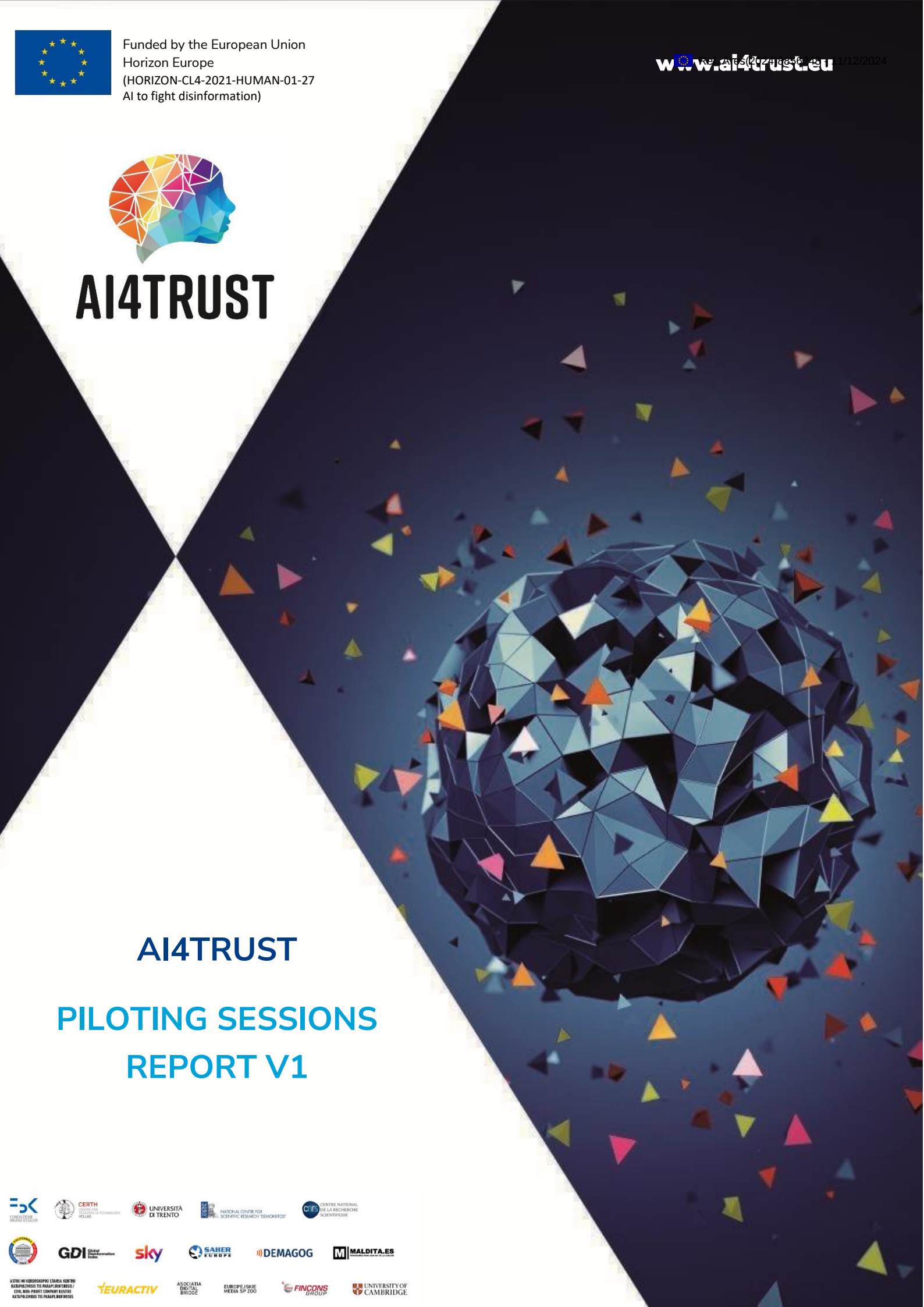
Funded by the European Union
Horizon Europe
(HORIZON-CL4-2021-HUMAN-01-27
AI to fight disinformation)

www.ai4trust.eu

AI4TRUST



AI4TRUST

PILOTING SESSIONS REPORT V1

Project acronym:	AI4TRUST
Project full title:	AI-based-technologies for trustworthy solutions against disinformation
Grant info:	Project ID: 101070190
Deliverable name:	D6.2 - Piloting Sessions Report v1
Dissemination level:	Public
Version:	v5.0
Status:	Final Version
Due date of deliverable:	30/11/2024
Actual submission date:	30/11/2024
Work Package:	WP6
Lead partner for this deliverable:	NCSR-D
Partner(s) contributing:	Ellinika Hoaxes, Demagog, EMS, SKYTG24, ADB, Maldita, EURACTIV
Main author(s):	Athanasios Davvetas, NCSR-D Dora Katsamori, NCSR-D
Contributor(s):	Yannis Delimaris, Ellinika Hoaxes Kinga Klich, Demagog Julia Ablewska, EMS Andrea Dambrosio, SKYTG24 Nicola Bruno, SKYTG24 Manuela Preoteasa, ADB Claudia Ocaranza Abascal, Maldita Judith Baars, EURACTIV

Evaluator(s):	Camille Roth, CNRS Serena Bressan, FBK Danilo Giampiccolo, FBK

Summary of modifications

V.	DATE	AUTHOR(S)	SUMMARY OF MAIN CHANGES
1	22/10/2024	Athanasios Davvetas, NCSR-D Dora Katsamori, NCSR-D	Conclusion of all AI4TRUST piloting sessions, collection and analysis of qualitative and quantitative evaluations carried out based on feedback from end-users
2	26/11/2024	Athanasios Davvetas, NCSR-D Dora Katsamori, NCSR-D	Finalisation of the first version of the deliverable for review
3	27/11/2024	Camille Roth, CNRS Serena Bressan, FBK Danilo Giampiccolo, FBK	Review of the deliverable by the QA Leader and the Coordinator
4	29/11/2024	Athanasios Davvetas, NCSR-D Dora Katsamori, NCSR-D	Implementation of review requests and finalisation of the deliverable
5	30/11/2024	Riccardo Gallotti, FBK Serena Bressan, FBK Danilo Giampiccolo, FBK	Final review and formatting of the deliverable by the Coordinator for submission on the European Commission Participant Portal

Table of contents

List of abbreviations

Abbreviation	Meaning
DPO	Data Protection Officer
NCSR-D	National Centre for Scientific Research “Demokritos”
WP	Work Package
SoMe	Social Media
UI	User Interface

Executive summary

This report constitutes Deliverable D6.2 - *"Piloting Sessions Report v1"* of Work Package (WP) 6, *"Piloting, Assessment & Fact-checking"*, led by NCSR-D, within the European project AI4TRUST: *"AI-based Technologies for Trustworthy Solutions Against Disinformation"*. It includes an update on the planning and methodology developed in Task (T) 6.1, *"Piloting Requirements, Coordination, and KPIs Definition"*, and presents the results of the first piloting phase of the AI4TRUST Platform v1 (D5.5), conducted as part of T6.3, *"Pilot Deployment and Operation"*. This phase took place in September and October 2024, involving all consortium end-users, including fact-checkers and media professionals.

This document first provides an essential recap of the methodology as well as the indicators and criteria specified in D6.1. The report then provides an overview of the evaluation process carried out, concludes with a summary of both quantitative and qualitative results, and includes a review of the monitoring process for the pilot workshops.

As already mentioned in D6.1, 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 and implementation 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 pilot requirements for the AI4TRUST Platform, was thus integral to our project's success.

With respect to the state-of-the-art analysis described in the D6.1, which provides a thorough understanding of existing technologies, methodologies, and best practices in the communities covered by AI4TRUST, as well as the AI4TRUST Platform Quality Model, we proceeded to the design of the pilot workshop format. The aim of the pilot workshops was to evaluate both the existing AI4TRUST Platform and its tools and try to demonstrate the value of the application of these technologies in the field of mis/disinformation by providing opportunities and meeting needs that arise in the process of assessing the reliability of news in the context of the fact checkers and media professionals' daily workflow. The pilot workshops were addressed to various end-users and organised with the support of the AI4TRUST partners. An overview of the qualitative and quantitative evaluation results is provided, alongside a more detailed analysis.

In conclusion, participants expressed a very positive attitude towards using these tools and the whole platform, emphasising the value they can add to their daily workflows. Their feedback has been invaluable in identifying ways to enhance the tools' functionality and performance.

1. Introduction

This section offers background information on the pilot workshops and their corresponding report. It outlines the purpose and scope of the report, along with the design choices and structure of the pilot workshops. Additionally, it details the objectives of the workshops while also providing a broader contextual perspective.

1.1. Purpose and Scope

The goal of this deliverable is to illustrate the evaluation of the pilot workshops performed within the AI4TRUST project. This document is part of WP6 – *"Piloting, Assessment & Fact-checking*) and in particular T6.3 – *"Pilot deployment and operation"*. In the following, we present the overview of the evaluation process as planned in D6.1 – "Pilot Planning Report" and related indicators and criteria. This deliverable is released on month (M) 23 of the project and reports on activities performed during the first evaluation round of the AI4TRUST Platform v1 (D5.5) by the end-users of the project, i.e., fact checkers and media professionals. The detailed piloting requirements, planning, and KPI definitions were outlined in D6.1 – *"Pilot Planning Report"*.

A continuous evaluation process was implemented to assess the AI4TRUST Platform, adopting a co-design approach in collaboration with the consortium's end-users, who are anticipated to become the future adopters of the developed technological solutions. An iterative process of co-design and evaluation offers significant benefits when developing a technological platform. By conducting multiple piloting phases and gathering feedback from end-users, the platform can be continuously refined to better meet their needs. This approach ensures that the design remains user-centred, improves functionality based on real-world use, and fosters greater engagement and trust among participants, ultimately leading to a more effective and impactful solution.

For these reasons, it was decided to evaluate the AI4TRUST Platform in two rounds for potential users to actively participate in the co-design of the AI4TRUST tools with the main purpose of their effective use in their daily work and the facilitation of their workflow.

1.2. Pilot Specifications

This section delves into the essential components that lay the foundation for a comprehensive and successful pilot design. The design takes into account and builds on the current fact-checking state-of-the-art and desirable criteria and features that were gathered within the previous D6.1.

1.2.1. Workshop Design Approach and Methodology

This deliverable provides an evaluation of the project results according to methodologies and criteria defined in D6.1. The evaluation is structured from two perspectives:

- a *macroscopic perspective* of the AI4TRUST Platform that focuses on the platform as a whole;
- a *microscopic perspective* of the AI4TRUST Platform that focuses on the validation tools offered by the platform.

NCSR-D, the leading organisation of WP6 and leader of Task 6.3 — *"Pilot deployment and operation"*, is responsible for the organisation and coordination of the first and the subsequent evaluation rounds within the AI4TRUST project. To this end, they adopted the concept of specialised workshops for the evaluation round, accommodating multiple participants. This approach aims to actively involve all participants through plenary discussions and collaboration within the group, as well as to better support them during their interactions with the tools. Therefore, the workshop is interactive by design based on the principles of adult learning, where participants need to share their thoughts and experiences on issues that concern them, and thus being the main actors. Additionally, one of the goals of the pilot workshops is the co-design process where participants, as the potential users of the AI4TRUST platform, have a crucial role in the decision-making process. Therefore, discussions are often driven by them.

Besides the two main goals mentioned before, gathering participants within the planned workshop is a good opportunity to engage their vision of future improvements as part of the co-design process. To do so, the workshop aims to better understand the potential AI4TRUST platform's users' needs and expectations. It also gives the opportunity to media professionals to be involved in the incremental process of developing a platform that targets their needs, as well as, to interact with novel tools and approaches to their daily workflows.

The workshop was designed with the objective to attract and engage media professionals. In compliance with the AI4 TRUST Grant Agreement (GA), the workshop targets the following stakeholders: fact-checkers, journalists (or media professionals in more general terms), policymakers, and researchers. The first piloting phase, the subject of this deliverable, involved fact checkers and journalists. A detailed presentation of the involved organisations takes place in Section 2.

Previous interactions with fact-checkers and journalists within the project's consortium had highlighted that these two participant/user groups have vastly different needs and interests. These interactions took place during the various planned project meetings, e.g., plenary consortium, WP6 and bilateral meetings as well as during the planned preparatory workshop. The above interactions suggest that the goal of fact-checkers is to debunk a false claim, and thus are interested in all activities, tools and approaches that can aid their pursuit of debunking claims. As a result, they are most of the time more experienced with the use of digital tools and often critical in the technical aspects of proposed methods such as user navigation, interaction, experience, and interface. Due to already established workflows, they are more sceptical of new solutions and approaches to dealing with mis/disinformation.

On the other hand, dealing with disinformation or misinformation is not the main goal of journalists but is a more of a side effect within the modern dissemination of news and rapid exchange of information. As journalists prepare informative articles that cover news and stories worldwide, they often deal with emerging misinformative content and narratives that have to be carefully circumvented in order to provide the best quality of coverage. In these days and age characterised by an abundance of information and a rapid dissemination of content through social media platforms, keeping up with traditional methods of journalism is not scalable. Therefore, journalists have started to use more and more digital tools to deal with disinformation in order to reduce effort in understanding the narratives and focus more on preparation of articles and news coverage. Thus, participants/user groups are less critical of the features of relevant digital tools and approaches and more prone to exploring and integrating digital solutions.

Despite the design and implementation of the workshops, several other relevant activities had to take place preliminary to the workshop, during and after their implementation. In coordination with the original piloting plan that was laid out within D6.1 — *"Pilot Planning Report"*, the following activities were prioritised before implementing the pilot workshops:

- Definition of a preliminary set of KPIs for the pilot testing;
- Setting up communication channels by establishing points of contact;
- Preparation of relevant communication material for the participants (i.e., invitation letter and consent form);
- Set up the testing environment by ensuring the participants access to the AI4TRUST Platform (provide credentials);
- Coordination with the partners for the workshops preparation;
- Creation and population of the evaluation corpus (an Excel sheet with samples of news items in all supported modalities, i.e., video, audio, text, and image, to serve as input during testing of the Platform within the workshops);
- Organisation of training workshops for participants and points of contact to get familiar with the AI4TRUST Platform and the structure of the pilot workshop (referred to as "preparatory workshop").

Table 1 portrays the above actions sorted by the date that they completed.

Completed at: (Project Month/Date)	Phase	Activity
M8/Aug. 23 Updated: M17/May 24	Preparatory	Definition of a preliminary set of KPIs for the pilot testing
M10/Oct. 23	Preparatory	Setting up communication channels by establishing points of contact

M19/July 24	Preparatory	Set up the testing environment by ensuring the participants access to the AI4TRUST Platform (e.g., providing credentials)
M19/July 24	Preparatory	Organisation of training workshop for participants and points of contact to get familiar with the Platform and the structure of the pilot workshop (referred to as “preparatory workshop”)
M19/July 24	Preparatory	Set up the testing environment by ensuring the participants access to the AI4TRUST Platform (provide credentials)
M21/Sep. 24	Preparatory	Preparation of relevant material for the pilot workshops (including evaluation corpus)
M22/Oct. 24	Execution	Execution of piloting workshops
M22/Oct. 24	Execution	Feedback collection through survey
M23/Nov. 24	Execution	KPI assessment through data analysis
M23/Nov. 24	Post-pilot	Preparation of pilot summary report

Table 1: Overview of the timeline that took place within the first evaluation round

The definition of the pilot testing KPIs took place at M8, with the initial version and planning report being laid out within D6.1. After feedback received from the project's first review with the external experts, the KPIs along with the pilot planning were updated at M19. The KPIs involve both qualitative and quantitative indicators. The establishment of pilot points of contact refers to the activity of assigning a representative from each organisation which facilitate and aid with communication, dissemination of relevant content, and internal organisation of the participants. This activity was concluded at M10.

Within M19, the following activities were completed: organisation and implementation of preparatory workshop, setting up testing environment, and preparation of relevant material. The preparatory workshop aimed to familiarise the participants and respective points of contact with the Platform and the process/activities that would take place within the pilot workshops. After completion of the preparatory workshop, relevant material and activities were sent to the consortium members to revise, adjust and optimise. The relevant material was packaged within an email, and it contained the recorded preparatory workshop session, the Platform, and toolbox

questionnaires, along with an Excel file for points of contact to fill in their availability for the pilot workshop. At the same time, NCSR-D along with its DPO internally prepared the consent form, which was also cross-checked along with SAHER, the leading partner in legal issues within the consortium (T1.4 — *"Data Management, Ethical, Legal and Security Compliance Framework"*).

An incremental approach for the preparation of relevant material was adopted. An initial version of the relevant material was prepared during M13 and M14, which was later updated during M19 and finalised at M21. This approach was essential in order to adjust the content to the incremental development of the AI4TRUST Platform. In collaboration with the AI4TRUST leaders of WP3 — *"AI-driven Data Analysis Methods"* (CERTH), responsible for the participatory development of the project's technological tools, and WP5 — *"Technical Implementation of the Platform & Security Framework"* (FINC), also acting as the system integrator, Platform access credentials were distributed to the end-users.

After discussion within the above-mentioned preparatory workshop, it was decided by the Work Package leaders of WP5 (FINC) and WP6 (NCSR-D) to take advantage of the early access to facilitate any zero-day breaking bugs for the seamless implementation of the pilot workshops. In addition to granting access to the Platform, an Excel sheet was distributed to serve to report any critical bugs, which could then be patched prior to the piloting workshops.

Furthermore, within M18, the process of gathering input data from the daily workflow of all participating organisations started. The produced resource called "Evaluation Corpus" contains misinformative content with the following structure: topic, content type, language, description, mirror link, original source link, a flag depicting whether the content is verified by other means, an open field describing what tools were used to verify said content, verdict/rating and notes. The Guidelines provided for the creation of the evaluation corpus for each participating organisations were the following:

- 10 samples for each module (audio, text, image, video) per participating organisation:
    - Organisations that works with 1 language (e.g., Italian): 20% English (2/10 text samples) - 80% Italian (8/10 text samples);
    - Organisations that work with 2 languages (e.g., German, French): 20% English (2/10 text samples) - 40% German (4/10 text samples) - 40% French (4/10 text samples).
- The use of widely known databases is discouraged as they may have been involved in the training of the tools.
- If organisations already gather data as such automatically, gathering once is acceptable, there is no need to update the evaluation corpus constantly.

The evaluation corpus was finalised during M21. To ensure the seamless operation of all the data samples within the pilot workshops an English version with backup data samples was created by

NCSR-D. The rest of the activities described within Table 1 can be found within the relevant sections of the deliverable.

