase

The pilot execution phase will occur over a period of 1 month and involve the following:

1. Testing sessions: parallel testing sessions will run continuously throughout the month of execution: a) the fact-checkers and editors use of the tool to assess sources, monitor infodemic risk, and support fact-check content production; b) media literacy experts to evaluate the use of the platform for content development; c) public policy experts for use of the tool to enhance their activities; d) Maldita provision of data to the platforms.
2. Monitoring and Moderation: Qualified facilitators and moderators will guide participants through the testing process. At weekly meetings, stakeholders will report back on the testing implementation, address technical issues, and facilitate discussions.
3. Feedback Collection: Feedback will be collected through a survey at the close of the pilot for each stakeholder type, occasional screen-sharing sessions, and note-taking during the weekly feedback meetings. Participants will be encouraged to provide detailed and constructive feedback.
4. Data Analysis for KPIs: The collected feedback will be analysed to assess the defined Key Performance Indicators (KPIs). Metrics related to user satisfaction, ease of use, and effectiveness to improve current processes and address will be evaluated.

■ 3.3.3 MALDITA post-pilot phase

The post-pilot phase will occur over the period of 1 month:

1. Feedback Aggregation: The feedback from all testing sessions will be aggregated and compiled into a comprehensive report. The report will highlight key findings, insights, and recommendations.
2. Evaluation Team: The report will be forwarded to the evaluation team responsible for assessing the AI4TRUST platform's performance against the defined objectives and KPIs.
3. Post-Event Communication: Participants will be informed of the outcomes and impact of their feedback via email and in a final review meeting. They will also be facilitated with the evaluation report.

3.4 DEMAGOG Pilot execution and community building plan

■ 3.4.1 DEMAGOG Preparatory Phase

The preparatory phase will include:

1. Designated Demagog Coordinator to be the main point of contact for internal communication.
2. Selection of the internal team based on specific expertness.
3. Channels and introduction materials set up to ensure smooth communication.
4. Inform the team about planned activities, deadlines, schedule of the execution phase and KPIs.

■ 3.4.2 DEMAGOG Execution Phase

The Execution phase will include:

1. The pilot team will participate in pilot workshops aimed to test the AI4TRUST platform.
2. The feedback will be provided in questionnaires updated on ad hoc analysis to ensure relevant feedback is recovered.
3. Conclusions will be based on the experience of working with the tool by individual analysts - considering specialisation and experience in particular information sectors. One-on-one meetings and an open discussion summarising the conclusions.

■ 3.4.3 DEMAGOG post-pilot Phase

The post-pilot phase will include:

1. Feedback Aggregation: Analysis of the Questionnaires and metrics by the Project Coordinator and support staff in relation to the KPIs determined during the Preparatory phase; summarising conclusions in the report.
2. Evaluation Team: The report will be shared with the team to agree on general conclusions.

3. Post-Event Communication: Evaluation Report will be shared with the Work Package 5 Leader.

3.5 SKYTG24 Pilot execution and community building plan

■ 3.5.1 SKYTG24 Preparatory Phase

1. Communication channels setup: Sky TG24 will name a coordinator, in charge of organising the Execution phase and set up proper communication channels with involved stakeholders.
2. Participant Invitation: Sky TG24's coordinator will select a group of journalists (as per 2.3.2) that will take part in the Execution phase.
3. Contents preparation: Sky TG24's coordinator, alongside other professionals involved in the project, will prepare a list of instructions for the benefit of stakeholders. These instructions explain the nature of the project, how the tool works and the goals of Piloting sessions.
4. Platform access: Stakeholders involved in the Execution phase will be granted access to the designed platform. To reduce risks of potential problems in the Execution phase, the ability to access the platform will be checked in advance.
5. Stakeholders training: Stakeholders will be provided with all the necessary information to successfully run the tests. At least one meeting will take place between coordinator and stakeholders, to share potential issues and to verify the correct functionality of communication channels.

