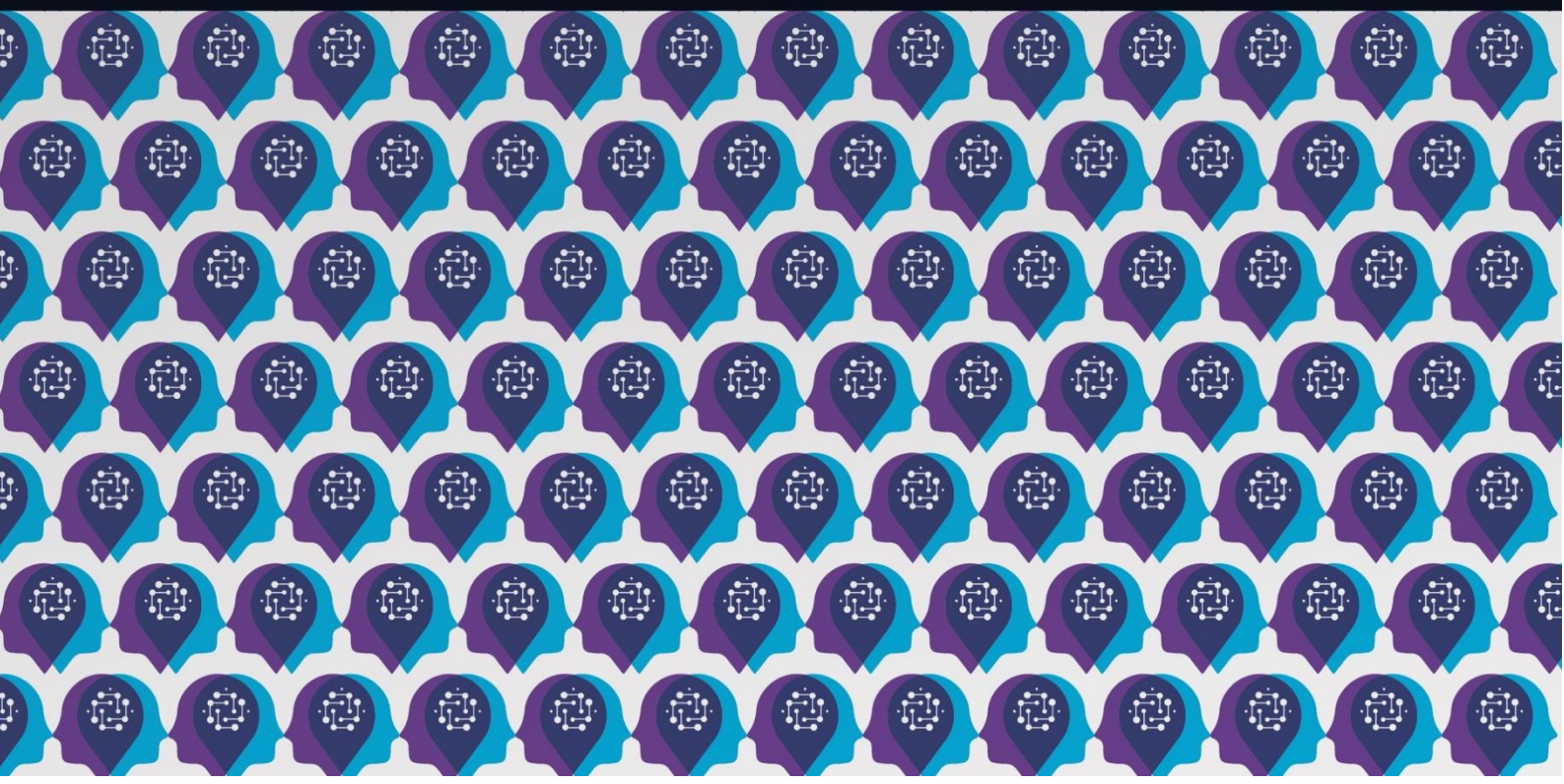




# AI4Debunk

D8.2 First report on the development of  
multimodal adaptable AI modules  
October 2025