1.2.2. Pilot Workshop Objectives

To design the objectives of the workshop, an alignment with all the tasks and broader objectives of WP6 had to be performed. T6.1 has the objective to plan the testing and to implement the pilot activities in order to validate the AI4TRUST Platform in different settings, namely different user groups, languages and distinct scenarios of use (i.e., journalism, fact-checking, policy making processes, research activities, etc). At the same time, the following activities are also of interest to WP6: (i) engaging with community of media practitioners, fact-checkers, policy makers; (ii) assessing the current efforts and methods adopted by fact-checkers and journalists to tackle disinformation and misinformation; and, (iii) testing and validating the AI4TRUST Platform with a community of end-users across Europe. Besides the above, the design of the objectives had to additionally ensure consistency with the methodology and sociotechnical requirements (WP2), and provide input to WP3-WP4-WP5 for the platform development.

In alignment with the above, the pilot workshops objectives, as communicated to the participants as part of their call for participation, were to explore their views regarding:

a. their expectations when visiting an online platform focusing on assessing the veracity of information, including effective validation activities, as part of their everyday workflow;
b. their evaluation after their engagement and interaction with the AI4TRUST platform, including the AI4TRUST tools, using real samples from their everyday workflow.

Additionally, part of the evaluation process was the specification of the piloting configurations to meet the partners' conditions and desirable criteria for testing AI4TRUST tools and their impact in their daily workflow. Thus, the goal of the pilot workshops except for the evaluation of the validation was also to better understand the potential AI4TRUST platform users' needs and expectations.

1.2.3. AI4TRUST Quality Model

Within Section 5 of D6.1, there is a description of the evaluation aspects of the AI4TRUST Platform Quality Model, including the KPIs appropriate for the success of the platform. The quality of a platform is determined by its ability to meet the explicit and implicit needs of its target stakeholders. By adapting the ISO/IEC 25010 model[1], these requirements can be categorised into broad groups known as characteristics, which are further divided into sub-characteristics. This structured approach ensures that all stakeholder needs are comprehensively addressed.

Considering the unique aspects of the AI4TRUST platform, including its components, stakeholders, and potential use cases, we have identified a set of relevant characteristics and sub-characteristics. These constitute the AI4TRUST Platform Quality Model, which has been used to evaluate the platform's quality throughout its piloting phases. The model is based on the ISO/IEC 25010

[1] https://www.iso.org/standard/78176.html & https://www.iso.org/standard/78175.html

standards, tailored to the specific functionalities and use cases of the AI4TRUST platform. The AI4TRUST Platform Quality Model is illustrated in the following Figure 1. We have enhanced certain characteristics with additional evaluation aspects to better align with the project's objectives.

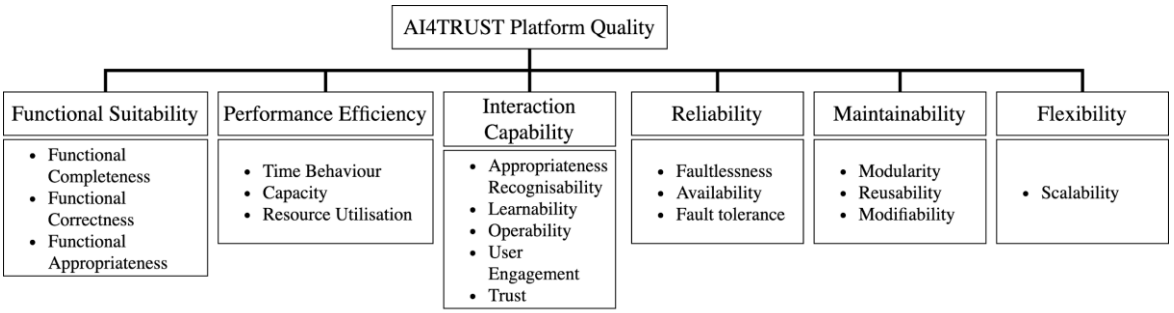


Figure 1: AI4TRUST Platform Quality Model

2. Workshop Design and Implementation

This section provides an overview of the pilot workshops that took place during the first evaluation round of the AI4TRUST platform. In addition to the workshops' structure, other related tasks, such as data collection methods and discussion-oriented approaches, key themes and topics covered. Among other topics, the workshop explores key themes such as data privacy and security, AI transparency and explainability, user-centred AI design, the impact of AI on society and the workforce, and trust in AI systems. Additionally, descriptions of the participant groups and their backgrounds are provided.

2.1. Workshop Description

Media professionals are the target group of the piloting sessions, including but not limited to fact-checkers and journalists. As most media work tasks are conducted by these two groups, they are particularly well-represented in the pilot workshops. Despite, including both groups allows a level of diversity, as the daily workflows of each group are vastly different from each other.

The workshops were designed with the aim of physical implementation, along with a respective moderator/facilitator from the leading organisation of WP6 (NCSR-D). NCSR-D and the respective participant organisations conducted most of the workshops in physical form (~70%, 5 out of 7 relevant organisations participated physically). Despite the best efforts of all partners, for reasons of force majeure, two of these workshops had to be conducted online. For this purpose a virtual meeting (in Zoom) was set up in advance, with all the participants, including the NCSR-D facilitators being online. The meetings were recorded and relevant material were digitised (e.g., digital writing board in the form of shareable document).

Within the summer of 2024, a preparatory workshop took place to inform the partners on the activities that will take place during the workshop and a smaller scale simulation of the workshop

took place. This allowed the pilot leaders to prepare the participants on the activities which they would participate in.

Each workshop lasted approximately three hours, which provided ample time to: (i) express their needs and expectations regarding the use of an information veracity validation platform, as well as interact with each other, and (ii) interact with the AI4TRUST Platform with a focus on the four different validation tools, each specialising in a different modality. The workshop schedule included a break that allowed time for the users to focus and reflect on different aspects of the AI4TRUST Platform, i.e., the platform as a whole and each individual group of tools within its toolbox. At the same time, this split the workshop into distinct phases. Figure 2 presents an overview of the pilot workshop in the form of flowchart.

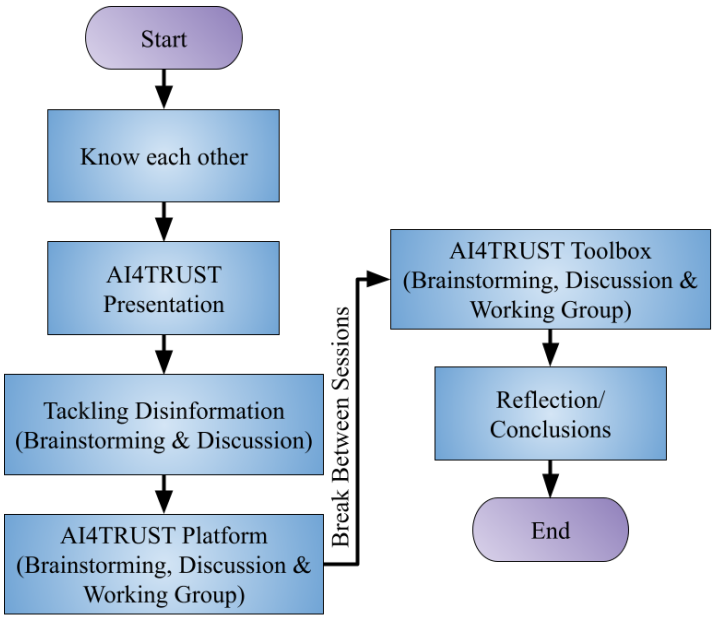


Figure 2: Pilot workshop sessions in the form of flowchart

Table 2 displays the various sessions that make up the structure of the workshop in more detail along with the duration, description, and tools/techniques used.

Activity	Duration	Description	Tools/Techniques
Know each other	15’	Participants introduced themselves and their work/involvement in the project	1. Discussion
Short Presentation	10’	Introduction to the AI4TRUST project and the Platform and presentation of the objectives of the pilot workshop/data collection process	1. Presentation

Needs and Expectations	15’	Discussion with media professionals: the aim of the session was to explore (i) the role of fact-checkers and journalists, (ii) their daily workflows, (iii) the challenges that they face in dealing with disinformation, and (iv) how digital tools can support their effort	1. Brainstorming 2. Discussion 3. Needs/ Expectations Board
AI4TRUST Platform	45’	Introduction of the AI4TRUST Platform through a presentation followed by an interactive evaluation segment. Participants provided feedback and discussed key criteria for platform assessment	1. Discussion 2. Brainstorming 3. Working Group 4. Questionnaire
Short Break (15’)			
AI4TRUST Toolbox	45’	Introduction and presentation to the AI4TRUST toolbox, interaction with the various tools using data samples from their everyday workflow. The session concluded with dedicated evaluation questionnaires	1. Discussion 2. Brainstorming 3. Working Group 4. Questionnaires
Reflection/ Conclusions	15’	Collaborative session to shape future platform release. Participants were encouraged to share what functionality is missing, type of tools or services needed, and how the platform can contribute to the project’s and end-users’ goals	1. Discussion

Table 2: Overview of the pilot workshop design

The workshop started with some initial discussion that revolved around the background of the participants. The moderators/facilitators of the workshop also introduced themselves at this stage. The duration of this session usually took around 15 minutes to complete. A short presentation of the project followed the “Know each other” session. During this part, an *ad hoc* presentation was utilised to introduce the AI4TRUST project to the participants. Furthermore, the objectives of the workshop and the data collection process were introduced. Once the participants were provided with the context of the project, the workshop, and how their feedback would be handled, the process of gathering needs and expectations began.

To encourage and to elicit the needs and expectations of users that interact with an information veracity validating platform, a brainstorming session followed. After this initial brainstorming and discussion, participants were asked to write their needs and expectations in a Needs/Expectations

board. In cases of physical workshops, this took place using a large piece of paper and post it notes, while for online meetings an online collaborative document was created (Google Document).

After completion of the above session, time to interact with the AI4TRUST Platform was allocated. Firstly, the platform was presented to the participants through screen sharing to kickstart the discussion and brainstorming to initialise the critical thinking of the participants. After the brainstorming, participants were encouraged to interact with the platform. Lastly, a questionnaire focused on the evaluation of the platform was shared to be filled by the participants. A short break followed this session and marked the end of the first half of the workshop.

After the short break, the second half of the workshop kick started with the evaluation of the AI4TRUST toolbox. Initially, all the tools developed within WP3 and later integrated into the platform in WP5 were presented to the participants, using an example for each modality to demonstrate the tools' functionality and provide a brief tutorial to help them understand the required inputs. Once the functionality was presented, a short brainstorming and discussion session was held. Participants were then encouraged to interact with the tools using the evaluation corpus they had prepared prior to the pilot workshop. Subsequently, they were asked to fill in a questionnaire aimed at evaluating the toolbox. Finally, a short discussion and reflection session took place to review the evaluation of the platform and the toolbox, providing participants with the opportunity to express their final remarks.

2.2. Evaluation Process Design

Section 5.1 of D6.1 includes the preliminary evaluation scenarios that were initially used for observation-based testing and were subsequently adjusted for use during the first phase of the evaluation process. The evaluation approach included the collection of quantitative and qualitative data through brainstorming and discussion processes and the use of online questionnaires. A more detailed description follows below. This evaluation approach, including the tools used, will remain unchanged throughout the planned subsequent evaluation rounds to ensure consistency of the results. Furthermore, the evaluation approach was connected with the desirable criteria in an attempt to connect the required needs of the stakeholders with the functionalities of the platform at a higher level.

As already mentioned in the previous section, before the workshops NCSR-D, as the piloting evaluation coordinator, sent to the points of contact of every organisation participating in the first evaluation round an email including information and preparatory steps. Among others, these emails included a consent form to be disseminated and filled in by the workshop participants, which included information about the:

- Research activity organiser
- Point of contact regarding personal data and participant rights
- Goals of the workshop
- Participation terms

- Third parties processing the data
- Data storage

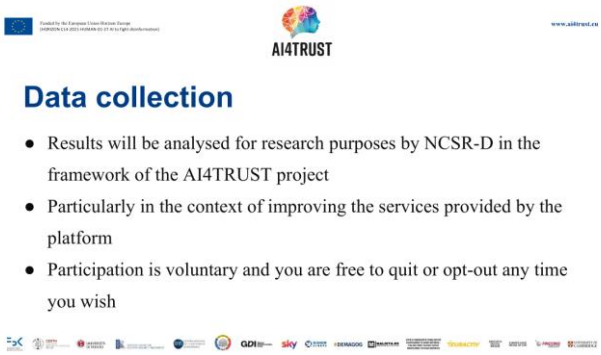


Figure 3: Data collection slide included in the complementary presentation to allow for discussion and presentation of the data collection terms

The template of the consent form has been attached in Appendix I. At the same time, during the workshop, reference was made again to the participants about the voluntary nature of their participation, as well as about the procedure that would be followed for the collection, processing and storage of the evaluation data. The relevant slide is shown in Figure 3.

2.3. Data Collection

During the workshop design, NCSR-D provisioned the gathering of both qualitative and quantitative feedback from the workshop's participants. The quantitative feedback was gathered mostly with the use of relevant questionnaires, i.e.,: *(i) platform evaluation (PEQ)*, *(ii) toolbox evaluation (TEQ) and an optional (iii) workshop evaluation (WEQ).* NCSR-D produced the final version as a Google survey to facilitate easy participation and response collection, following the dissemination and gathering of feedback from the project consortium and relevant partners.

2.3.1. Platform Evaluation (PEQ)

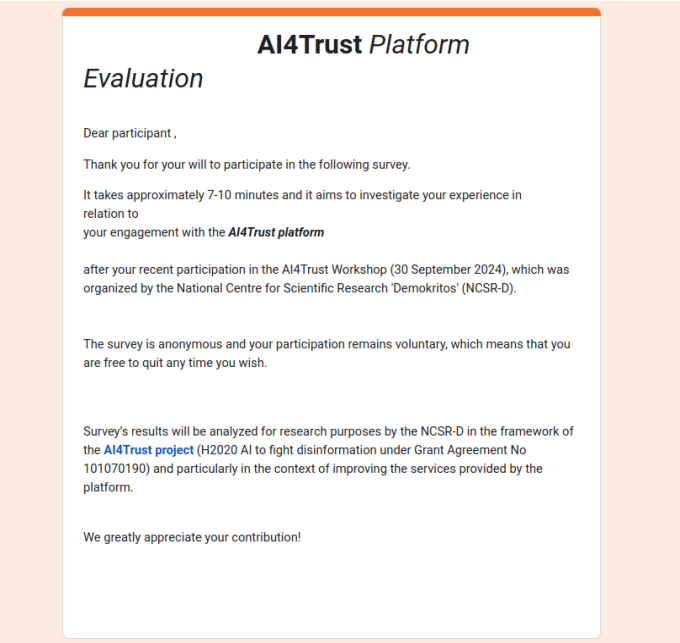
The scope of the Platform Evaluation Questionnaire (PEQ) was to gather feedback regarding the platform as a whole including the *navigation*, *interaction*, *and usability*. To be consistent with the relevant literature and to set a comparative set of questions, PEQ included questions from the relevant System Usability Scale (SUS)[2] and E*ffectiveness*, *Efficiency*, *Satisfaction and Learnability* (EESL)[3] questionnaires. Figure 4 displays the starting page of PEQ as a Google Form. The main sections of PEQ are:

- Generic/Demographics Info
- Expectations/Experience
- Usability of the Platform (based on SUS)

[2] J. Brooke, "SUS: A Retrospective" *Journal of Usability Studies*, vol. 8, no. 2, pp. 29–40, 2013.
[3] 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

- Platform Functionality
- User's Satisfaction (based on EESL)

Participants were first asked to visit the AI4TRUST platform, before signing in to the toolbox, explaining that the homepage was under construction. After a discussion about their expectations and needs when visiting such a platform, they were asked to fill in the online questionnaire. Their thoughts and suggestions during the discussion were recorded by the facilitator and will be presented in Section 4 of this document as well as the results of the questionnaire.


AI4Trust *Platform*

Evaluation

Dear participant ,

Thank you for your will to participate in the following survey.

It takes approximately 7-10 minutes and it aims to investigate your experience in relation to
your engagement with the ***AI4Trust platform***

after your recent participation in the AI4Trust Workshop (30 September 2024), which was organized by the National Centre for Scientific Research 'Demokritos' (NCSR-D).

The survey is anonymous and your participation remains voluntary, which means that you are free to quit any time you wish.

Survey's results will be analyzed for research purposes by the NCSR-D in the framework of the **AI4Trust project** (H2020 AI to fight disinformation under Grant Agreement No 101070190) and particularly in the context of improving the services provided by the platform.

We greatly appreciate your contribution!

Figure 4: Google Form Starting Page of PEQ

2.3.2. Toolbox Evaluation (TEQ)

Toolbox Evaluation Questionnaire (TEQ) follows the same structure as PEQ. The major difference between the two questionnaires is that TEQ is repeated for each modality of tools (i.e., image/video, audio and text). At the same time, TEQ includes additional questions regarding the performance of the tools, such as reporting on the total amount of data samples tested and whether the tool produced a false prediction.

Upon entering the toolbox, participants were given a brief presentation of the tools developed within WP3 and their functionalities. Afterwards, participants were invited to interact with them using real examples from their daily workflow, which the points of contact had either collected beforehand in the evaluation corpus excel or brought in on the day. Upon completion of their tasks, and after they had tried each tool at least once to get an idea of their functionalities, they were asked to fill in the relevant questionnaire (TEQ).

2.3.3. Workshop Evaluation (WEQ)

After the conclusion of each workshop, NCSR-D facilitators sent a relevant “Thank You” email to the points of contact that included the Workshop Evaluation Questionnaire (WEQ) to be disseminated to the workshop participants. WEQ includes the following sections:

- Generic/Demographics Info
- Workshop's Evaluation (structure, duration, content)
- Expectations/Experience

The aim of WEQ was to get the feedback of participants to adjust or change any aspects of the workshop in view of the following rounds of evaluation. All three questionnaires predominantly used the Likert scale for most of the questions. They are all included in Appendix II.

2.3.4. Moderator Report

Questionnaires are an agile and easy to deploy form of gathering quantitative feedback. However, one key aspect of performing evaluation through pilot workshops is to receive qualitative feedback. Qualitative feedback through discussion enables the prioritisation of tasks during development, while also facilitating the gathering of suggestions, needs, and expectations.