■ 3.5.2 SKYTG24 Execution Phase

1. Piloting sessions: Sky TG24's stakeholders will run the Piloting sessions focusing on day-to-day realistic scenarios, applying the AI4TRUST tool - whenever possible - during the daily news reporting activities of Sky TG24
2. Monitoring and moderating: Sky TG24 team's coordinator will take care of monitoring and moderating the process, addressing potential issues throughout the piloting session.
3. Feedback collection: Feedbacks will be collected based on journalists' experience throughout the session. To ensure a better understanding of each stakeholder's

experience, individual and group meetings are expected to take place. Questionnaires shared among other partners will play an important role in gathering all the stakeholder's experience.

3.5.3 SKYTG24 post-pilot phase

1. Feedback aggregation: Sky TG24 team coordinator and support team will prepare a report based on the feedback received during the Execution phase. A report draft will be shared with stakeholders to ensure the best possible reporting on the Piloting session, gathering both feedback on the tests and suggestions on possible improvements of the tool.
2. Post-event communication: Final report will be shared with WP leader and other relevant stakeholders of the project.

3.6 ELLINIKA Pilot execution and community building plan

3.6.1 ELLINIKA Preparatory Phase

The AI4TRUST project manager will oversee the coordination of the workshops. He will cooperate closely with the team editor to familiarise themselves with the tool, prepare the introductory presentation, plan the execution, anticipate possible issues or questions. The process is expected to last 1 month or less.

3.6.2 ELLINIKA Execution Phase

The workshops will run locally in our main office. In the first workshop, an introduction will be presented by the project manager and team editor regarding the functions and advantages of the tool. Then, they will show a few practical examples of how the tool can be used.

Afterwards, they will aid every team member, one by one, to work with the tool. Next, each team member will try to explore the tool and apply it to everyday work. The editor and manager will note any questions, difficulties, possible improvements, and address live what they can. They will keep notes of the process, which will be recorded with video and audio.

For the next workshops, each team member will pick different functionalities of the tool to test in depth, which are more pertinent to their everyday work. The editor and manager will again be ready to note and help, as before.

For the last sessions, the team will try to develop new guidelines of how the tool can be integrated with its already established processes.

Finally, each team member will fill questionnaires with their experiences, will note their own opinions and possible improvements of the tool. This process is expected to last about 1 month.

■ 3.6.3 ELLINIKA post-pilot phase

The previous notes will be aggregated, cross-checked, and discussed with the notes of the manager and editor. They will compile the findings into a report, with both a summary and detailed sections. An emphasis will be put to practical insights and recommendations.

This report will be shared with the team members for any final remarks. Then, it will be forwarded to the WP5 leader. This process is expected to last 1 month or less.

3.7 Euractiv Pilot execution and community building plan

■ 3.7.1 EURACTIV pilot Preparatory Phase

The preparatory phase will include:

1. Designated EURACTIV Coordinator to be the main point of contact for communication with journalists.
2. Channels set up to ensure smooth communication.
3. Evaluation of primary topics and concerns to be included during the Execution and Post-Pilot phase through KPIs. Results to be used in creating the Questionnaires.
4. Selection of journalists based on specific expertise and languages.
5. Discussion with journalists as to what the Execution phase will include - determining timing and availability. Briefing on what journalists should consider when using the tool.

■ 3.7.2 EURACTIV Execution Phase

The Execution phase will include:

1. Over the course of a month, 3 journalists will take part in testing workshops to use the AI4TRUST tool, while being monitored by the designated EURACTIV Coordinator to answer any questions.
2. Questionnaires will be updated based on ad hoc analysis to ensure relevant feedback is recovered.
3. Feedback will be received based on the journalists experiences - taking into account the different controls of topics and languages. One-on-one discussions with the journalists to go over the results of their Questionnaires to ensure all points were made.