Grant Agreement No.: 101135757  
 Call: HORIZON-CL4-2023-HUMAN-01-CNECT  
 Topic: HORIZON-CL4-2023-HUMAN-01-05  
 Type of action: HORIZON Innovation Actions

## D8.2 FIRST REPORT ON THE DEVELOPMENT OF MULTIMODAL ADAPTABLE AI MODULES

<b>Project Acronym</b>	AI4Debunk
<b>Project Number</b>	101135757
<b>Project Full Title</b>	Participative Assistive AI-powered Tools for Supporting Trustworthy Online Activity of Citizens and Debunking Disinformation
<b>Work package</b>	WP 8
<b>Task</b>	Task 2
<b>Due date</b>	31/10/2025
<b>Submission date</b>	24/10/2025
<b>Deliverable lead</b>	Partner UMONS
<b>Version</b>	1
<b>Authors</b>	Kevin El Haddad (Partner UMONS)
<b>Contributors</b>	Axel Brando (Partner BSC), Alvaro Parafita (Partner BSC)
<b>Reviewers</b>	Chun Fei Lung (Partner HU)
<b>Abstract</b>	The proliferation of multimodal news, which combines text, images, and audio, necessitates advanced systems capable of analyzing information across different formats to combat sophisticated disinformation. This report details the foundational development of the AI4DEBUNK decision support framework. We present three core components: (1) a multimodal information extraction system for audio and visual entity identification, designed to populate a central Knowledge Graph; (2) a cross-modal coherence analysis module to detect Out-of-Context (OOC) disinformation by verifying the semantic consistency between images and their captions; and (3) the initial version

of the AI4DEBUNK Platform. This platform is a modular, extensible architecture that orchestrates various analysis modules—including deepfake detection, claim similarity, and coherence checking—to process multimodal news items and compute an aggregate "DisinfoScore." While these components provide a strong foundation, future work will focus on the deep integration of the Knowledge Graph to provide the rich contextual data required to unify the system and enhance its adaptive intelligence.

---

<b>Keywords</b>	Multimodal Disinformation, Decision Support System, Cross-Modal Analysis, Coherence Analysis, Out-of-Context (OOC), Information Extraction, Knowledge Graph, Modular Architecture, AI4DEBUNK, Fake News Detection
-----------------	---

---

## DOCUMENT DISSEMINATION LEVEL

**Dissemination level**

---

<b>X</b>	PU - Public
	SEN - Sensitive

---

## DOCUMENT REVISION HISTORY

Version	Date	Description of change	List of contributor(s)
0.1	24/10/2025	First draft	UMONS
0.2	25/10/2025	Internal Quality Assessment Review	HU
0.5	29/10/2025	Implementation of suggestions	UMONS
1	30/10/2025	Final version	UMONS

## STATEMENT ON MAINSTREAMING GENDER

The AI4Debunk consortium is committed to including gender and intersectionality as a transversal aspect in the project’s activities. In line with EU guidelines and objectives, all partners – including the authors of this deliverable – recognise the importance of advancing gender analysis and sex-disaggregated data collection in the development of scientific research. Therefore, we commit to paying particular attention to including, monitoring, and periodically evaluating the participation

of different genders in all activities developed within the project, including workshops, webinars and events but also surveys, interviews and research, in general. While applying a non-binary approach to data collection and promoting the participation of all genders in the activities, the partners will periodically reflect and inform about the limitations of their approach. Through an iterative learning process, they commit to plan and implement strategies that maximise the inclusion of more and more intersectional perspectives in their activities.

## DISCLAIMER

The AI4Debunk project has received funding from the European Union's Horizon Europe Programme under the Grant Agreement No. 101135757.

Views and opinions expressed are, however, those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

## COPYRIGHT NOTICE

### © AI4Debunk - All rights reserved

No part of this publication may be translated, reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the written permission of the publisher or provided the source is acknowledged.