The Moderator Report is a structured document created for on-site note-taking by the workshop facilitator. During the first evaluation round, representatives from the WP leader organisation, NCSR-D, facilitated the workshop’s workflow and materials. The objective of the Moderator Report was to act as an overview of each individual workshop by archiving useful suggestions, testing observations, issues identified, overall notes, and other general notes that may have arisen during the various discussions that took place throughout the pilot workshop sessions. The Moderator Report includes the following sections:

- Session details (e.g., date/time, moderator name(s), No. of participants)
- Preparatory notes (e.g., checks on test data and evaluation scenarios, list of identified issues)
- Testing session observations (e.g., evaluation scenarios and tasks, feedback, technical issues, lead time, suggestions)
- Overall impressions (e.g., strengths, weaknesses, potential improvements)
- General notes

A template of this document, which was adapted for each pilot workshop moderator's report, is attached in Appendix III.

2.3.5. Needs and Expectations Map

The design of the pilot workshop included the extraction of needs and expectations from the participants. To facilitate this process, during the first part of the workshop NCSR-D facilitators designed a session where participants were encouraged to express their needs and expectations regarding the potential use of the AI4TRUST Platform and its tools as part of their everyday

workflow. Their thoughts were recorded both by themselves using post it notes or by the facilitators during the brainstorming session. This approach yielded the following advantages:

- Structured archiving of feedback;
- Engaging participants and active participation;
- Provoking the process of critical thinking and reflection.

Activities that encourage participants to express their views and experience add value to the evaluation process as they increase participants' engagement and encourage a sense of co-creation and co-design. Often, participants were also encouraged to participate in smaller discussion groups in order to support all of them to express themselves - often in their native language - and make suggestions, allowing at the same time a seamless and enjoyable extraction of feedback.

Lastly, dedicated discussion and hands-on sessions such as the needs and expectations mapping board provoked critical thinking towards the current state of platform and tools. Participants, in turn, began reflecting on the current state and expressed suggestions that would meet their expectations or highlight needs that were currently unmet.

2.4. Workshop Schedule & Activities

Although the intention of the NCSR-D facilitators was that all the pilot workshops would take place on site, taking into account the limitations set by the participant organisations, we finally managed to schedule 4 of the 6 workshops on site and the other 2 online, while maintaining the same format and encouraging the active participation and involvement of the participants, also thanks to the adaptation of methods and materials to the online or offline form.

After all the participating organisations had communicated their availability for September and October, the NCSR-D team internally created an initial pilot workshop schedule including all organisations. Emails to all points of contact were sent to inquire on their availability for the selected choice of dates. After a positive response from the respective organisation, a bilateral meeting between NCSR-D and the respective organisation took place to clarify the workshop details such as location, time, number of participants, and to clear out any additional questions. Following the bilateral coordination meeting, an email was sent to the points of contact containing details about the workshop, including its objectives, duration, and a brief agenda. The email also served as a reminder to finalise the evaluation corpus and provide any feedback regarding the questionnaires. The above process led to the finalisation of the pilot workshop schedule.

Some days before the scheduled days of pilot workshops, the points of contact were informed to disseminate the consent forms to all participants in the session and to ask them to complete and sign these documents, before the participation in the workshop. Finally, after the successful completion of each workshop, an email was sent containing the workshop evaluation questionnaire to assess the workshop's structure, duration, content, and overall success, as well as to gather additional feedback to refine the workshop process for future evaluation rounds.

In the following table, the exact timeline of each workshop is displayed, including the name of the participating pilot institution, the date and time, and the location where the sessions took place:

Participant Org.	Date/Time	Location
Ellinika Hoaxes	30 Sep. 2024, 09:30 CET	Ellinika Hoaxes Premises (Thessaloniki)
Demagog & EMS	03 Oct. 2024, 11:00 CET	Representation of the European Commission in Poland (Warsaw)
SKYTG24	09 Oct. 2024, 10:00 CET	Online (Zoom Meeting Room)
ADB	11 Oct. 2024, 10:00 CET	Representation of the European Commission in Bucharest
Maldita	14 Oct. 2024, 12:30 CET	Online (Zoom Meeting Room)
EURACTIV	16 Oct. 2024, 14:00 CET	EURACTIV Premises (Brussels)

Table 3: 1st evaluation round timeline

The workshops were mainly conducted between the end of September and October 2024, and for the vast majority in person. Initially, NCSR-D facilitators suggested the organisation of the workshops within the premises of each organisation. The idea behind that is to enable the seamless participation of the relevant colleagues as well as to avoid the need of further resources.

Due to the decision to combine two partners, Demagog and EMS, both based in Poland, into a single workshop, and because ADB ended up engaging with more stakeholders than initially planned, an alternative solution was required. To meet these new needs, these organisations took the initiative to organise their workshops at the European Commission's representation offices in their respective cities. This arrangement allowed the partners to avoid using additional resources while ensuring that suitable rooms with the necessary equipment were available for the smooth operation of the workshops.

2.5. Participant Demographics

Out of the seven organisations participating in the first piloting evaluation round of AI4TRUST, four – namely SKYTG24, EMS, ADB, and EURACTIV – primarily employ journalists, and their focus is on combating disinformation within the journalistic process. The remaining three organisations – Maldita, Demagog, and Ellinika Hoaxes – predominantly employ fact-checkers, and their interest

lies in using the tools and platform to support and enhance the fact-checking process. The above organisations cover a broad spectrum of languages, such as Spanish, Polish, Greek, Italian, Romanian, French, and German. Below is a brief introduction to each organisation:

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.

The *Demagog Association* is a Polish non-profit organisation that fights disinformation through fact-checking, debunking of fake news, media literacy and education, technology development, research, and policy action.

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.

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 NBC Universal.

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 38 expert opinions and commentary. Striving to build resilience against disinformation, EMS is also extensively involved in fact-checking activities.

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).

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.

The following figures and subsections portray the different participant distribution in terms of gender, age group, role, and experience in that role. We based the figures on the total of 44 responses within the relevant questionnaire.

2.5.1. Gender & Age Group

49 media professionals in total participated in the pilot workshops and 44 of them evaluated the AI4TRUST Platform and tools. Figure 5 presents the gender distribution of the participants. For the purposes of quick identification, we visualise the results with a pie chart. The figure suggests that the pilot workshops were able to achieve a near-even gender distribution. In particular, 59% (N=26)

of participants were male, while 41% (N=18) were female. This near-even split allows for a fairly balanced and diverse evaluation that takes into consideration any potential biases or gender-specific trends.

Figure 6 displays the age group distribution of the participants. We utilise a pie chart for the purposes of visualisation. The participants of the pilot workshops seem to be fairly balanced within the various age groups. Specifically, 18.2% of the participants are younger than 25, 36.4% of the participants are between the ages of 26 and 30, 15.9% of the participants are between the ages of 31 and 40, while the rest (29.5%) are older than 41. The above percentages display a rather balanced split between age groups.

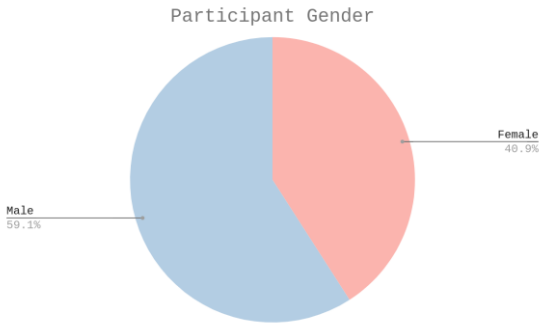


Figure 5: Participant Gender Distribution

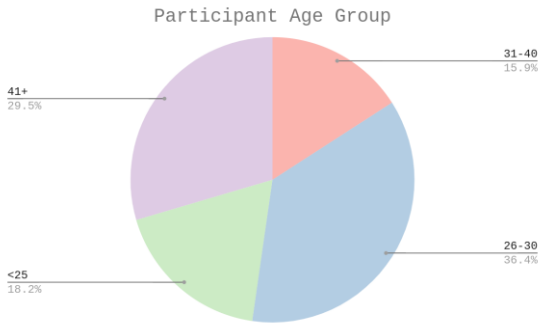


Figure 6: Participant age groups

2.5.2. Roles and Experience

Figure 7 visualises the role distribution of the participants using a bar chart to quickly identify large groups. 34% of the participants were journalists (15/44), while 29.5% were fact-checkers (13/44).

Figure 8 depicts the experience distribution of the participants in each of the aforementioned roles. As discussed during the previous section, within the age group of participants, a clear trend in the workshop participant emerges. The vast majority of the participants (63.6%) have less than 10 years of experience in their particular role, while the next big category is those with over 20 years of experience (20.5%), leaving the rest of participants with 10-20 years of experience at 15.9%.

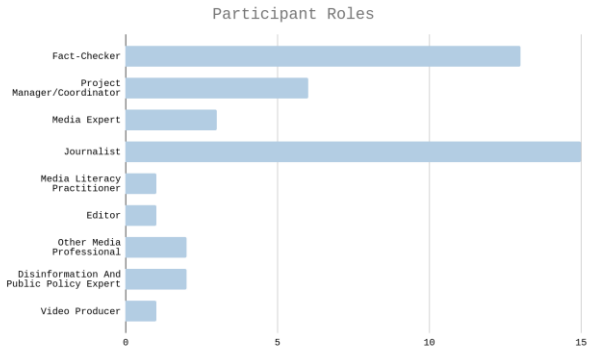


Figure 7: Participant roles

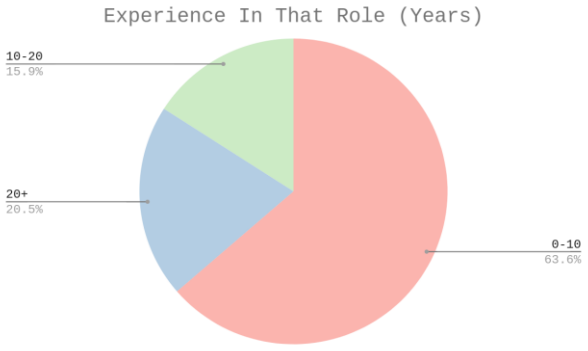


Figure 8: Participant Experience

3. Evaluation Results

As already mentioned, the evaluation's methodology aims to gather insights on the *usability, functionality, user experience and attitudes of users* with regards to both the AI4TRUST Platform itself and its functions (v1 - D5.5), including the validation of its tools (v1 - D3.1). In the following sections we will present them and in Section 4 we will make some correlations and interpretations.

In this part of the workshop, insights were collected from two different phases; (i) first, after some warm-up questions we had a discussion regarding the potential needs and expectations of the users; and then (ii) a questionnaire was proposed regarding the specific enablers and barriers that participants had encountered while interacting with the AI4TRUST Platform and tools. Specifically, different variables were evaluated (e.g., user friendliness of the visualisation, efficiency, accessibility, value of the functionality and output, attractiveness of the interface, language usage, task flow, and trustworthiness of the tool), where the specific enablers and barriers were examined. Below is an overall presentation of the findings per category of questions asked to the participants.

3.1. AI4TRUST Platform - Findings

The first questions were about needs and expectations of the potential users of the AI4TRUST platform, where responses were collected both as part of the brainstorming session and through the questionnaire. As evidenced by the majority of responses, users' expectations regarding the platform focused on the ease of use and navigation, as well as on creating a sense of security and trustworthiness, in order to be able to trust its use. At the same time, the needs in terms of daily workflow focused mainly on reducing the time they spent visiting a platform where all the tools are centralised and there is a history/archive per user, again focusing on the issue of reliability and trustworthiness. The following are some representative answers. Figures 9 and 10 display the word cloud of Q5 and Q6 respectively.

Q5. What are your expectations in the field of tackling disinformation and misinformation by using the AI4TRUST Platform?

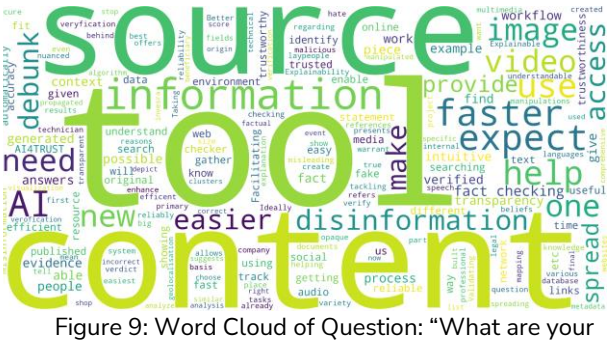


Figure 9: Word Cloud of Question: "What are your expectations in the field of tackling disinformation and misinformation by using the AI4TRUST Platform?" excluding the term "platform"

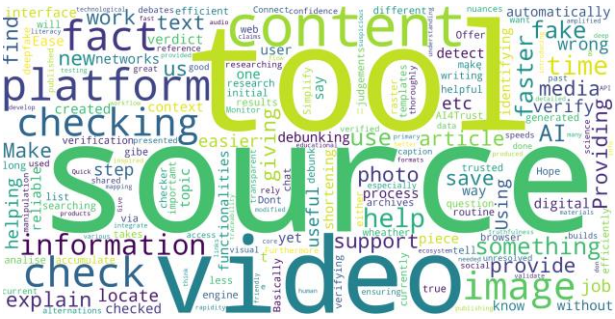


Figure 10: Word Cloud of Question: "How can the AI4TRUST platform support your everyday workflow?"

U1: A secure, fast, efficient one-size-fits-all platform to track AI-generated content or check the veracity of some content.

U2: Transparency and explainability of the platform's content providing the evidence reliably.

U3: A way to reach out to the primary or original source of the content and how this content is spread in the various social networks.

U4: To support and enhance the current workflow with trusted tools; efficient verification process.

U5: A platform that unifies several individual tools and creation of a database that includes verified content.

U6: Support fact-checking process by helping the user to debunk false information and to identify misleading statements.

U7: Provide the appropriate documentation in which this content is spread: who is the source of it, why it's spreading right now, is it used to depict a specific event in a correct or incorrect way, etc.

U8: A list of trusted sources for any information (news).

Q6. *How can the AI4TRUST platform support your everyday workflow?*

U1: Being responsive and time-efficient.

U2: Helping me support why some content is AI generated or not and find the initial source of the content.

U3: Providing trustworthy tools and supporting the verdict generation or reliability check of content.

U4: Providing a platform that will accumulate all different tools in one (like archives, debunk templates, list of sources, deep fake checking, research engine for checking photos).

U5: Monitor the content that is shared on social networks.

U6: Giving a better understanding of the media ecosystem.

U7: Providing us an API so we could integrate your platform's core functionalities into our workflow.

During the brainstorming session on the elements they would expect from a platform in order to trust it and integrate it into their daily workflow, as well as the needs it should address, the responses again centred on its design, functionality, and content, aiming to make it both appealing and trustworthy. In more detail:

- **Design**

Easy to use and simplistic; tutorial/guidelines regarding its use; visually appealing website including colour/liveliness, more icons/images, visible logos and titles, graphics, etc; an overview of

its content; appropriate font and button size (e.g., sign in); visible tabs with the platform's functions (e.g., toolbox, project's website, users' reviews/suggestions, etc).

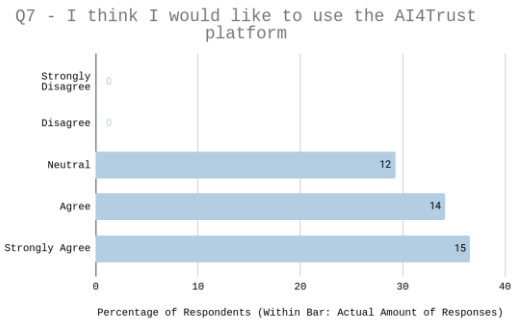
- **Content**

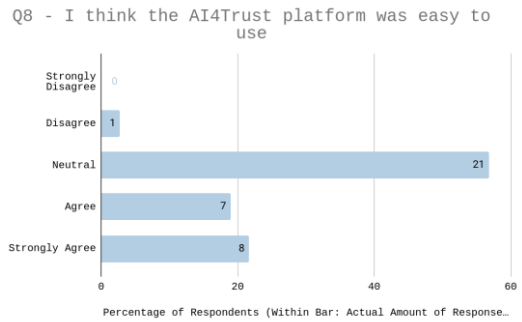
Presentation of the consortium, project, funding, and scope; contact details; data protection and security policy; terms of confidentiality; cookies policy; transparency of results & explainability; proper documentation; providing guidance on how to deal with misinformative narratives or content; providing context of the news stories and sources of the news; reliability (reproducibility), time response, accessibility, explainability what it includes and the production of the results.

- **Functions**

Easy navigation - improving user experience; reporting the limitations of the platform; verdicts to be based on a variety of classifiers; detecting a variety of disinformation signals; responsive platform for different devices, accuracy, simplicity and trustworthiness; user accessibility & simple UI; tools benchmarked against humans; time-efficient response; including verified sources; integrating internal databases within the platform to directly feed content that is found within; common database across fact-checkers to store material worked on and consult with other fact-checkers; archive of misinformative content.

The following questions (Q7-Q16) shown in Figures 11 and 12 are linked to the *System Usability Scale (SUS)* regarding the usability of the platform. At this point it should be stressed that the platform evaluation in terms of usability, as well as the other indicators below, is not particularly representative due to the level of the platform's readiness. Since the first version of the platform primarily focused on hosting the individual tools, participants found it challenging to distinguish between evaluating the platform as a whole and assessing its individual components (such as the AI4TRUST validation tools) due to lack of content. This was evident through qualitative feedback as at the time many participants found the questionnaire unnecessary, thus some of the participants omitted to answer some questions or gave a *neutral answer*. Nevertheless, this process aided the gathering of desirable features that will additionally contribute positively to the design and integration process of the platform as well as to serve as a baseline for comparative studies in subsequent piloting rounds.