■ 3.7.3 EURACTIV post-pilot phase

The post-pilot phase will include:

1. Analysis of the Questionnaires by the EURACTIV Coordinator and support staff in relation to the KPIs determined during the Preparatory phase.
2. Overall results will be shared with the EURACTIV journalists as a peer review and to ensure the overall conclusion matches their experience.
3. Evaluation Report will be shared with the Work Package 6 Leader.

3.8 ADB Pilot execution and community building plan

■ 3.8.1 ADB pilot Preparatory Phase

1. designating the coordinator of the testing unit of the AI fact-checking tools and the communication contact
2. designating the AI tools testing internal team (2 journalists, 1 AI fact-checking tools coordinator, 1 media expert-journalist)
3. Briefing of the journalists on the tools to be tested in accordance with the KPIs. Internal discussions for the schedule of the execution phase.

■ 3.8.2 ADB Execution Phase

The pilot unit will participate in testing workshops on the AI4TRUST tool. They will provide feedback and will be available for answering questions Questionnaires will be periodically revised following ad hoc analyses, ensuring that the collected feedback remains pertinent and valuable.

Input will be gathered through the perspectives of the journalists. Subsequent one-on-one discussions with the journalists will delve into the outcomes of their questionnaires, ensuring that all points are comprehensively addressed.

■ 3.8.3 ADB post-pilot phase

Analysis of the Questionnaires, summarising the main conclusions in conjunction with EURACTIV project partners and in relation to the KPIs determined during the Preparatory phase. Preliminary results will be shared with partners. Evaluation Report will be shared with the Work Package 6 Leader.

3.9 EMS Pilot execution and community building plan

■ 3.9.1 EMS pilot Preparatory Phase

- Designating a coordinator, who is the main contact person, responsible for facilitating all types of communication, such as contacts with journalists.
- Maintaining regular communication throughout the project (both within the team and externally).
- Keeping a dialogue regarding the implementation phase: Determining the exact timing and availability of journalists as well as preparing instructions for the well-thought-out application tool.

■ 3.9.2 EMS Execution Phase

The pilot unit plans to participate in a test workshop on the implementation of the AI4TRUST tool. During these workshops, it will offer feedback and answer any questions. The questionnaires will be reviewed regularly based on specific analysis, ensuring that the feedback collected remains relevant and valuable. EMS is ready to proceed as soon as it receives the tools for testing and the results from the technical teams.

■ 3.9.3 EMS post-pilot phase

The questionnaires will be reviewed and summarised into key findings in collaboration with EURACTIV project partners, while aligning with the predefined key performance indicators (KPIs) established during the Preparatory phase. Preliminary findings will be shared with

our partners for feedback. A comprehensive Evaluation Report will then be shared with the Work Package 6 leader.

4. Evaluation methodology for AI4TRUST

4.1 Evaluation objectives

The evaluation objectives are:

- Understand the ease of use for new users of the tool
- Evaluate the efficiency of the use for experienced users accomplishing tasks
- Determine what if any errors exist in the tool
- Compare the use of the tool with users' existing mechanisms of doing the same task
- Evaluate users' satisfaction with the tool
- Gather information about existing practices to refine the tool

4.2 Evaluation methodology

Evaluation of the pilot will entail online or in person walk-throughs with the user. Testers will come from the AI4TRUST team. The type of users should include experienced users and potential users for the final AT4Trust tool (i.e., both fact checkers and people who work with policy information). Qualitative and quantitative data should be captured.

■ 4.2.1 Questions for users

- How was the experience of using the product to complete this task?

  [Probe:] What are your thoughts on the language used?

  [Probe:] How easy or difficult was it to navigate?

  [Probe:] What are your thoughts on the design and layout?

- How would you describe your overall experience with this tool?
- What did you like the most about using this tool?
- What did you like the least?
- What, if anything, surprised you about the experience?
- What, if anything, caused you frustration?