How to cite this report: Kevin El Haddad (2025). AI4Debunk D8.2: First Report on the development of multimodal adaptable AI modules. <https://ai4debunk.eu/wp-content/uploads/2025/11/AI4Debunk-Deliverable-8.2.pdf>

The AI4Debunk consortium is the following:

<b>Participant number</b>	<b>Participant organisation name</b>	<b>Short name</b>	<b>Country</b>
1	LATVIJAS UNIVERSITATE	UL	LV
2	FREE MEDIA BULGARIA SRL	EUalive	BE
3	PILOT4DEV	P4D	BE
4	INTERNEWS UKRAINE	IUA	UA
5	CONSIGLIO NAZIONALE DELLE RICERCHE	CNR-IRPPS	IT
6	UNIVERSITA DEGLI STUDI DI FIRENZE	MICC/UNIFI	IT
6.1	CONSORZIO NAZIONALE INTERUNIVERSITARIO PER LE TELECOMUNICAZIONI	CNIT	IT
7	BARCELONA SUPERCOMPUTING CENTER CENTRO NACIONAL DE SUPERCOMPUTACION	BSC	ES
8	DOTSOFT OLOKLIROMENES EFARMOGES DIADIKTIOY KAI VASEON DEDOMENON AE	DOTSOFT	EL
9	UNIVERSITE DE MONS	UMONS	BE
10	UNIVERSITY OF GALWAY	UoG	IE
11	STICHTING HOGESCHOOL UTRECHT	HU	NL
12	STICHTING INNOVATIVE POWER	IP	NL
13	F6S NETWORK IRELAND LIMITED	F6S	IE

---

## TABLE OF CONTENTS

---

<b>1</b>	<b>INTRODUCTION .....</b>	<b>10</b>
1.1	OBJECTIVES.....	10
1.2	EXPECTED OUTCOME .....	10
<b>2</b>	<b>MULTIMODAL INFORMATION EXTRACTION SYSTEM FOR KNOWLEDGE GRAPH INGESTION .....</b>	<b>11</b>
2.1	AUDIO MODALITY: SPEAKER IDENTIFICATION.....	11
2.2	IMAGE MODALITY: FACE RECOGNITION .....	11
2.3	CURRENT STATUS AND FUTURE WORK .....	11
<b>3</b>	<b>CROSS-MODAL COHERENCE ANALYSIS .....</b>	<b>11</b>
3.1	TECHNICAL IMPLEMENTATION .....	12
3.2	LIMITATIONS AND FUTURE ENHANCEMENTS .....	12
<b>4</b>	<b>THE AI4DEBUNK PLATFORM .....</b>	<b>12</b>
4.1	IMPLEMENTED MODULES .....	13
4.2	DISINFOSCORE CALCULATION AND FUTURE DIRECTIONS .....	14
<b>5</b>	<b>CONCLUSION.....</b>	<b>14</b>

---

## LIST OF FIGURES

---

FIGURE 1 AI4DEBUNK PLATFORM ..... 13

---

## ABBREVIATIONS

---

WP	Work Package
LLM	Large Language Model
CLIP	Contrastive Language-Image Pre-training
ViT	Vision Transformer

---

---

## EXECUTIVE SUMMARY

---

This report details the foundational development of multimodal analysis systems for the AI4Debunk project, designed to address sophisticated, multi-format disinformation. The current work establishes three core pillars of a comprehensive decision support solution. A multimodal information extraction system uses facial and audio recognition to identify individuals in media for ingestion into a Knowledge Graph. A second component provides cross-modal coherence analysis, leveraging a contrastive language-image model to detect out-of-context disinformation by checking image-text consistency. These tools are integrated into the central AI4Debunk Platform, which serves as a modular and extensible orchestration layer. This platform dynamically activates specialized modules based on an incoming news item's specific modalities. It then aggregates their analyses to compute a single "Disinfoscore" that quantifies the assessed level of disinformation. While these systems provide a robust foundation, current limitations include the separate operation of identification systems and a lack of deep contextual awareness. Therefore, the next phase of development will prioritize the deep integration of the project's Knowledge Graph. This strategic goal is intended to provide the rich contextual data necessary to unify the modules and empower the entire platform with greater adaptive intelligence.