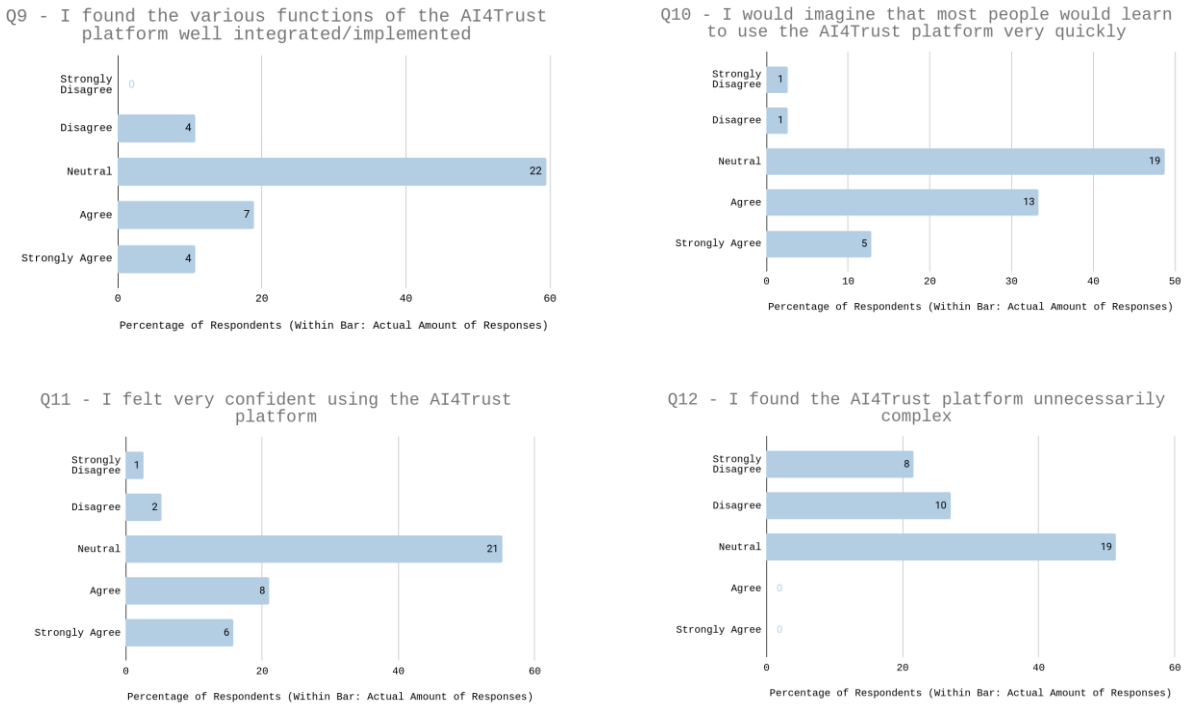


Figure 11: Percentage of responses distributed over the questions relative to SUS

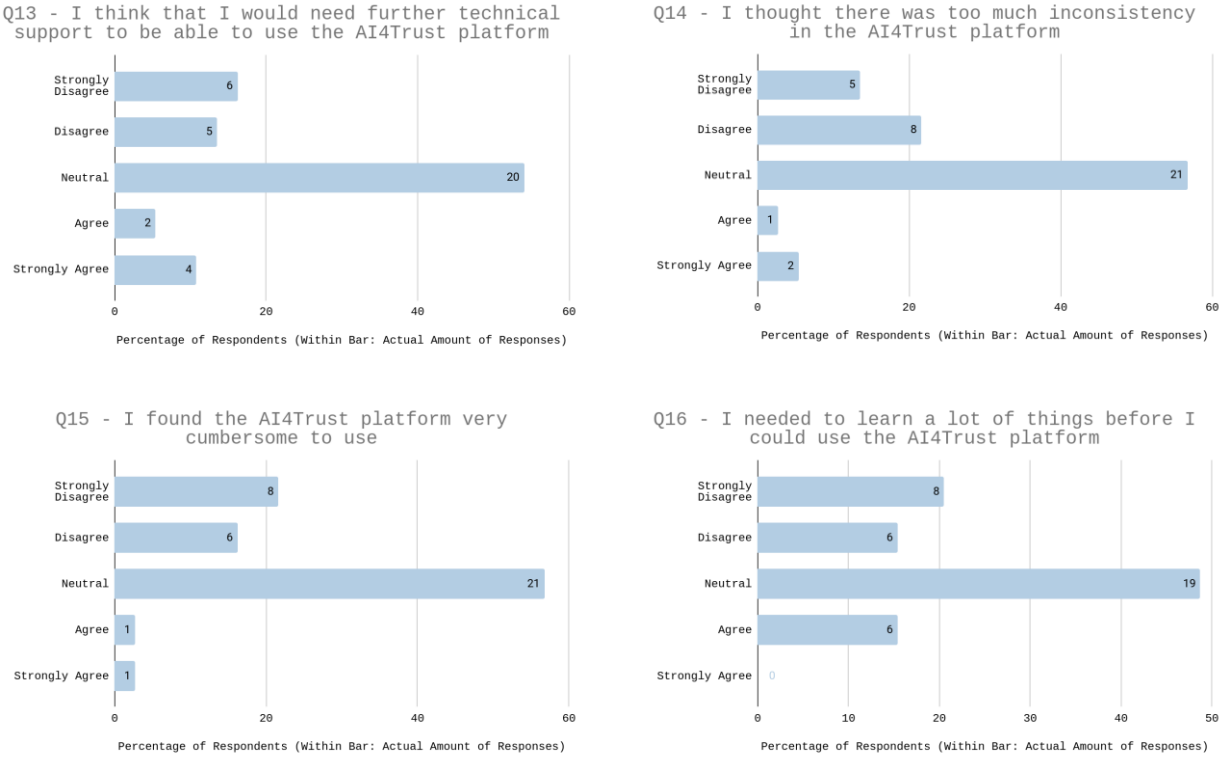


Figure 12: Percentage of responses distributed over the questions relative to SUS

Based on the results above, most participants (strongly) agreed that they would like to use the AI4TRUST Platform; however the majority seemed unable to assess the usability (neutral) due to

the level of readiness of the current time period. The following questions (Q17-Q22) shown in Figure 13 present users' opinions regarding the platform *functionality*, however taking into consideration the limitation already mentioned regarding the level of the platform's integration and readiness.

The participants were also asked to provide feedback regarding the platform's functionalities. According to the answers above, the majority (54,3%) chose 'user friendly navigation' as the function provided mainly by the platform and most of them (40,9%) expressed their confidence in the 'Integration of fact-checking activities in the daily journalistic process' through the platform. The majority also expressed difficulty in assessing the quality of the results. At the same time, some open-ended questions followed, where participants had the chance to provide us feedback about missing or/and additional functions that could be integrated in the platform.

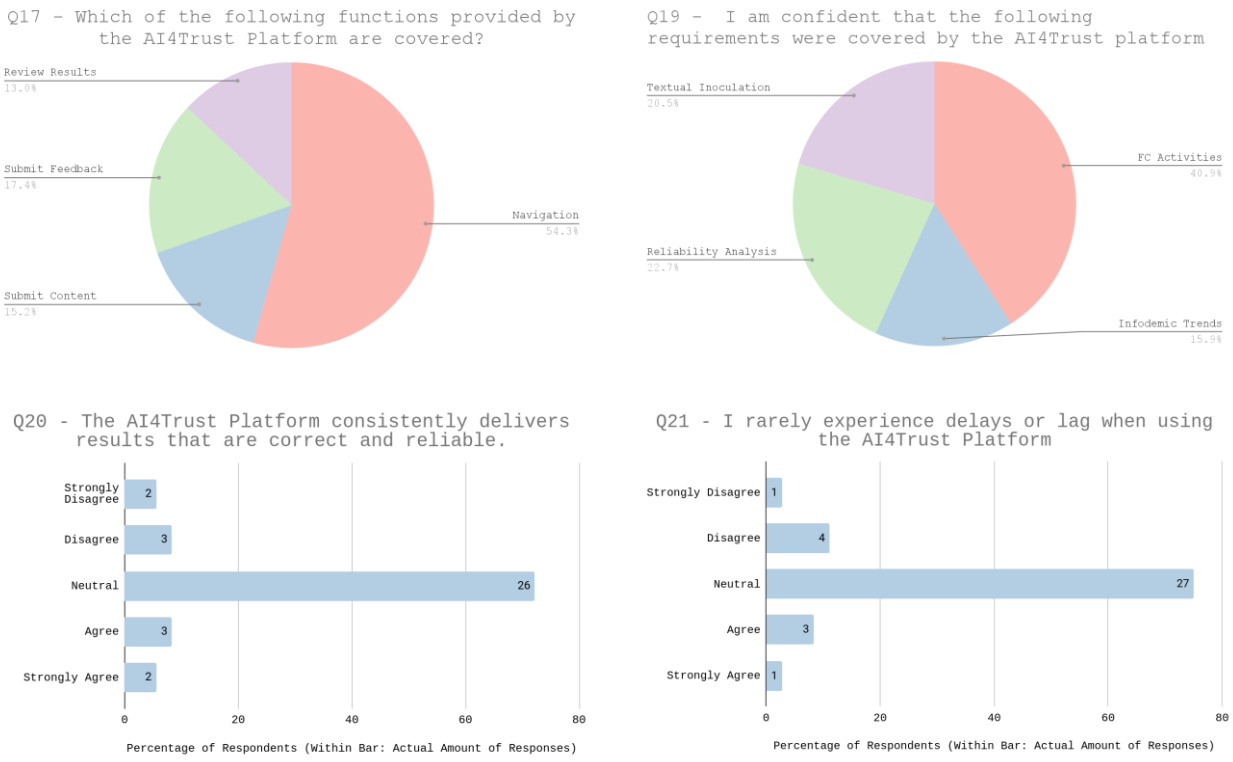


Figure 13: Percentage of responses distributed over the options of the question relative to functionality

Q18. Have you observed any missing functions from the AI4TRUST Platform?

Uploading content; reverse image search; terms of use (security & transparency terms); UI improvements; saving search results; tutorial-introduction on how to use the various functions; return to previous page (back button).

Q22. Do you have any suggestions about possible functionalities that can be added to the AI4TRUST Platform?

Uploading content from local storage; reverse image search; terms of use (security & transparency terms); pop-ups that guide the user in order to use the platform in the easiest way possible;

reliability analysis of a source, including social media profiles; providing detailed indication of which features within the video or text are manipulated; automatic archiving of the content; transparency on AI solutions and data policy; tutorial / help button; UI improvements to establish trust and ease of use, including limitations of the solutions

The following questions (Q23-Q37) are addressed to the indicators of *Effectiveness*, *Efficiency*, *Satisfaction and Learnability (EESL)* (also visualised in Figure 14 and 15 again taking into consideration the limitation already mentioned regarding the level of the platform's readiness.

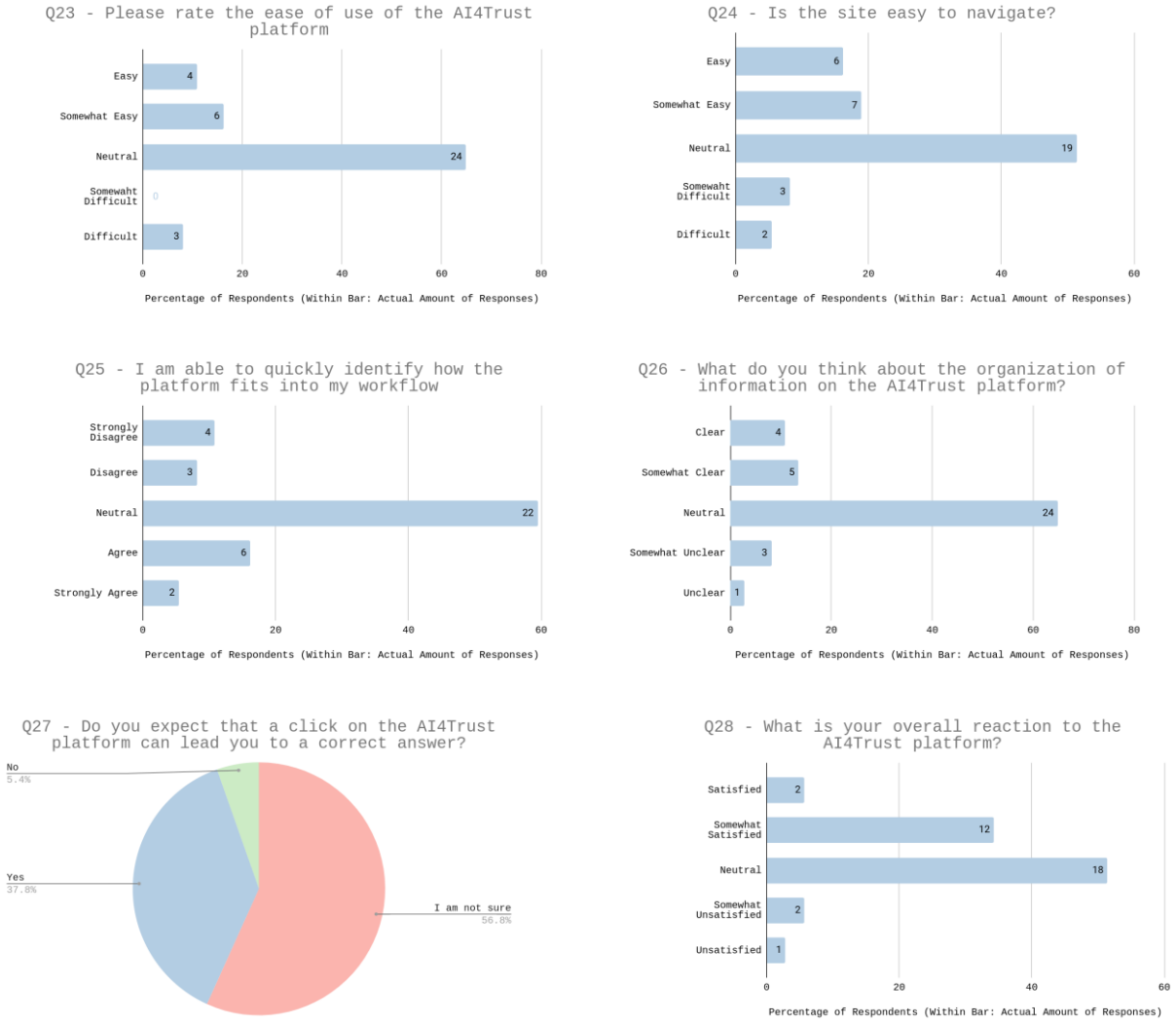


Figure 14: Percentage of responses distributed over the options of the question relative to EESL

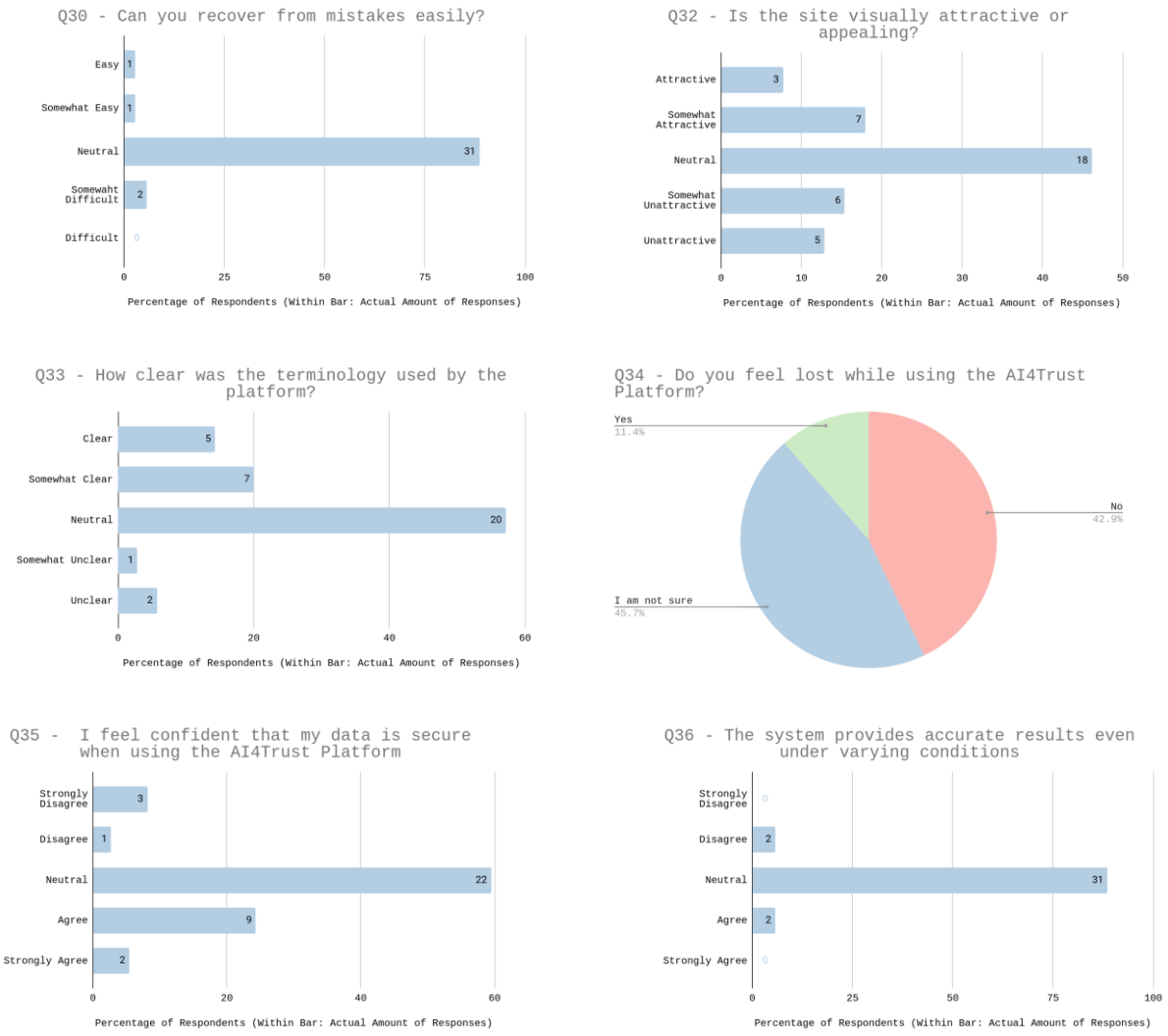


Figure 15: Percentage of responses distributed over the options of the question relative to EESL

As can be observed in the above responses, the majority of participants again expressed their difficulty (neutral) in evaluating the platform in terms of *Effectiveness*, *Efficiency*, *Satisfaction and Learnability*, due to the degree of its readiness. From the other responses, however, we observed a positive attitude on their part regarding its use.

Q37. Do you have suggestions on how to improve the overall usability of the AI4TRUST platform?

Clear instructions; better design - UI/UX friendly; data protection policy.

3.2. AI4TRUST Tools - Findings

Following the same structure as the platform's evaluation questionnaire, the first questions were about the needs and expectations of the potential users of the AI4TRUST tools, where responses were collected both as part of the brainstorming session and through the questionnaire. As

evidenced by the majority of the responses, users' expectations regarding the tools focused on the ease of use as well as the effectiveness and trustworthiness. At the same time, the needs in terms of daily workflow concentrated mainly on reducing the time of the fact checking process including the production of trustworthy content based on documentation and explainability. Some indicative responses are provided below. Figures 16 and 17 display the word clouds for Q5 and Q6, respectively.