- On a scale from 1 to 5 (1=not at all likely, 5=very likely), how likely are you to recommend this product to a colleague?
- How frequently would you use this product?

  [Option 1:] Never

  [Option 2:] Very Rarely (once per month)

  [Option 3:] Rarely (2-3 times/month)

  [Option 4:] Occasionally (2-3 times/week)

  [Option 5:] Frequently (1-2 times/day)

  [Option 6:] Very Frequently (3+ times/day)
- What could we do to improve this tool for your job or work?
- How do you see this tool being used in your work?

4.2.2 Questions for the tester's assessment

- Did the tool work?
- Overall rate the success of the task completion.
- Measure the time it takes for users to complete tasks.
- Note where and how challenges appear in the process.

4.3 Pilot KPIs for quantitative and qualitative assessment

The KPIs for the assessment are:

- Performance of the tool, measured by time it takes for users to complete a task.
- Ease-of-Use, measured by time it takes for users to complete a task and by qualitative questions above.

- Function, how users accomplish what they intend to do and how they assess the functionality.
- State-of-the-art improvement in workflow, measured by qualitative above.
- Operational KPIs, how this tool be integrated into workflow of detection and reporting this through assessment.

Conclusions

This report presents the piloting and evaluation framework for the AI4TRUST platform, via the organisation of relevant piloting sessions targeting different stakeholder types and their different needs. The deliverable contextualises the pilots, setting the main objective each of them pursues and the means and resources needed to accomplish it. Furthermore, it presents the methodology to be followed in order to collect meaningful feedback and properly analyse it and produce actionable and concrete insights for the further development of the platform.

As the AI4TRUST offerings mature, more specific assessment points will emerge, possibly leading to the relevant refinements on the pilot execution and feedback collection processes and tools. As such, the relevant methodologies will be likely re-assessed and calibrated for the second piloting cycle as foreseen in the work plan, always under the main principles and guidelines presented here.

References

1. https://tg24.sky.it/politica/2018/03/23/elezioni-twitter-bot
2. https://tg24.sky.it/argomenti/codice-youtube
3. https://tg24.sky.it/cronaca/2018/11/14/youtube-video-estremi-fake-news
4. https://tg24.sky.it/cronaca/2018/04/06/taurianova-fake-news
5. https://tg24.sky.it/politica/2018/03/08/fake-news-elezioni-2018
6. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023
7. https://reutersinstitute.politics.ox.ac.uk/digital-news-report/2023/italy
8. https://www.workplace.com/resources/tech/integrations/intro
9. https://www.youtube.com/playlist?list=PLrfXGOVayapnzAlBCtpR0sTreE-HjfMmU
10. https://docs.google.com/presentation/d/1T2iXCRsdO5plXqNEA46OlXge_9tC6glp_VJSgQuzNek/edit?usp=sharing
11. https://toolbox.google.com/factcheck/explorer
12. https://huggingface.co/
13. https://healthfeedback.org/process/
14. https://healthfeedback.org/evaluation/misleading-wall-street-journal-opinion-piece-makes-the-unsubstantiated-claim-that-the-u-s-will-have-herd-immunity-by-april-2021/.)
15. https://en.wikipedia.org/wiki/Link_rot
16. https://gyrovague.com/2023/08/05/archive-today-on-the-trail-of-the-mysterious-guerrilla-archivist-of-the-internet/
17. https://www.happyscribe.com/
18. https://www.europarl.europa.eu/news/en/press-room/20220704IPR34401/eu-citizens-trust-traditional-media-most-new-eurobarometer-survey-finds
19. https://www.balcanicaucaso.org/eng/Areas/Romania/Romania-the-interference-of-parties-in-the-media-221501
20. https://www.euractiv.ro/facts-not-fake
21. https://eldorado.tu-dortmund.de/handle/2003/40586
22. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2023-06/Digital_News_Report_2023.pdf