---

## 1 INTRODUCTION

---

The contemporary digital news landscape is characterized by its inherently multimodal nature, where text, imagery, and audio collectively contribute to the narrative. Consequently, analytical approaches that are restricted to a single modality, such as text, are insufficient for comprehensive disinformation detection. This necessitates the development of advanced systems capable of synthesizing and reasoning over information distributed across various modalities.

While existing state-of-the-art systems can process multimodal inputs, they are often engineered for narrow, task-specific applications (e.g., classification), which limits their operational flexibility. Conversely, while Multimodal Large Language Models (LLMs) offer greater versatility through prompt engineering, their significant computational overhead and deployment complexity present substantial barriers to practical application in real-world scenarios.

---

### 1.1 OBJECTIVES

---

The primary objective of this work package is to develop a suite of multimodal systems to serve as a decision support framework for disinformation analysis. This report details the initial implementation of the three core components developed to achieve this goal:

**Multimodal Information Extraction:** Systems designed to extract and normalize key information from various data modalities, primarily for ingestion into the project's knowledge graph.

**Cross-Modal Coherence Analysis:** A system engineered to verify the semantic consistency between an image and its corresponding textual caption. This component can function as a standalone tool or as an integrated module within the main platform.

**AI4Debunk Modular Platform:** A flexible, modular platform designed to orchestrate the analysis of multimodal news content. Its architecture allows for the seamless integration of new analytical modules as state-of-the-art technologies evolve.

---

### 1.2 EXPECTED OUTCOME

---

This report details the design and initial implementation of the aforementioned systems. These components constitute the foundational work for the comprehensive AI4Debunk decision support solution, with further development and integration planned for Task 9.2.

---

## 2 MULTIMODAL INFORMATION EXTRACTION SYSTEM FOR KNOWLEDGE GRAPH INGESTION

---

The core objective of this component is to perform robust, cross-modal entity identification to facilitate the linking of individuals across disparate news sources within the project's knowledge graph. Accurate multimodal recognition is critical for reliable identity resolution. This implementation focuses on two primary modalities: audio and facial imagery.

---

### 2.1 AUDIO MODALITY: SPEAKER IDENTIFICATION

---

Speaker diarization and recognition functionalities were implemented utilizing state-of-the-art pipelines from the **SpeechBrain** and **PyAnnote** toolkits. These frameworks were selected for their modular architecture and access to pretrained models that demonstrate high accuracy in segmenting and attributing identities within multi-speaker audio streams.

---

### 2.2 IMAGE MODALITY: FACE RECOGNITION

---

Facial recognition is performed using the **ArcFace** model, chosen for its demonstrated high performance on standard benchmarks (e.g., Labeled Faces in the Wild) and its computational efficiency. The model's use of an additive angular margin loss function enhances the discriminative power of its feature embeddings, enabling precise identity matching under real-world conditions with constrained computational resources.

---

### 2.3 CURRENT STATUS AND FUTURE WORK

---

Currently, the audio and facial recognition systems operate independently. Future work will focus on developing an integrated, multimodal identification system capable of fusing information across these modalities for enhanced accuracy and robustness.

---

## 3 CROSS-MODAL COHERENCE ANALYSIS

---

This system is designed to assess the semantic coherence between an image and its accompanying textual caption. Its primary function is to detect **Out-of-Context (OOC) disinformation**, a prevalent manipulation technique where genuine images or texts are repurposed to misrepresent or fabricate a narrative, thereby exploiting their perceived authenticity to mislead audiences.

---

### 3.1 TECHNICAL IMPLEMENTATION

---

The methodology leverages the **Contrastive Language-Image Pre-training (CLIP)** model. It employs CLIP's distinct encoders—a Transformer for text and a Vision Transformer (ViT) for images—to project each modality into a shared, high-dimensional embedding space. The cosine distance between the resulting text and image vectors is then calculated. If this distance exceeds a predefined threshold, the image-text pair is flagged as semantically incoherent.