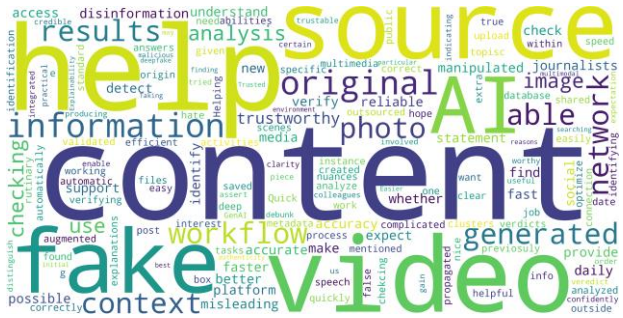


Figure 16: Word Cloud of Question: "What are your expectations in the field of tackling disinformation and misinformation by using AI4TRUST tools?" excluding the terms "tool", "tools", "fact", "checkers" and "fact-checkers"

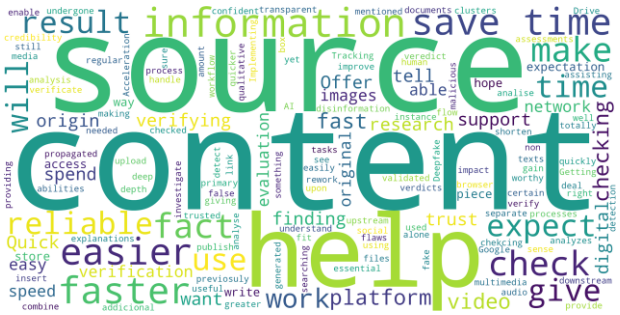


Figure 17: Word Cloud of Question: "How do you expect AI4TRUST tools to support your everyday work flow?" excluding the terms "tool" and "tools"

Q5. ***What are your expectations in the field of tackling disinformation and misinformation by using AI4TRUST tools?***

U1: Easy to use and help optimisation of workflows.

U2: Efficiency and trustworthy.

U3: I hope to have a standard tool to confidently assert that content has been generated or manipulated by AI.

U4: Save results and create a database .

Q6. ***How do you expect AI4TRUST tools to support your everyday work flow?***

U1: Save time. Tracking the flow of a certain disinformation - upstream (the original source of the information to be checked) and downstream (to investigate its impact).

U2: Check content from a variety of sources.

U3: Integrate with other platforms, e.g., connect with Google Drive to automatically fetch data from there.

U4: Provide fast, efficient and trustworthy answers.

U5: Providing a combination of individual tools.

As part of the brainstorming session regarding what functions they would expect to be integrated in the AI4TRUST validation tools, participants provided some very interesting views, mainly

focused on: (i) the time needed; (ii) the need for supporting a variety of sources as input, including social media; as well as (iii) the results to be supported by a narrative and additional documentation. Some of the answers per tool category are reported below.

- **Image:**

Including detection of manipulation by other means than AI; checking whether an image was used at other sites; easier/simpler way to communicate why the image is synthetic; text transcription from image; showing reverse results in chronological order; including detection of manipulation by other means; showing which part is manipulated rather than just a percent of yes/no.

- **Audio:**

AI generated voice detection - target recognition of VIP's/celebrities as their voices are frequently used for AI generation; ENG Transcriptions; video and audio analysis unification; highlighting the part that was modified or altered; editing/providing corrections on the transcription.

- **Text:**

Easier/simpler way to communicate why the text is misinformative (clickbait, etc.); directing scraping from a link/website; support for a variety of languages; highlighting what is check worthy within the text rather than only presenting a percentage; check-worthy claim should be more explicit on which parts are check worthy and why; the hate speech detector that accepts terms of hate speech and categorise them as hate speech would keep the platform updated.

- **Video:**

Supporting videos from a variety of sources, including SoMe TikTok, etc.; transcription; showing a collage of thumbnails before analysing the video to rapidly examine the content; image and video to be unified; progress bar or estimated time needed for analysis; video and audio analysis unification; highlight parts of the video that have been altered; detect other things than face swapping.

The following questions (Q7-Q16) shown in Figure 18 are linked to the *System Usability Scale (SUS)* regarding the usability of the AI4TRUSTt tools.

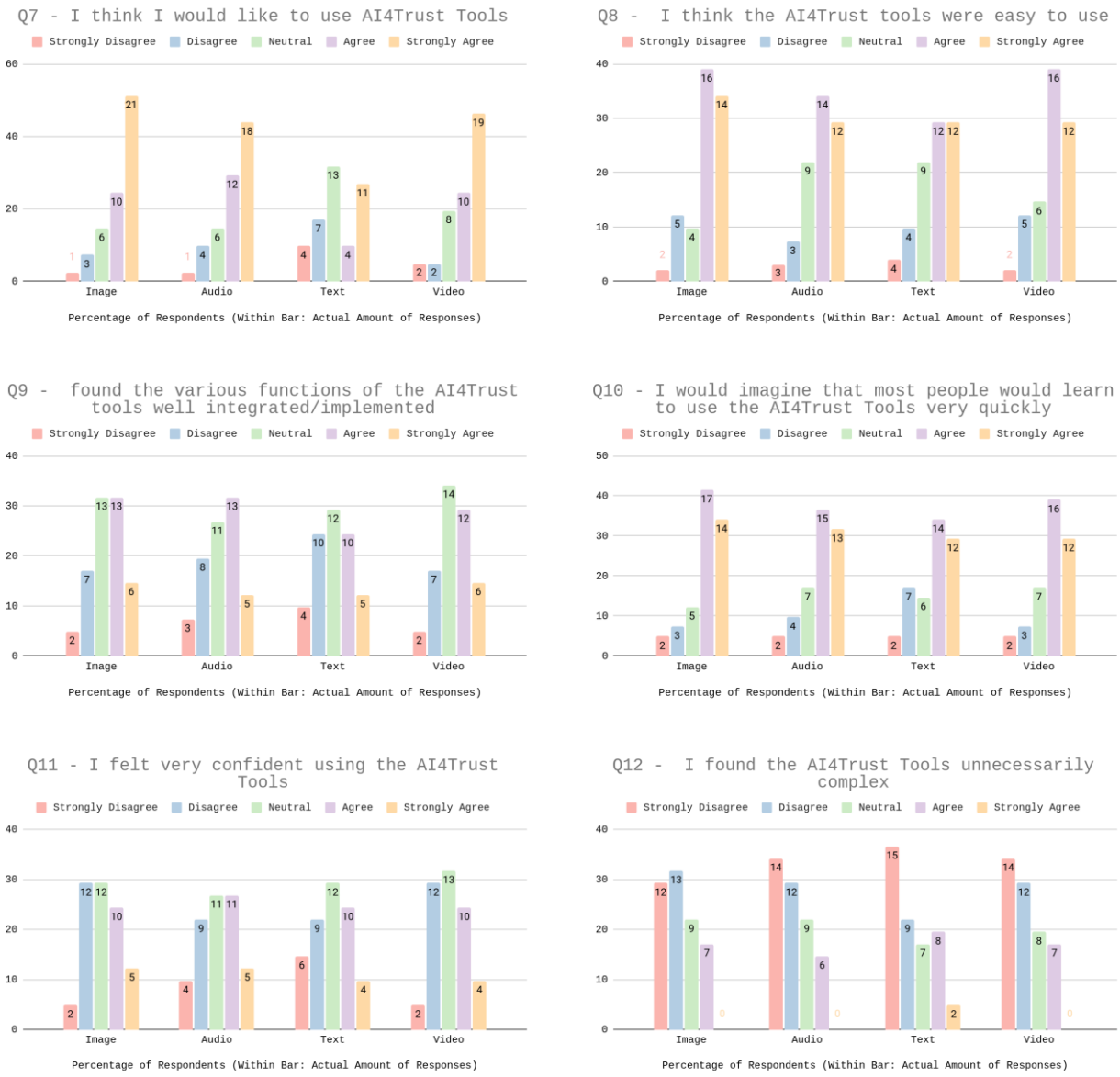


Figure 18: Percentage of responses distributed over the options of the question relative to SUS

According to the answers above, most of the participants seemed positive to use the AI4TRUST tools with a greater preference for image (31/41) and video (30/41) tools and a lesser preference for text tools (15/41), which is mainly linked to their daily needs. They also agreed that their use was easy and most of them argue that the various functions of the tools were well implemented. However, from their answers we understand that they did not feel confident to use them.

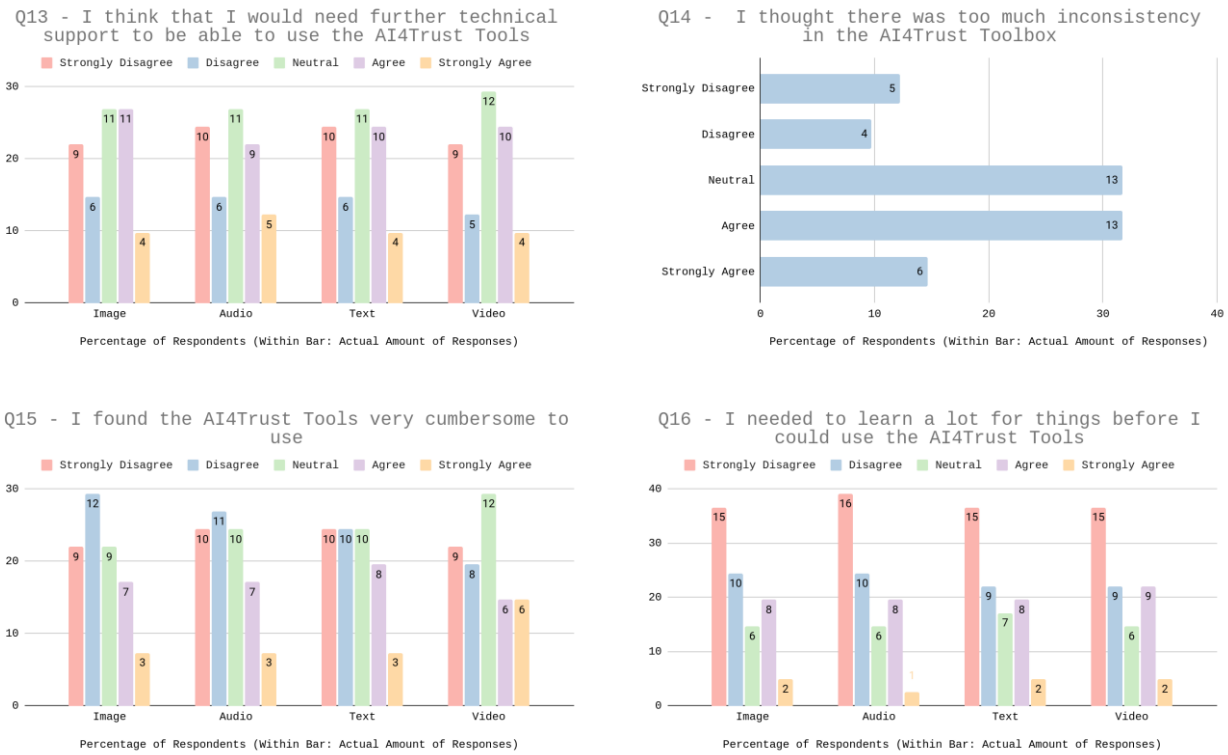


Figure 19: Percentage of responses distributed over the options of the questions relative to SUS

This is also confirmed by their responses as visualised in Figure 19, with the majority (14/41) stating that they would need additional support to use the tools and enough to declare neutrality (11/41), which is perhaps partly due to the fact that they thought that there is much inconsistency in the AI4TRUST Toolbox. However, the majority of participants (average: 20/41) did not find the tools cumbersome to use, nor did they believe that extensive training was required before using them (average: 25/41).

The following questions (Q17-Q22) are presenting users' opinions regarding the tools' *functionality*.

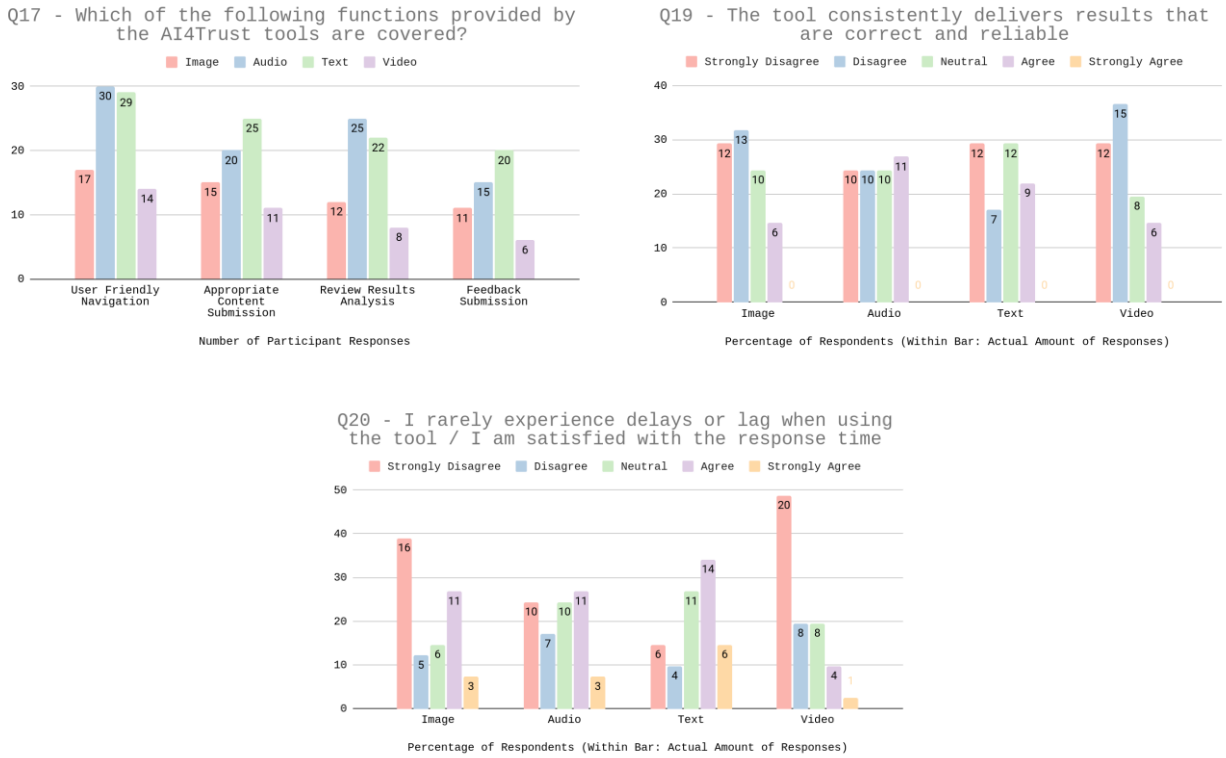


Figure 20: Percentage of responses distributed over the options of the questions relative to functionality

According to the answers above, the majority reported that the four functions provided (i.e., user friendly navigation, appropriate content submission, review results analysis & feedback submission) are mainly covered by the tools of audio and text, followed by the corresponding tools of image and video.

Regarding the quality of the results per tool, the majority expressed a lack of confidence in their correctness and reliability, especially about the tools of image and video, while expressing a dissatisfaction with the response time, with the possible exception of the text tool.

The participants were then asked the following 2 open-ended questions to suggest new functions and/or make suggestions for the existing ones:

Q18. Are there any other functions that you would like to be provided by each tool?

- **Image:**

Tutorial - how to put links within the tool; including images to be relevant concerning deep fake; highlighting areas that appear manipulated; extracting transcription; clearing definition of decision boundaries; checking if it was manipulated by other media (e.g., photoshop); progress bar or estimation of time needed; chronological order of results in reverse image search; searching for original source; filtering results (date, relevance or best match); file upload feature.

- **Audio:**

Optimised language support; audio and video analysis unification or a quick way to switch between them on a single data sample; file upload feature; automatically showing minutes next to transcription (so that it can be easily found) and showing what Bonafide and Anomalous are when it comes to factors; sending transcription to text analysis.

- **Text:**

Including support of all languages in all text tools; checking for plagiarism; detecting content manipulated by other means; when it comes to the clickbait or hate speech analysis, providing specific data may help to debunk it, otherwise just highlighting it as a suspicious claim; providing results in other languages as well; bigger character limit; comparison of the statements with other sources.

- **Video:**

Capability to faster identify fake scenes within longer videos; an alternative easy to use editing tool that allow user to trim and extract specific scenes of interest for analysis; detecting whether content has been manipulated by other media; sensitive content tag; progress bar or estimated time; unifying video and audio analysis or easy back and forth.

Q21. Do you have any suggestions about possible functionalities that could improve the tools?

- **General suggestions:**

A way to report errors, e.g., suggestion box - reporting button; archiving service; button to check additional content or different modality; functionality to check whether a piece of text and other content are related in terms of context; transparency in scales and standards used to determine the percents; how to base that something is AI-gen; creating database with verified materials; including all languages for all tools.

- **Suggestions for image tools:**

Considering checking the lighting within an image to improve the efficiency - fakes include inconsistent lighting; including other manipulation except AI.

- **Suggestions for audio tools:**

The tool should be able to identify the language particularities of the original character - let's name it the meta language of the original person/character voice as compared to the fake voice; checking whether the content of the audio - not only if the style is fake.

- **Suggestions for text tools:**

The tool should be able to identify the source of the original text of news; hate speech and possibly other classifiers are very sensitive and suggest that content is positive when it is not; verdict generation should be separated from this tab and have an individual tab.

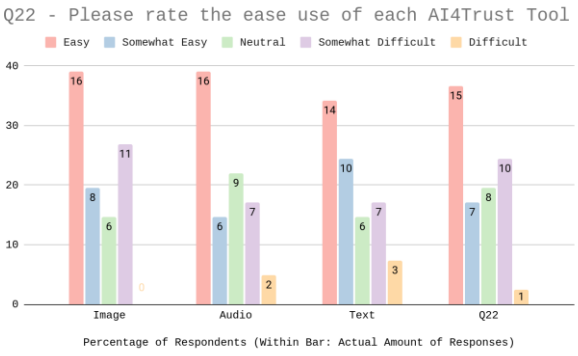
- **Suggestions for video tools:**

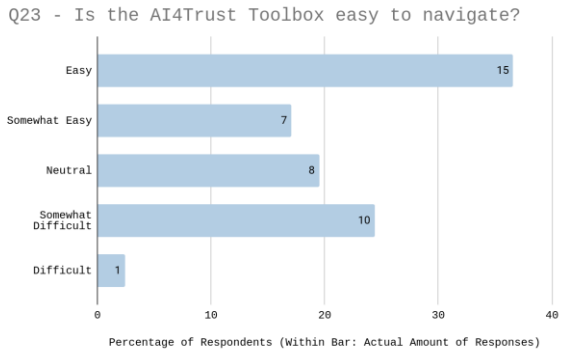
It should be faster.

The following questions (Q23-Q37) shown in Figure 21 are addressed to the indicators of E*ffectiveness, Efficiency, Satisfaction and Learnability (EESL)*.

According to the answers above, the majority described the tools as easy to use (average: 23.5/41), as well as navigating the toolbox (22/41). However, they seemed cautious about how to fit them into their workflow. Most of the participants (average: 17/41) described the organisation of information on each tool as clear or somewhat clear, although a large number (average: 11/41) declared neutrality. While in the case of the text tool, the number of participants (16/41) who described the organisation of the information as clear or somewhat clear coincides with the number of those (16/41) who reported that it is somewhat unclear or unclear.

Subsequently, the majority (average: 21.5/41) expressed distrust in the correctness of the answers of the tools. Regarding the overview evaluation of the tools, responses varied by tool, with satisfaction predominating in the case of the image (16/41) and audio (21/41) tools and dissatisfaction in the case of the text (17/41) and video tools (17/41), while a relatively high percentage declared neutrality (average: 9/41).





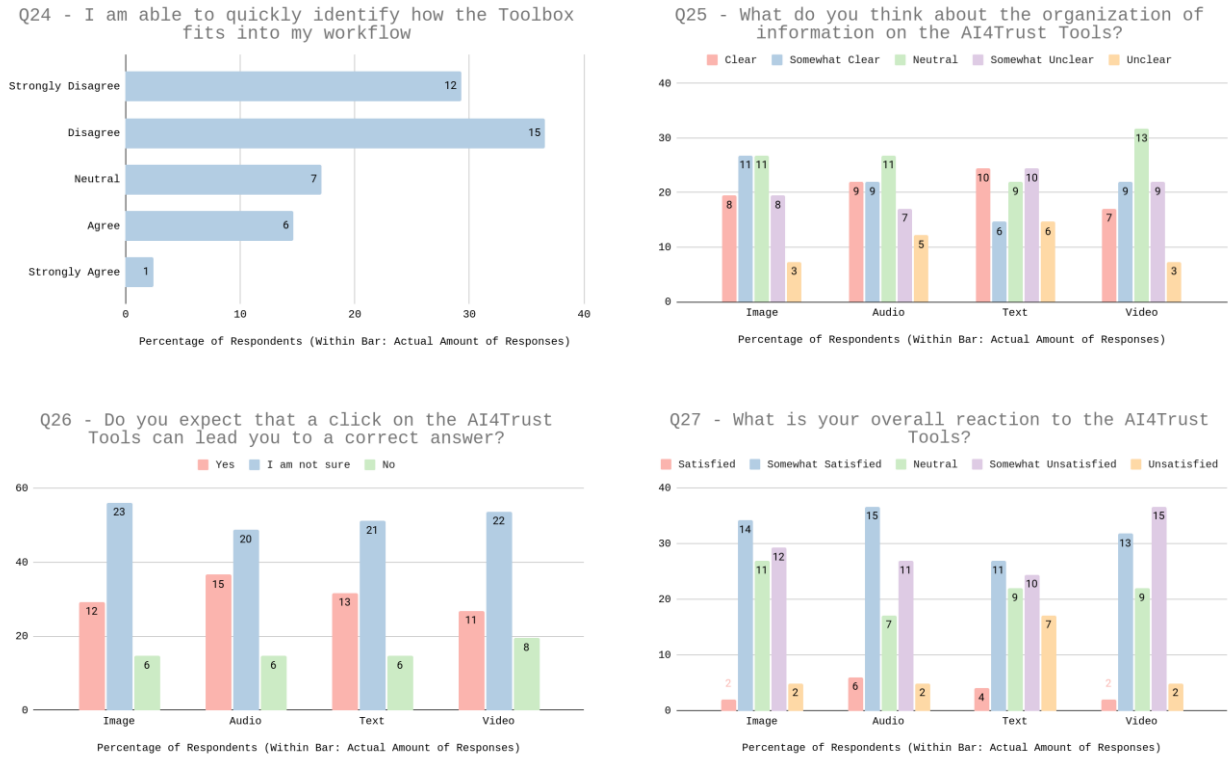


Figure 21: Percentage of responses distributed over the options of the question relative to EESL

The following responses shown in Figure 22 address the satisfaction regarding the efficiency and effectiveness of the AI4TRUST tools, as well as satisfaction of the interaction with them. Regarding the ease of recovering mistakes, most participants answered neutrally (average:18.5/41), as did most responses regarding the visual attractiveness of the AI4TRUST tools (16/41). This response is probably justified considering the fact that the tools are under development and participants were aware of this. Thus, they remained reluctant and expected their last version of the AI4TRUST tools before expressing a final verdict. In general, they feel satisfied with the terminology used (average: 21/41), although the percentages stating that the terminology is unclear or somewhat unclear are equally high (16/41).

The majority (58,5%) found the Toolbox easy to use, as they do not feel lost when using it. However, they expressed uncertainty (neutrality) regarding the security of their data (20/41). Finally, most participants (23/41) appeared to disagree with the idea that the Toolbox provides accurate results under varying conditions.

Figure 22: Percentage of responses distributed over the options of the question relative to EESL

The next question (Q28) focused on noticing mistakes in the evaluation of instances during the AI4TRUST tools' use. According to the responses (see Figure 23), in the case of the image and video tools, more mistakes were observed compared to the other two tools. The majority of mistakes can be categorised as the system producing false positives or false negatives when determining whether content has been artificially generated or manipulated. Specifically, this means the system may fail to identify manipulated content (false negative) or incorrectly classify non-manipulated content as altered (false positive).

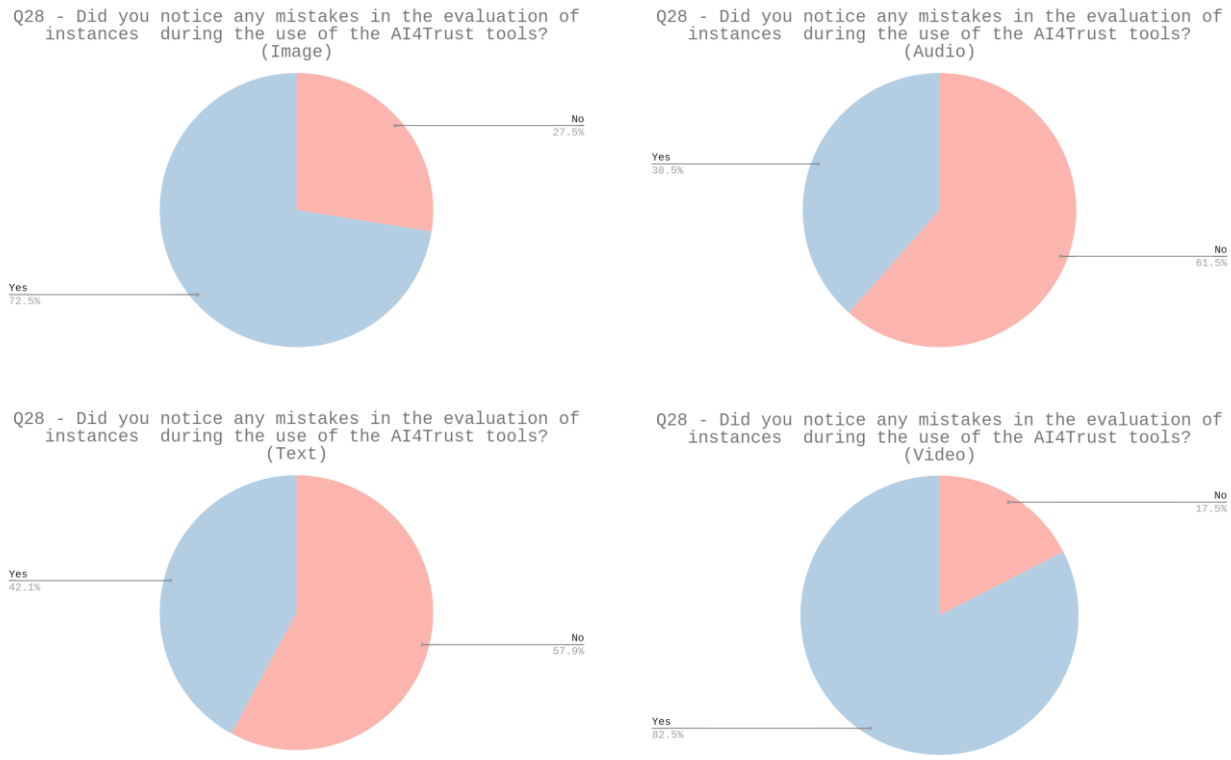


Figure 23: Percentage of responses distributed over the options of the question relative to EESL

Q36. Do you have suggestions on how to improve the overall usability of the AI4TRUST tools?

Progress bar / estimated time; tutorial regarding the use; improve response time; data policy.

4. Discussion and Recommendations

4.1. Findings Overview

In this Section, we aim to provide an overview of the findings in order to draw some conclusions that will contribute to improving both the AI4TRUST Platform (WP5) and the respective tools (WP3). Before presenting these results, it is interesting to focus on the demographic characteristics of the participants, highlighting some data and drawing some conclusions.

As mentioned in Section 2.5, most of the participants were within the age group of 26-30 (36,4%), which could perhaps lead to the conclusion that professionals in this age group are interested in new technologies that could potentially alleviate identified drawbacks or offer new services that are currently unavailable. At the same time, the next big age group that was older than 41 (29,4%) was interested in experiencing new tools that allow them to reduce their effort or assist them in

their established workflows. This does not mean that other age groups did not show the same interest in improving their daily workflow by experiencing new tools.

Before proceeding to further analysis, it would be useful to mention that deployment of the pilots is structured via two iterations in order to ensure maximal acceptance and impact of the AI4TRUST Platform and its tools at the final release. In order to obtain rich and complete results, the evaluation strategy adopted a mixed-method design: the first phase of the piloting focused on smaller groups of users to gain qualitative as well as quantitative insights on *usability*, *functionality*, *user experience and attitudes* and perceived *impact and acceptance* as part of their everyday work flow, while the second phase of piloting aspired to evaluate the AI4TRUST Platform and its tools with larger groups of users. The aim was again to gather additional quantitative as well as quantitative data into *usability, functionality, impact, and acceptance* of users regarding both the platform itself and the respective tools and their functions, in order to support with a productive and effective way their everyday tasks in the field of fact checking and news validation.

As already mentioned, for the evaluation of both the AI4TRUST platform and its tools, online questionnaires were used that had the same structure and the questions served to evaluate specific indicators that will be mentioned again below and are mainly related to *functionality* and *usability*. At the same time, however, it was important to investigate, on the one hand, the *attitude* of users towards these tools and, on the other hand, the *impact* they may have on their daily work flow. The answers to these questions therefore emerged largely both from the brainstorming sessions and from the discussions during the navigation of the AI4TRUST Platform and the interaction with its tools.

In conclusion, we can say that the design of an easy-to-use, trustworthy, and effective platform, where all validation tools are centralised while maintaining an analysis history/results archive for each user, would be highly appreciated. Similarly, the use of validation tools targeting different modalities—such as images, video, text, and audio—that provide reliable and comprehensive results, along with the appropriate documentation, would be particularly valuable in enhancing their daily workflow. These tools would help save time while simultaneously making the results of their work more robust and dependable.

The findings of this report would be utilised within the respective technical WPs to further improve their services in subsequent versions to ensure that the tools better align with the expectations of the end users (e.g. fact checkers and journalists), refining their functionality and user experience. Similarly, the platform will integrate the insights gained from these evaluations, further enhancing its design and performance. By addressing the requirements highlighted by the users, both the tools and the platform will be optimised to support more efficient workflows and provide even more reliable and comprehensive results for all stakeholders. The following sections provide a more detailed presentation of the findings, feedback, and suggestions for improvement.

4.1.1. AI4TRUST Platform

Regarding the piloting and evaluation of the AI4TRUST platform, it is worth reiterating that the quality of the collected data does not fully meet our needs due to the platform's level of readiness at the time (i.e., its focus on hosting individual tools and the lack of relevant content) during the workshops. This led to users encountering difficulties while navigating the platform. Nevertheless, some key conclusions based on the data collected for each indicator are outlined below.

- **System Usability Scale (SUS)**

Participants (strongly) *agreed that they would like to use the AI4TRUST Platform*; however, the platform's usability evaluation should be repeated when the level of readiness of the platform will be increased (within subsequent evaluation rounds and their reports).

- **Functionality**

Participants:

- recognised the '*user friendly navigation*' as the function *provided mainly* by the platform;
- expressed their *confidence* in the '*integration of fact-checking activities in the daily journalistic process*' through the platform;
- expressed *difficulty* in *assessing the quality of the results* due to the lack of transparency in the tools' results;
- as *suggested functionalities*, they mentioned the followings: *uploading content from local storage; reverse image search; terms of use & conditions (security & transparency terms); pop-ups that guide the user in order to use the platform in the easiest way possible; reliability analysis of a source, including social media profiles; automatic archiving of the content; transparency on AI solutions and data policy; tutorial / help button; UI improvements to establish trust and ease of use, including limitations of the solutions; tutorial-introduction on how to use the various functions; return to previous page (back button).*

- **Effectiveness, Efficiency, Satisfaction and Learnability (EESL)**

As mentioned in the case of the evaluation of the AI4TRUST Platform in terms of SUS, the majority of participants again expressed their difficulty (neutral) in evaluating the platform in terms of *Effectiveness, Efficiency, Satisfaction and Learnability*, due to the degree of its readiness. Although according to the replies of the others, we observe a *positive attitude* on their part regarding its use.

4.1.2. AI4TRUST Tools

The data collected was largely reliable, as the AI4TRUST tools were sufficiently developed, allowing participants to engage with them using real examples from their daily workflows. They were then able to provide feedback through discussion sessions and the questionnaire provided. As previously mentioned, users' expectations for the tools centred on *ease of use*, as well as *effectiveness and trustworthiness*. At the same time, their needs in terms of daily workflow primarily focused on reducing the time spent on the fact-checking process, including the production of trustworthy content based on documentation and explainability. Overall, we can characterise

their attitude towards the AI4TRUST tools as positive, foreseeing the potential positive impact that an improved version of the tools (v2 - D3.2) could have on their daily work. Some key conclusions based on the data collected for each indicator are outlined below.

- **System Usability Scale (SUS)**

Participants:

- ➢ seemed **positive** to *use the AI4TRUST tools* with a greater preference for **image** (31/41) and **video** (30/41) ***tools***, which is mainly linked to their daily needs;
- ➢ **agreed** that their *use was easy* and **argued** that the *various functions* of the tools were ***well implemented;***
- ➢ ***did not feel confident yet*** to *use the tools*, as they would need ***additional support*** to use them;
- ➢ **thought** that there is *much inconsistency in the AI4TRUST tools, although* ***they did not find*** *the tools cumbersome to use* and deemed *a lot of training is not required* before using them.

- **Functionality**

According to the majority of responses:

- ➢ *all four functions* (user friendly navigation, appropriate content submission, review results analysis, and feedback submission) provided are *mainly covered by the tools of audio and text***,** followed by the corresponding tools of image and video;
- ➢ there is *a lack of confidence* regarding the tools' results about their correctness and reliability, especially about the tools of *image and video;*
- ➢ there is also a *dissatisfaction* with the *response time* except perhaps in the case of the *text tool.*

A summary of both the suggested and missing functionalities for each modality is provided below:

- **Image:** tutorial regarding the functions & use of each tool; including images to be relevant concerning deep fake; highlighting areas that appear manipulated; text transcription; learn definition of decision boundaries; clarification if it was manipulated by other media (e.g., photoshop); progress bar or estimation of time needed; chronological order of results in reverse image search; search for original source; filter results (date, relevance or best match); file upload feature.
- **Audio:** optimised language support; audio and video analysis unification or a quick way to switch between them on a single data sample; file upload feature; automatically showing minutes next to transcription (so that it can be easily found) and showing what Bonafide and Anomalous are when it comes to factors; sending transcription to text analysis.
- **Text:** including support of all languages in all text tools; checking for plagiarism; detecting content manipulated by other media; when it comes to the clickbait or hate speech analysis, providing specific data may help to debunk it, or otherwise, just highlighting it as a

suspicious claim; bigger character limit; comparison of the statements with other sources; identifying the source of the original text of news; verdict generation should have an individual tab.

- **Video:** capability to faster identify fake scenes within longer videos; an alternative easy to use editing tool that allow user to trim and extract specific scenes of interest for analysis; detecting whether content has been manipulated by other media; sensitive content tag; progress bar or estimated time; unifying video and audio analysis or easy back and forth.

From the responses we could distinguish some suggestions that are common to all the tools and that would be valuable taking into consideration in the context of improving the AI4TRUST tools. These suggestions could be listed as follows:

(i) tutorial regarding the functions & use of the tools; (ii) progress bar or estimation of time needed; (iii) searching for original source; filter results (date, relevance or best match); (iv) file upload feature; (v) archiving service - create database with verified materials as well as misinformative content; (vi) button to check additional content or different modality; (vii) including all languages for all tools; (viii) reviewing report after its use regarding the validation of the results.

4.2. KPI Assessment

This section presents a KPI assessment after completion of the first piloting round. We evaluate the progress made toward meeting the preliminary targets set in the previous planning phase (pilot planning within D6.1). Table 4 presents the current progress toward the set KPIs, reflecting both quantitative and qualitative measurements.

The alignment of the preliminary KPIs with the characteristics of the AI4TRUST Quality Model provides an indication of the current state of the platform. In terms of functional suitability, the platform has met the planned KPIs, except for the sub-characteristic of functional completeness. Of the four initial pilot requirements — reliability analysis of information sources, integration of fact-checking activities into the daily journalistic process, infodemic trends for specific public interest issues, and evidence-based textual inoculation — only the first two are addressed by the first version of the platform.

Regarding the platform's efficiency, it appears that the AI4TRUST Platform is capable of handling large volumes of requests and responding timely to them. The resource utilisation is rather efficient, with proper adjustments under stress level and low power consumption. The initial assessment of appropriateness and recognisability revealed that several user requirements have not been met by this version of the platform (currently covering 9 out of 24), although the platform seems relatively easy to learn, as most participants were able to grasp its functionality within a few minutes. In terms of operability, the first assessment is close to the desired target (SUS score of 68), but slightly

lower, with an average score of 62.9, indicating that while usability is satisfactory, there is room for improvement. Lastly, user engagement was extensive, due to dedicated interaction sessions.