---

### 3.2 LIMITATIONS AND FUTURE ENHANCEMENTS

---

A primary limitation is that while CLIP excels at encoding high-level semantic content, preliminary experiments suggest it may not adequately capture complex contextual details, such as verifying the temporal alignment between an event described in the text and the date an image was created.

To address this, the following steps are planned:

- **Benchmarking:** Ongoing work in Task 8.2, to be continued in Task 9.2, involves the development of a dedicated benchmark to quantitatively evaluate the system and validate this hypothesis.
- **Knowledge Graph Integration:** Future enhancements will focus on integrating contextual information from the WP6 Knowledge Graph to enrich the input data and improve coherence assessment.
- **Advanced Model Development:** A more ambitious research direction involves fine-tuning an existing model or training a new model from scratch, specifically designed to incorporate a wider range of contextual cues beyond basic semantic similarity.

---

## 4 THE AI4DEBUNK PLATFORM

---

An initial version of a modular, multimodal platform for processing and analyzing news content has been implemented, as displayed in **Figure 1**. The platform is architecturally composed of two primary elements:

- A collection of **specialized, independent modules**, each designed to process a specific data modality or perform a distinct analytical function.
- An **orchestration layer** that dynamically activates relevant modules based on the modalities present in an incoming news item. This layer also aggregates the outputs from activated modules to compute a composite **Disinfoscore**, quantifying the assessed level of disinformation.



The platform's architecture is explicitly designed for modularity and extensibility. New modules can be integrated, or existing ones modified, with minimal impact on the overall system, provided they adhere to the platform's defined API specifications (e.g., input/output formats, class methods). In the current implementation, module activation is predicated solely on the modalities present in the input data.

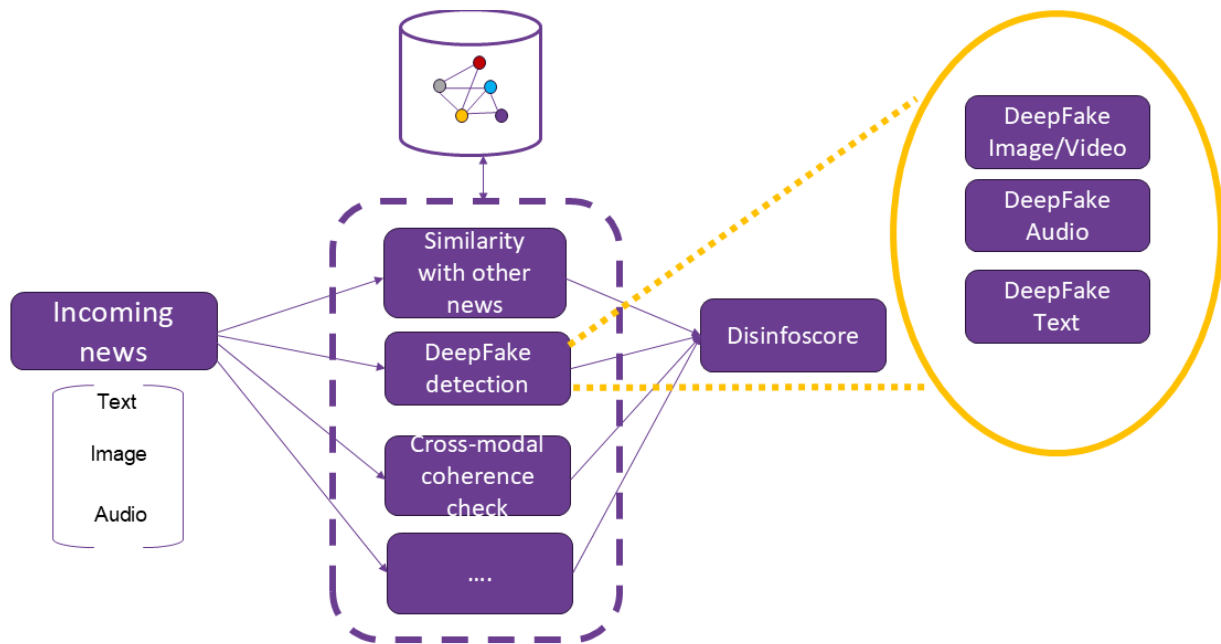


FIGURE 1 AI4DEBUNK PLATFORM

## 4.1 IMPLEMENTED MODULES

The current version of the platform, illustrated with orange-highlighted components in **Figure 1**, integrates the following analytical modules:

1. **Claim Similarity Estimation:** Compares an input news item against a knowledge base to identify supporting or opposing claims (detailed in D8.1).
2. **Deepfake Detection Suite:** A collection of tools for detecting manipulated media across modalities (detailed in D8.1), including:
  - **Text-Based DeepFake Detection:** Analyzes textual content for indicators of artificial generation or manipulation.
  - **Image/Video DeepFake Detection:** Examines visual media for signs of deepfakes, providing outputs such as localization heatmaps and authenticity scores on a frame-by-frame basis for video.
  - **Audio DeepFake Detection:** Processes audio signals to detect synthetic or cloned voices.

3. **Cross-Modal Coherence Analysis:** The module described in Section 3, which assesses the semantic alignment between images and their captions.

---

## 4.2 DISINFOSCORE CALCULATION AND FUTURE DIRECTIONS

---

Modules can operate as standalone tools or contribute to the platform's aggregate analysis. The Disinfoscore is currently calculated as a weighted average of the normalized outputs from each activated module. A detailed description of this mechanism is provided in deliverable D8.4. The platform's modularity enables significant adaptability. However, the performance of individual modules on novel data or in new contexts can be limited. To address this, future work, continuing from WP8 into WP9, will focus on deeper integration with the project's **knowledge graph**. By providing rich, structured contextual information and semantic relationships, the graph can enhance module performance and enable dynamic adaptation to new entities, terminology, and evolving events.

---

## 5 CONCLUSION

---

This report has detailed the design and initial implementation of the core multimodal components for the AI4Debunk project. We have successfully developed foundational capabilities for **multimodal information extraction**, established a mechanism for **cross-modal coherence analysis** to detect Out-of-Context disinformation, and engineered the first version of the **AI4Debunk platform**.

The primary contribution of this work is the establishment of a **modular and extensible architecture**. This platform provides the flexibility to integrate, update, and deploy diverse analytical tools in response to the rapidly evolving landscape of digital disinformation. The systems for entity identification and coherence analysis serve as both powerful standalone tools and integral initial modules for this overarching framework.

Looking ahead, the next phase of development, to be conducted in **Task 9.2**, will focus on addressing current limitations and enhancing system intelligence. The central strategic priority is the deep **integration of the Knowledge Graph**. By leveraging the rich, structured context provided by the graph, we aim to:

- **Enhance module performance** by supplying the contextual information that models like CLIP currently lack.
- **Unify the separate audio and facial identification systems** into a cohesive, truly multimodal entity recognition pipeline.
- **Improve the platform's adaptability** to new events, entities, and disinformation narratives.

The work completed in this task establishes a robust foundation. Future efforts will build upon this groundwork to create a sophisticated, context-aware decision support system for effectively combating multimodal disinformation.

---

---

## Review Sheet of Deliverable Report

### D8.2 First report on the development of multimodal adaptable AI modules

---

---

<b>Editor(s):</b>	Kevin El Haddad
<b>Responsible Partner:</b>	University of Mons (UMONS)
<b>Status-Version:</b>	Draft - v0.1
<b>Date:</b>	24/10/2025
<b>Distribution level (CO, PU):</b>	Public
<b>Reviewer (Name/Organization)</b>	Chun Fei Lung (HU)
<b>Review date</b>	29/10/2025

*Disclaimer: This assessment reflects only the author's views and the European Commission is not responsible for any use that may be made of the information contained therein"*

Mark with X the corresponding column:

<b>Y= yes</b>	<b>N= no</b>	<b>N = not applicable</b>
---------------	--------------	---------------------------