Initial factors for trust in the platform did not seem to be covered as from the qualitative feedback received from the participants it appears that the lack of terms that clear explain who is behind the platform, under which guidelines it operates and how their data are handled is a pressure point that does not enable an environment of trust. At the same time, to integrate the tools in their daily work requires additional explainability to understand how the tool reaches particular conclusions and visual aid that suggests parts of the content that seem to be manipulated, fake or tampered with using any means.

Lastly, in terms of reliability, the AI4TRUST Platform demonstrated no performance degradation over several weeks of uptime. The other characteristics were not applicable in the current version, as only a single component is part of the system. Since the platform operates on a single web server that hosts both the front-end and back-end, experimenting with various faults, modifications, or other simulations is not feasible at this stage. This will be revisited in future versions of the platform.

Quality Characteristic	Test	Preliminary KPIs	Assessment based on 1st evaluation round
Functional Suitability			
Functional Completeness	Observation tests via usage scenarios	Cover 80% of pilot requirements	50% of pilot requirements are covered
Functional Correctness	Observation tests via usage scenarios	Over 80% scenario accuracy rate[4]	100% scenario accuracy rate
Functional Appropriateness	Observation tests via usage scenarios	Over 80% task suitability rate [5]	100% task suitability rate
Performance Efficiency			
Time Behaviour	Performance testing (Single user)	• Throughput: At least 100 requests/second	Throughput: 20 requests per second[6]

[4] Scenario Accuracy: Correctly Executed Scenarios / Total Scenarios Tested
[5] Task Suitability Rate: Suitable Tasks / Total Tasks Tested
[6] Tested RPS are lower due to hardware limitations of the testing equipment (work laptop could not go past 60 RPS)

		• Response: less than 2 seconds for front-end • Lead time: Decrease compared to previous methods	Response: less than a second (0.65ms)
Capacity	Load stress (Multiple users)	• Throughput: At least 200 requests/second • Response: less than 2 seconds	• Throughput: 60 requests per second7 • Response: less than a second (0.65ms)
Resource Utilisation	Performance Counters	• Less than 80% utilisation of resources (CPU, Memory, Disk, Network) • Estimation of power consumption under idle and under stress	• Idle: ○ CPU: 1.5% (AVG) 38.1% (MAX) ○ Memory: 40.49% (AVG) 40.18% (MAX) ○ Disk: No significant use to report (<20%) ○ Power: At 50% CPU utilization (2 vCPUs), the estimated power consumption is 4.85W • Under Stress: ○ CPU: 18.79% (AVG) 100% (MAX) ○ Disk: No significant use to report (<20%) ○ Memory: 41.54% (AVG) 41.90% (MAX) ○ Power: At 100% CPU utilization (2

[7] Ibid.

			vCPUs), the estimated power consumption is 9.7W
Appropriateness Recognisability			
Appropriateness Recognisability	Observation tests via usage scenarios	Cover 80% of user requirements	37.5% of user requirements are covered
Learnability	Observation tests via usage scenarios	Less than 1 hour to complete a task without assistance on average	On average it took participants less than 3 minutes for participants to complete a task (excluding waiting for system's response)
Operability	System Usability Scale (SUS)	Over 68 System Usability Score	Average of 62.90 System Usability Score
User Engagement	Design Evaluation	Over 3 minutes of session length	45 minutes of interactive sessions
Trust	Observation tests via usage scenarios	Cover most factors dictated as important for assessing trustworthiness by the end-users.	• Data policies and security terms & terms of use are not included • Explainability of the toolbox results is not adequate
Reliability			
Faultlessness	Longevity testing	Retain scenario accuracy and task suitability after 2 weeks of uptime	It did retain these characteristics after 2 weeks of uptime
Availability	Longevity testing	Retain throughput rates after 2 weeks of uptime	It did retain throughput rates

Fault tolerance	Fault simulation	Over 95% of time the platform is available and operational	Not applicable in this version
Maintainability			
Modularity	Observation tests via usage scenarios	Change in a particular component should affect the availability of less than 25% of the rest of the components	Not applicable in this version
Reusability	Longevity testing	Over 70% of reusable components.	Not applicable in this version
Modifiability	Modification simulation	Over 95% of time the platform is available and operational Retain accuracy and task suitability during modifications	Not applicable in this version
Flexibility			
Scalability	Observation tests via usage scenarios	Over 80% of automatic adjustment of resources.	Not applicable in this version

Table 4: 1st evaluation round KPI assessment

4.3. Post- Pilot Activities / Community Building

Section 6.3 of D6.1 enumerates the activities entailed in the final phase of the piloting process, with the aim to appreciate the participants for their contribution and especially to highlight the importance of the community as part of the co-design process of the AI4TRUST Platform and to keep their engagement and their interest alive. In summary, these activities concern:

Acknowledgment and recognition of the community members who actively participate through public acknowledgments, certificates of participation, or incentives to show gratitude for their time and effort.

Regular updates and progress reports for keeping the community informed about the progress of the pilots, the AI4TRUST Platform and its tools.

Privacy and data protection for community members participating in online testing should be ensured by clearly informing them about how their personal information will be handled, stored, and used, in compliance with relevant legislation such as the EU General Data Protection Regulation (GDPR), the Digital Services Act (DSA), and the AI Act.

Continuous engagement beyond the pilot workshops, by maintaining a community platform or forum where participants can stay connected, share their experiences, and continue to provide input for future developments.

With the completion of the first phase of the evaluation, the need to maintain communication with the community was also highlighted by the participants themselves, who expressed their interest in continuing to interact with the AI4TRUST Platform and its tools and to be informed about their functional updates, as well as to be invited to participate subsequent evaluation rounds. Some participants also suggested organising online discussions per group partner organisation to discuss the results of the evaluation and offer further clarifications to their responses. On behalf of NCSRD, as WP6 leader and host organisation, the following activities have been/will be initiated:

- Publication of relevant announcements on the social media of the NCSR-D and the project, expressing our thanks to the participants and respecting the terms and conditions stated in the signed Consent Forms;
- Schedule an online meeting among involved partners, where the main findings of the evaluation will be presented, followed by a discussion and comments on them and a joint decision on the next steps regarding the improvement of the AI4TRUST Platform and its tools;
- Creation of a thank-you letter including the key findings of the evaluation and informing the next steps in terms of improving the AI4TRUST Platform and its tools. This letter will be sent to the partners who hosted the pilot workshops with the kind request that it be forwarded to all participants, highlighting their significant contribution to the co-design and co-creation process;
- As the AI4TRUST Platform and its tools undergo functional improvements, an update will be sent to the community, encouraging members to make further use of the tools and provide comments and suggestions that could lead to additional enhancements;
- Invitation to participate in subsequent evaluation rounds;
- Managing their personal data with respect and confidentiality, in accordance with the aforementioned applicable legislation, and using and publishing the findings of the evaluation for research purposes, strictly in accordance with the conditions outlined in the Consent Forms they have signed, i.e., only in aggregated form.

4.4. Roadmap of the Pilot Activities

This section acts as a brief overview of the piloting plan as implemented during the first evaluation round and an indicative timeline and activities for subsequent evaluation rounds.

4.4.1. Timeline Overview

This section presents the roadmap of the pilot activities related to the first evaluation round in order to act as a template for subsequent evaluation rounds. The activities of the first evaluation round have been described in detail in previous sections, Figure 24 is an overview, which captures the whole evaluation process in the form of a flowchart.

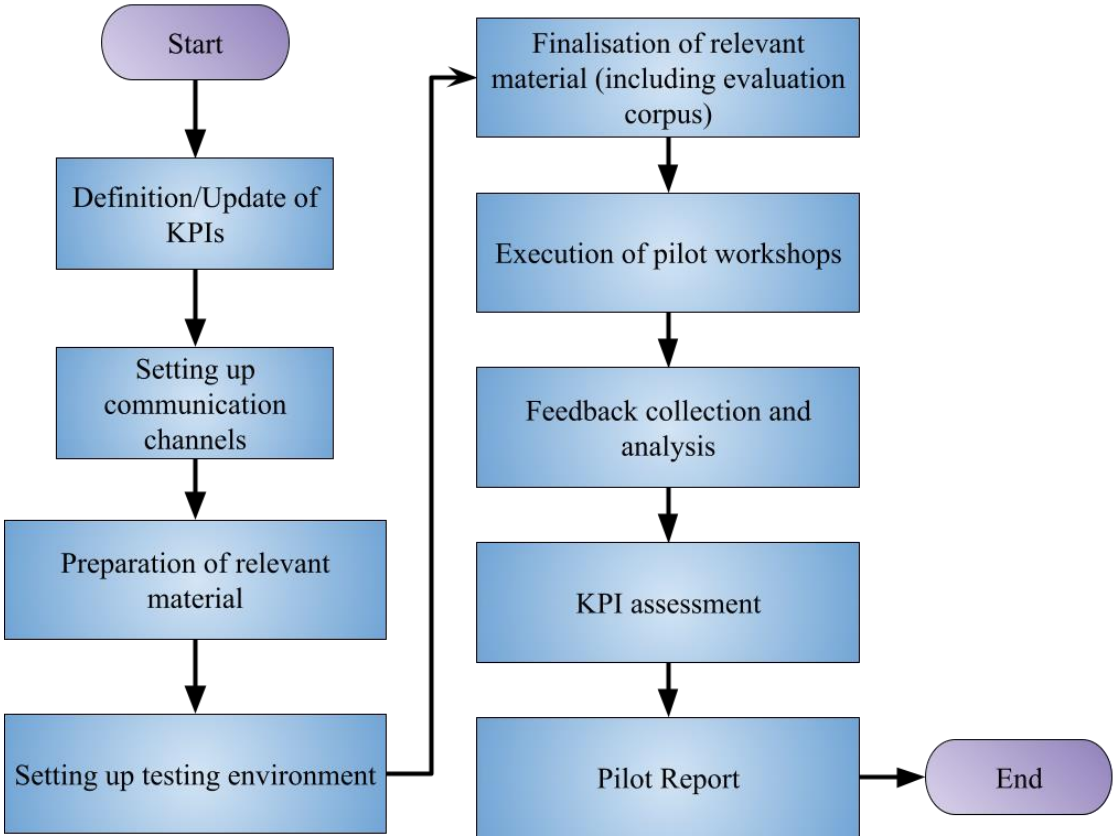


Figure 24: Timeline of first evaluation round

4.4.2. Indicative Actions for Subsequent Evaluation Rounds

This Section outlines the stakeholder engagement efforts, highlighting the participation of organisations during the subsequent piloting phases. It also includes a timeline and some suggestions for its format, based on the findings from the first phase evaluation.

Stakeholder engagement

In Section 7.1 of D6.1, there is a comprehensive overview of the involved entities and their background, as well as their human resources provided for the piloting session. As mentioned, engaging stakeholders is a continuous process that evolves throughout the lifecycle of the piloting phases and in turn of the project. Strong efforts in engaging stakeholders provides enrichment of the various stages of the piloting session. Thus, during subsequent phases of stakeholder engagement, our aim is to target additional participants besides media professionals, such as policymakers and/or researchers. The Task Leader of T6.2 – *"Community engagement and state of the art of fact-checking methodology and indicators"*, namely MALDITA, as well as the rest of the

consortium members will contribute to that initiative by identifying a list of stakeholders who can provide additional users to participate in the pilot workshops, while other dissemination and promotional activities will be planned to attract community interest, such as media publications and campaigns led by EURACTIV through WP7. This broadening of the community aims both:

i. To ensure as much and diverse data as possible in the evaluation process, which will ensure that further development of the platform will effectively meet the users' needs, and
ii. To foster a sense of collaboration and partnership between the project and stakeholders in order to keep the community's interest active.

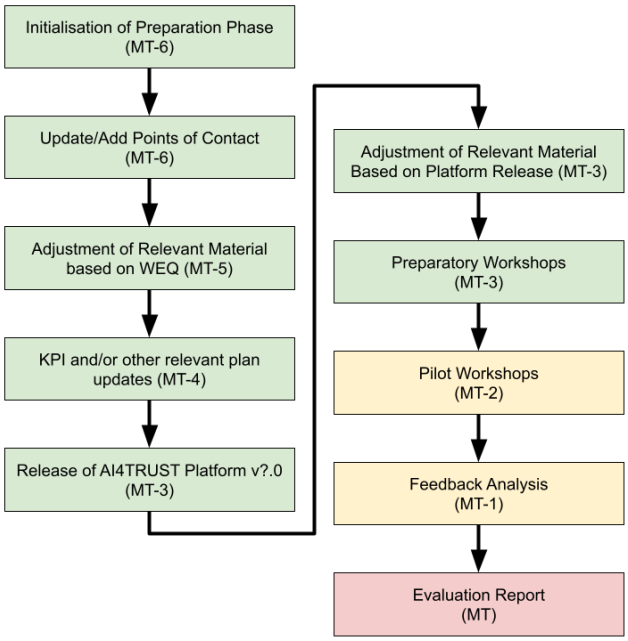


Figure 25: Indicative Timeline for Next Evaluation Rounds

Timeline and Suggestions for improvement

As mentioned in D6.1, during the subsequent piloting phases, the approach will be refined based on insights and outcomes from the first round. The structure established in the initial phase will remain intact, which will ensure continuity and consistency as well as the uniformity of the data in order to make any comparisons and correlations. However, taking into account the feedback we received from the first pilot workshops, minor changes will be included regarding:

i. The brainstorming session, which will either be dropped or focus on other topics;
ii. The duration of each session, devoting more time to the interaction with the AI4TRUST Platform and its tools and the necessary report after each use;
iii. Revisit the total duration of the workshop.

The structure of the subsequent pilot workshops will be presented in the D6.3. The evaluation tools (questionnaires), however, will remain the same. A suggestive timeline that can serve as a template for subsequent evaluation rounds is presented in Figure 25. Within this figure green boxes represent activities that can partake within the preparation phases. Yellow boxes represent pilot

execution phase activities, while the red box represents the post-pilot activities. The initialisation, end and implementation of each activity is referenced considering the project month that the pilot report has to be submitted. Thus MT is the report submission project month, while MT-1 is a month beforehand, and so on. Therefore, this indicative/template timeline can be applied for any subsequent evaluation rounds while at the same time it leaves room for necessary adjustments based on the feedback received during the first evaluation round.

5. Conclusion

This report outlines both qualitative and quantitative findings from the first piloting and evaluation of the AI4TRUST Platform v1 (D5.5) and its associated tools (v1 - D3.1), developed within the AI4TRUST project so far. It mainly focuses on evaluating the *usability*, *functionality*, and overall *user experience*, assessing how these tools fit into the daily workflows of media professionals such as fact-checkers and journalists. Additionally, the evaluation provides insights into the users' perceptions of the platform's *effectiveness*, *trustworthiness*, and the *impact* it has on their work.

A mixed-method evaluation strategy, integrating both qualitative feedback and quantitative data, proved successful. Given the platform's early stage of development (with a focus on hosting individual tools but lacking full content integration), smaller, closed-group evaluations enabled a detailed assessment. These results will directly inform the next phase of development (D3.2, D5.6), which will involve a wider audience to provide more diverse feedback (D6.3). This approach will help refine the platform's functionalities, addressing issues like content completeness, trustworthiness, and the overall user interface. The feedback will also contribute to enhancing the tools' capability to support users' daily work, making them more efficient and reliable.

Mis/disinformation continues to be a significant issue across digital platforms, as highlighted by several ongoing efforts to counteract fake news. The above is quite evident by multiple collaboration, initiatives, and investments made by big tech companies along with programs and organisations in the broader field of media. At the same time, the use of digital tools that aid the work of media professionals turns out to be increasingly important in the pursuit of dealing with artificially generated content, such as deep fakes.

This first evaluation round indicates that initial development of the AI4TRUST platform is on the right track as participants signal their willingness to use it. At the same time, the respective AI-driven tools left a positive mark on the participants due to being reliable, sufficiently developed and fairly easy to use. However, the first round also indicates an initial lack of content (e.g., terms of use or security and transparency terms). The lack of content, along with desired UI improvements and integration of tutorials made participants less inclined to trust and less willing to incorporate the platform in their daily work. A similar pattern follows in the respective tools where a need for more explainability and improvements in terms of providing additional support of the results/verdicts is highlighted as necessary to increase the confidence of participants when using these tools.

Ultimately, as the AI4TRUST Platform and its tools continue to evolve, incorporating these insights and aligning with global efforts to combat mis/disinformation, its development will not only support fact-checkers and journalists but contribute to a broader societal effort to restore trust in digital media.

References

- Bangor, Aaron, Philip T. Kortum, and James T. Miller. 2008. “An Empirical Evaluation of the System Usability Scale.” International Journal of Human-Computer Interaction 24 (6): 574–94. (https://doi.org/10.1080/10447310802205776)
- Brooke J. SUS-A quick and dirty usability scale. Usability evaluation in industry. 1996 Jun 11;189(194):4-7.
- Chakraborty, Hrishikesh, Nicole Solomon, and Kevin J Anstrom. 2023. “A Method to Estimate
- Intra-Cluster Correlation for Clustered Categorical Data.” Communications in Statistics - Theory and Methods 52 (2): 429–44. https://doi.org/10.1080/03610926.2021.1914660)
- Jeng, Judy. 2005. “Usability Assessment of Academic Digital Libraries: **Effectiveness, Efficiency, Satisfaction, and Learnability**.” Libri 55 (2–3). https://doi.org/10.1515/LIBR.2005.96.
- Lewis, James R. 2018. “The System Usability Scale: Past, Present, and Future.” International Journal of Human–Computer Interaction 34 (7): 577–90. (https://doi.org/10.1080/10447318.2018.1455307)
- Schepman A, Rodway P. The General Attitudes towards Artificial Intelligence Scale (GAAIS): Confirmatory validation and associations with personality, corporate distrust, and general trust. International Journal of Human–Computer Interaction. 2023 Aug 9; 39(13): 2724-41.
- Vlachogianni, Prokopia, and Nikolaos Tselios. 2022. “Perceived Usability Evaluation of Educational Technology Using the System Usability Scale (SUS): A Systematic Review.” Journal of Research on Technology in Education 54 (3): 392–409. (https://doi.org/10.1080/15391523.2020.1867938)

Appendix I.

Consent Form

Date: **xx xxx 2024** Place: **xxx**

Title of Project: AI4TRUST - AI-based-technologies for trustworthy solutions against disinformation

The **AI4TRUST project**, funded by the European Union's Horizon 2020 AI to fight disinformation under Grant Agreement No 101070190, aims to combat misinformation and disinformation in the EU by creating a trust-based environment that integrates the automated monitoring of social and news media with advanced AI-based technologies to enhance the work of human fact-checkers.