ELEMENT TO REVIEW	Y	N	NA	COMMENTS
<b>FORMAT: Does the document ... ?</b>				
...include editors, deliverable name, version number, dissemination level, date, and status?	X			
...contain a license (in case of public deliverables)?	X			
...include the names of contributors and reviewers?	X			
...has a version table consistent with the document's revision?	X			
... contain an updated table of contents?	X			
... contain a list of figures consistent with the document's content?	X			
... contain a list of tables consistent with the document's content?			X	
... contain a list of terms and abbreviations?	X			
... contain an Executive Summary?	X			
... contain a Conclusions section?	X			
... contain a List of References (Bibliography) in the adequate format, if relevant?			X	
... use the fonts and sections defined in the official template?	X			
... use correct spelling and grammar?	X			
... conform to length guidelines (50 pages maximum (plus Executive Summary and annexes)	X			
... conform to guidelines regarding Annexes (inclusion of complementary information)	X			

ELEMENT TO REVIEW	Y	N	NA	COMMENTS
... present consistency along the whole document in terms of English quality/style? (to avoid accidental usage of copy&paste text)	X			
<b>About the content...</b>				
ELEMENT TO REVIEW	Y	N	NA	COMMENTS
Is the overall style of the deliverable correctly organized and presented in a logical order?	X			
Is the Executive Summary self-contained, following the guidelines and does it include the main conclusions of the document?	X			
Is the body of the deliverable (technique, methodology results, discussion) well enough explained?	X			
Are the contents of the document treated with the required depth?	X			
Does the document need additional sections to be considered complete?		X		
Are there any sections in the document that should be removed?		X		
Are all references in the document included in the references list?			X	
Have you noticed any text in the document not well referenced? (copy and paste of text/picture without including the reference in the reference list)		X		
<b>SOCIAL and TECHNICAL RESEARCH WPs (WP4, 5, 12, 13, 14)</b>				
ELEMENT TO REVIEW	Y	N	NA	COMMENTS
Is the deliverable sufficiently innovative?				
Does the document present technical soundness and its methods are correctly explained?				
What do you think is the strongest aspect of the deliverable?				

ELEMENT TO REVIEW	Y	N	NA	COMMENTS
What do you think is the weakest aspect of the deliverable?				
Please perform a brief evaluation and/or validation of the results, if applicable.				
<b>AI AND TECNOLOGICAL WPS (WP6 – WP11 )</b>				
ELEMENT TO REVIEW	Y	N	NA	COMMENTS
Does the document present technical soundness and the methods are correctly explained?	X			
What do you think is the strongest aspect of the deliverable?				This report provides a succinct, readable high-level overview of the work that has been done, and limitations of the current approach are properly addressed.
What do you think is the weakest aspect of the deliverable?			X	
Please perform a brief evaluation and/or validation of the results, if applicable.			X	
<b>DISSEMINATION AND EXPLOITATION WPs (WP15 – WP17)</b>				
ELEMENT TO REVIEW	Y	N	NA	COMMENTS
Does the document present a consistent outreach and exploitation strategy?				
Are the methods and means correctly explained?				
What do you think is the strongest aspect of the deliverable?				
What do you think is the weakest aspect of the deliverable?				
Please perform a brief evaluation and/or validation of the results, if applicable.				
<b>DISSEMINATION AND EXPLOITATION WPs (WP18)</b>				
ELEMENT TO REVIEW	Y	N	NA	COMMENTS

ELEMENT TO REVIEW	Y	N	NA	COMMENTS
Does the document present the main ethical aspects regarding the use of methods and human involvement?				
What do you think is the strongest aspect of the deliverable?				
What do you think is the weakest aspect of the deliverable?				
Please perform a brief evaluation and/or validation of the results, if applicable.				

**SUGGESTED IMPROVEMENTS**

PAGE	SECTION	SUGGESTED IMPROVEMENT

**CONCLUSION**

Mark with X the corresponding line.

X	Document accepted, no changes required.
	Document accepted, changes required.
	Document not accepted, it must be reviewed after changes are implemented.

Please rank this document globally on a scale of 1-5 (1 = poor, 5= excellent) – using a half point scale. Mark with X the corresponding grade.

Document grade	1	1.5	2	2.5	3	3.5	4	4.5	5
									X