Research Activity Organiser:

Institute of Informatics & Telecommunications, National Centre for Scientific Research "Demokritos" (NCSR-D), Greece

For any query regarding your personal data, your rights (access, information, edit, deletion, rectification, portability and further privacy issues you may contact the designated Data Protection Officer Mr Alexandros Nousias (alexandros.nousias@iit.demokritos.gr).

Things I need to know

1. **The goal of this Workshop is to**

(a) ***better understand*** how the people involved in fact checking process (media practitioners, fact-checkers, policy makers, journalists) perceive their role and what are their needs and expectations during their everyday work flow in the field of tackling disinformation and misinformation;

(b) ***better understand*** the potential AI4TRUST Platform's users' needs and expectations; and

(c) ***evaluate*** AI4TRUST planned fact-checking and validation activities in order to increase professionals' 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

2. **Voluntary Participation**: Your participation is voluntary, which means that you do not have to take part, if you do not want to and you may leave the group at any time for any reason. All your information will then be destroyed.

3. **Third parties**: AI4TRUST may share the discussion recordings solely with academic institutions within the context of the present research. No communication of your inputs or communication data to the public will take place. Your image data in photo and/or audiovisual format may be communicated to the public. If you do not wish to appear in any photographs or videos which may be viewed outside the project, please tell us.

4. **Storage**: AI4TRUST shall store your data securely for three (3) years at the Institute of Informatics & Telecommunications, NCSR "Demokritos", Greece

After having received explicit information regarding the objectives and format of the Workshop, I will participate in the online workshop on [DATE] 2024 in the context of the AI4Trust project whose specific and limited purposes have been explained to me,

I hereby confirm that:

[1] I have read this consent form and understand the aims of the Workshop __Yes __ No

[2] I agree and provide specific and informed consent to participate in the research conducted by the AI4Trust project __Yes __ No

[3] I agree and provide specific and informed consent for my contributions to the AI4Trust project workshop to be recorded for internal use __Yes __ No

[4] I agree and provide explicit and informed consent for the processing of my personal data (namely name, professional, etc) solely for communication purposes in the context of the research activity __Yes __ No

[5] I agree and provide explicit and informed consent for the processing and sharing of my image data in photo and/or audiovisual format solely for communication and dissemination purposes among the AI4TRUST research team partners.

5 I agree and provide explicit and informed consent for such data to be viewable by interest parties outside the AI4TRUST consortium __Yes __ No

[6] I agree and provide explicit and informed consent for the processing and sharing of my named feedback in via the AI4TRUST repository __Yes __ No

Name and email of the participant:	**Signature:**

Appendix II.

Platform Evaluation Questionnaire

Dear participant ,

Thank you for your will to participate in the following survey.

It takes approximately 7-10 minutes and it aims to investigate your experience in relation to your engagement with the AI4Trust platform after your recent participation in the AI4Trust Workshop (30 July 2024), which was organized by the National Centre for Scientific Research 'Demokritos' (NCSR-D).

The survey is anonymous and your participation remains voluntary, which means that you are free to quit any time you wish.

Survey's results will be analysed for research purposes by the NCSR-D in the framework of the AI4Trust project (H2020 AI to fight disinformation under Grant Agreement No 101070190) and particularly in the context of improving the services provided by the platform.

We greatly appreciate your contribution!

Generic/Demographics Info

1. Gender

_ Female - Male _ Prefer not to say _ Other

2. Age Group

_ <25 _ 26-30 _ 31-40 _ 41+

3. Profession

_ Fact-checker _ Journalist _ Editor _ media literacy practitioner _ computer engineer/technical specialist _ disinformation and public policy expert _ video producer _ digital product manager _ media expert _ project manager/coordinator _ other media professional

4. Years of employment in this profession

Expectations/Experience

5. What are your expectations in the field of tackling disinformation and misinformation by using the AI4Trust Platform?
6. How does the AI4Trust platform support your everyday work flow?

Usability of the Platform (SUS)

7. I think I would like to use the AI4TRUST Platform (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
8. I think the AI4Trust Platform was easy to use (image/video, audio, text) (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
9. I found the various functions of the AI4Trust platform well integrated/implemented

(1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)

10. I would imagine that most people would learn to use the AI4TRUST Platform very quickly
    (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
11. I felt very confident using the AI4Trust Platform
    (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
12. I found the AI4Trust Platform unnecessary complex
    (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
13. I think that I would need further technical support to be able to use the AI4Trust Platform
    (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
14. I thought there was too much inconsistency in the AI4Trust Platform
    (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
15. I found the AI4Trust Platform very cumbersome to use
    (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
16. I needed to learn a lot for things before I could use the AI4TRUST Platform
    (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)

Platform's Functionality

17. Which of the following functions provided by the AI4Trust Platform are covered?

_ user friendly navigation _ appropriate content submission _ review results analysis _ feedback submission

18. Have you observed any missing functions from the AI4Trust Platform?
19. I am confident that the following requirements were covered by the AI4Trust platform

_ Reliability analysis of information sources _ Evidence-based textual inoculation _ Integration of fact-checking activities in the daily journalistic process _ Infodemic trends for specific public interest issues

20. The AI4Trust Platform consistently delivers results that are correct and reliable.
    (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
21. I rarely experience delays or lag when using the AI4Trust Platform
    (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
22. Do you have any suggestions about possible functionalities that can be added to the AI4Trust Platform?

User's Satisfaction (EESL Questionnaire)

23. Please rate the ease of use of the AI4Trust platform
    (1 - Easy, 2: - Somewhat Easy, 3. - Neutral, 4. - Somewhat Difficult, 5 - Difficult)
24. Is the site easy to navigate?
    (1 - Easy, 2: - Somewhat Easy, 3. - Neutral, 4. - Somewhat Difficult, 5 - Difficult)
25. I am able to quickly identify how the platform fits into my workflow.
    (1 - Strongly Disagree, 5 - Strongly Agree)
26. What do you think about the organization of information on the AI4Trust platform?
    (1 - Clear, 2: - Somewhat Clear, 3. - Neutral, 4. - Somewhat Unclear, 5 - Unclear)
27. Do you expect that a click on the AI4Trust platform can lead you to a correct answer?
    (Yes, No, I am not sure)

28. What is your overall reaction to the AI4Trust platform? (1 - Satisfied, 2: - Somewhat Satisfied, 3. - Neutral, 4. - Somewhat Unsatisfied, 5 - Unsatisfied)
29. Did you notice any mistakes in the evaluation of instances during the use of the AI4Trust platform? (e.g. I have annotated 3 instances of hate speech in an article, but the system submitted 4 of which it found the 2 of them, so 50%) Yes, No
    If yes, how many instances in comparison with the total number? (e.g. 2 out of 4)
30. Can you recover from mistakes easily? (1 - Easy, 2: - Somewhat Easy, 3. - Neutral, 4. - Somewhat Uneasy, 5 - Uneasy)
31. How many mistakes (relevant retrieved instances) could be recalled? (e.g. 2 out of 4)
32. Is the site attractive? (1 - Attractive, 2: - Somewhat Attractive, 3. - Neutral, 4. - Somewhat Unattractive, 5 - Unattractive)
33. What do you think about the terminology used in the site?
    (1 - Clear, 2: - Somewhat Clear, 3. - Neutral, 4. - Somewhat Unclear, 5 - Unclear)
34. Do you feel lost while using the AI4Trust Platform? (Yes, No, I am not sure)
35. I feel confident that my data is secure when using the AI4Trust Platform
    (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
36. The system provides accurate results even under varying conditions.
    (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
37. Do you have suggestions on how to improve the overall usability of the AI4Trust platform?

Thank you for your contribution!

****

Toolbox Evaluation Questionnaire

Dear participant ,

Thank you for your will to participate in the following survey.

It takes approximately 7-10 minutes and it aims to investigate your experience in relation to your engagement with the AI4Trust Toolbox (validation tools) after your recent participation in the AI4Trust Workshop (30 July 2024), which was organized by the National Centre for Scientific Research 'Demokritos' (NCSR-D).

The survey is anonymous and your participation remains voluntary, which means that you are free to quit any time you wish.

Survey's results will be analysed for research purposes by the NCSR-D in the framework of the AI4Trust project (H2020 AI to fight disinformation under Grant Agreement No 101070190) and particularly in the context of improving the services provided by the platform.

We greatly appreciate your contribution!

Generic/Demographics Info

1. Gender

_ Female _ Male _ Prefer not to say _ Other

2. Age Group

_ <25 _ 26-30 _ 31-40 _ 41+

3. Profession

_ Fact-checker _ Journalist _ Editor _ media literacy practitioner _ computer engineer/technical specialist _ disinformation and public policy expert _ video producer _ digital product manager _ media expert _ project manager/coordinator _ other media professional

4. Years of employment in this profession

Expectations/Experience

5. What are your expectations in the field of tackling disinformation and misinformation by using AI4Trust tools?
6. How do you expect AI4TRUST tools to support your everyday work flow?

Usability of the Toolbox (SUS)

(each question addresses each AI4Trust tool)

7. I think I would like to use AI4Trust tools (image/video, audio, text) (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
8. I think the AI4Trust tools were easy to use (image/video, audio, text) (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
9. I found the various functions of the AI4Trust Tools well integrated/implemented (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
10. I would imagine that most people would learn to use the AI4Trust Tools very quickly (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
11. I felt very confident using the AI4Trust Tools (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
12. I found the AI4Trust Tools unnecessarily complex (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
13. I think that I would need further technical support to be able to use the AI4Trust Tools (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
14. I thought there was too much inconsistency in the AI4Trust Toolbox (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
15. I found the AI4Trust Tools very cumbersome to use (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
16. I needed to learn a lot of things before I could use the AI4Trust Tools (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)

Toolbox's Functionality

17. Which of the following functions provided by the AI4Trust toolbox (image/video, audio, text) are covered?

_ user friendly navigation _ appropriate content submission _ review results analysis _ feedback submission

18. Are there any other functions that you would like to be provided by each tool (image/video, audio, text)?
19. The tool consistently delivers results that are correct and reliable (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
20. I rarely experience delays or lag when using the tool / I am satisfied with the response time (1 - Strongly Disagree, 2: Disagree, 3. Neutral, 4. Agree, 5 - Strongly Agree)
21. Do you have any suggestions about possible functionalities that could improve the tools?

User's Satisfaction (EESL Questionnaire)

22. Please rate the ease of use of this AI4Trust Tool (image/video, audio, text) (1 - Easy, 2: - Somewhat Easy, 3. - Neutral, 4. - Somewhat Difficult, 5 - Difficult)
23. Is the Toolbox easy to navigate? (1 - Easy, 2: - Somewhat Easy, 3. - Neutral, 4. - Somewhat Difficult, 5 - Difficult)
24. I am able to quickly identify how the Toolbox fits into my workflow. (1 - Strongly Disagree, 5 - Strongly Agree)
25. What do you think about the organisation of information on the AI4Trust Tools? (1 - Clear, 2: - Somewhat Clear, 3. - Neutral, 4. - Somewhat Unclear, 5 - Unclear)
26. Do you expect that a click on the AI4Trust Toolbox can lead you to a correct answer? (Yes, No, I am not sure)
27. What is your overall reaction to the AI4Trust Tools? (1 - Satisfied, 2: - Somewhat Satisfied, 3. - Neutral, 4. - Somewhat Unsatisfied, 5 - Unsatisfied)
28. Did you notice any mistakes in the evaluation of instances during the use of the tool? *(e.g. I have annotated* 3 instances of hate speech in an article, but the system submitted 4 of which it found the 2 of them, so 50%) Yes, No
    If yes, how many instances in comparison with the total number? *(e.g. 2 out of 4)*
29. Can you recover from mistakes easily? (1 - Easy, 2: - Somewhat Easy, 3. - Neutral, 4. - Somewhat Uneasy, 5 - Uneasy)
30. How many mistakes (relevant retrieved instances) could be recalled? *(e.g. 2 out of 4)*
31. Are the AI4Trust Tools visually attractive or appealing? (1 - Attractive, 2: - Somewhat Attractive, 3. - Neutral, 4. - Somewhat Unattractive, 5 - Unattractive)
32. What do you think about the terminology used in the Toolbox? (1 - Clear, 2: - Somewhat Clear, 3. - Neutral, 4. - Somewhat Unclear, 5 - Unclear)
33. Do you feel lost while using the AI4Trust Toolbox? (Yes, No, I am not sure)
34. I feel confident that my data is secure when using the AI4Trust Toolbox (1 - Strongly Disagree, 2 - Disagree, 3 - Neutral, 4 - Agree, 5 - Strongly Agree)
35. The system provides accurate results even under varying conditions (1 - Strongly Disagree, 2 - Disagree, 3 - Neutral, 4 - Agree, 5 - Strongly Agree)
36. Do you have suggestions on how to improve the overall usability of the tools?

Thank you for your contribution!

****

Workshop Evaluation

Dear participant ,

Thank you for your will to participate in the following survey.

It takes approximately 5 minutes and it aims to investigate your experience in relation to your recent participation in the *AI4Trust preparatory Workshop* (30 July 2024), which was organized by the National Centre for Scientific Research 'Demokritos' (NCSR-D).

The survey is anonymous and your participation remains voluntary, which means that you are free to quit any time you wish.

Survey's results will be analysed for research purposes by the NCSR-D in the framework of the AI4Trust project (H2020 AI to fight disinformation under Grant Agreement No 101070190) and particularly in the context of improving its format in view of the pilot workshops in autumn 2024.

We greatly appreciate your contribution!

Generic/Demographics Info

1. Gender
2. Age Group

Workshop's Evaluation

3. How would you evaluate the workshop about the content? (1: I didn't like at all, 2: I didn't like, 3: I don't know, 4: I liked, 5: I liked very much)
4. How would you evaluate the workshop about the structure? (1: I didn't like at all, 2: I didn't like, 3: I don't know, 4: I liked, 5: I liked very much)
5. How would you evaluate the training about the duration? (1: I didn't like at all, 2: I didn't like, 3: I don't know, 4: I liked, 5: I liked very much)
6. How would you evaluate the following parts of the Workshop?
   [6.1] Introduction - Know each other (1: I didn't like at all, 2: I didn't like, 3: I don't know, 4: I liked, 5: I liked very much)
   [6.2] Project's presentation (1: I didn't like at all, 2: I didn't like, 3: I don't know, 4: I liked, 5: I liked very much)
   [6.3] Tackling disinformation & misinformation process – needs and expectations (1: I didn't like at all, 2: I didn't like, 3: I don't know, 4: I liked, 5: I liked very much)
   [6.4] AI4Trust Platform (Overview & discussion) (1: I didn't like at all, 2: I didn't like, 3: I don't know, 4: I liked, 5: I liked very much)
   [6.5] AI4Trust Toolbox (Overview & discussion) (1: I didn't like at all, 2: I didn't like, 3: I don't know, 4: I liked, 5: I liked very much)
   [6.6] Reflection-Conclusions (1: I didn't like at all, 2: I didn't like, 3: I don't know, 4: I liked, 5: I liked very much)
7. How much do you agree/disagree with the following?
   [7.1.] During my participation in the workshop I had time to express my thoughts (1: I fully

disagree, 2: I disagree, 3: I don't know, 4: I agree, 5: I fully agree)

[7.2.] During my participation in the Workshop I was able to interact with my colleagues and exchange thoughts and viewpoints (1: I fully disagree, 2: I disagree, 3: I don't know, 4: I agree, 5: I fully agree)

Expectations/Experience

8. What were your expectations from the Workshop?

_ obligation (part of my contribution to the project) _ interest for the topic _ personal need for gaining new knowledge/learning new tools _ need to get informed and transfer knowledge to colleagues _ all the above _ Other

9. Did the workshop manage to meet your expectations? (_ Yes _ No _ I am not sure)
10. The workshop gave me the confidence to better understand the use of the AI4Trust platform and promote it my colleagues (1: Strongly disagree, 2: Disagree, 3: I am not sure, 4: Agree, 5: Strongly agree)
11. The organizers/facilitators were well prepared and helped me understand the underlying concepts (1: Strongly disagree, 2: Disagree, 3: I am not sure, 4: Agree, 5: Strongly agree)
12. Please, feel free to share with us any thoughts/feelings (what did we miss? What can we do differently in the pilot workshop?)

Thank you for your contribution!

Appendix III.

Moderator's Report

Session Details	● Date/Time ○ xxxx, 2024 xxx CET ● Moderator(s) ○ Athanasios Davvetas ○ Dora Katsamori ● No. of Participants ○ xx
Preparatory Notes	● Test data prepared ● Briefed about tasks and evaluation scenarios **Issues Identified** ● Issue 1: ...
Testing Session Observations	**Platform (Evaluation Scenario 2)** ● Tasks performed Access homepage Access login Perform login Access content analysis page ● Feedback Positive Neutral Negative ● Lead time (and User engagement) ○ Less than xxx min. ● Technical Issues ○ xxx ● Suggestions ○ User xxx **Video Tools** ● Analyse news items (Evaluation Scenario 1+3) ○ Tasks performed ● Access content analysis page ● Select appropriate modality ● Submit content ● Compare manual assess of veracity to automatic analysis ○ Feedback Positive Neutral Negative

	○ Lead time: ■ xx min, xx min, xx min even ○ Technical Issues ■ xxx ○ Suggestions ■ xx **Image Tools** ● Analyse news items (Evaluation Scenario 1+3) ○ Tasks performed ● Access content analysis page ● Select appropriate modality ● Submit content ● Compare manual assess of veracity to automatic analysis ○ Feedback Positive Neutral Negative ○ Lead time: ■ Under xxx min ○ Technical Issues ■ xxx ○ Suggestions ■ xx **Audio Tools** ● Analyse news items (Evaluation Scenario 1+3) ○ Tasks performed ● Access content analysis page ● Select appropriate modality ● Submit content ● Compare manual assess of veracity to automatic analysis ○ Feedback Positive Neutral Negative ○ Lead time: ■ 2 xxx, 5 xxx, 15 xxx min ○ Technical Issues ■ xx ○ Suggestions ■ xxx

	Text Tools ● Analyse news items (Evaluation Scenario 1+3) ○ Tasks performed ● Access content analysis page ● Select appropriate modality ● Submit content ● Compare manual assess of veracity to automatic analysis ○ Feedback Positive Neutral Negative ○ Lead time: ■ Under xxx min ○ Technical Issues ■ xx ○ Suggestions ■ xx
Overall Impressions	Strengths: Weaknesses: Potential Improvements:
General Notes	● xxxxx ● Workflow (Journalists): ○ Disinformation xxx ● Workflow (Fact-checking): ○ xxxx