

# WISG

## Workshop interdisciplinaire sur la sécurité globale

14 & 15 mars **2024** • Couvent des Jacobins  
Rennes



## Session: «Quelle recherche dans un cadre européen ?»

Emmanuel Lazega (Sciences Po)

Yasmine Hourri (ENSAE)

Camille Roth (CNRS)

Paola Tubaro (CNRS)

## Trois plateformes européennes en construction (program Resilient Society)

AI for Citizen Intelligent  
Coaching against  
Disinformation

TITAN

vera.ai: VERification Assisted  
by Artificial Intelligence

vera.ai

AI-based-technologies for  
trustworthy solutions  
against disinformation  
Riccardo Gallotti, PI,  
Fondazione Bruno Kessler

AI4TRUST

HORIZON-CL5-2021-D1-01-05:  
AI to fight disinformation (RIA)  
Total budget: 7,177,089 euros

Trustworthy :

- ... IA explicable
- ... IA transparente
- ... IA éthique
- ... IA responsable/auditable
- ... IA juste
- ... IA sûre



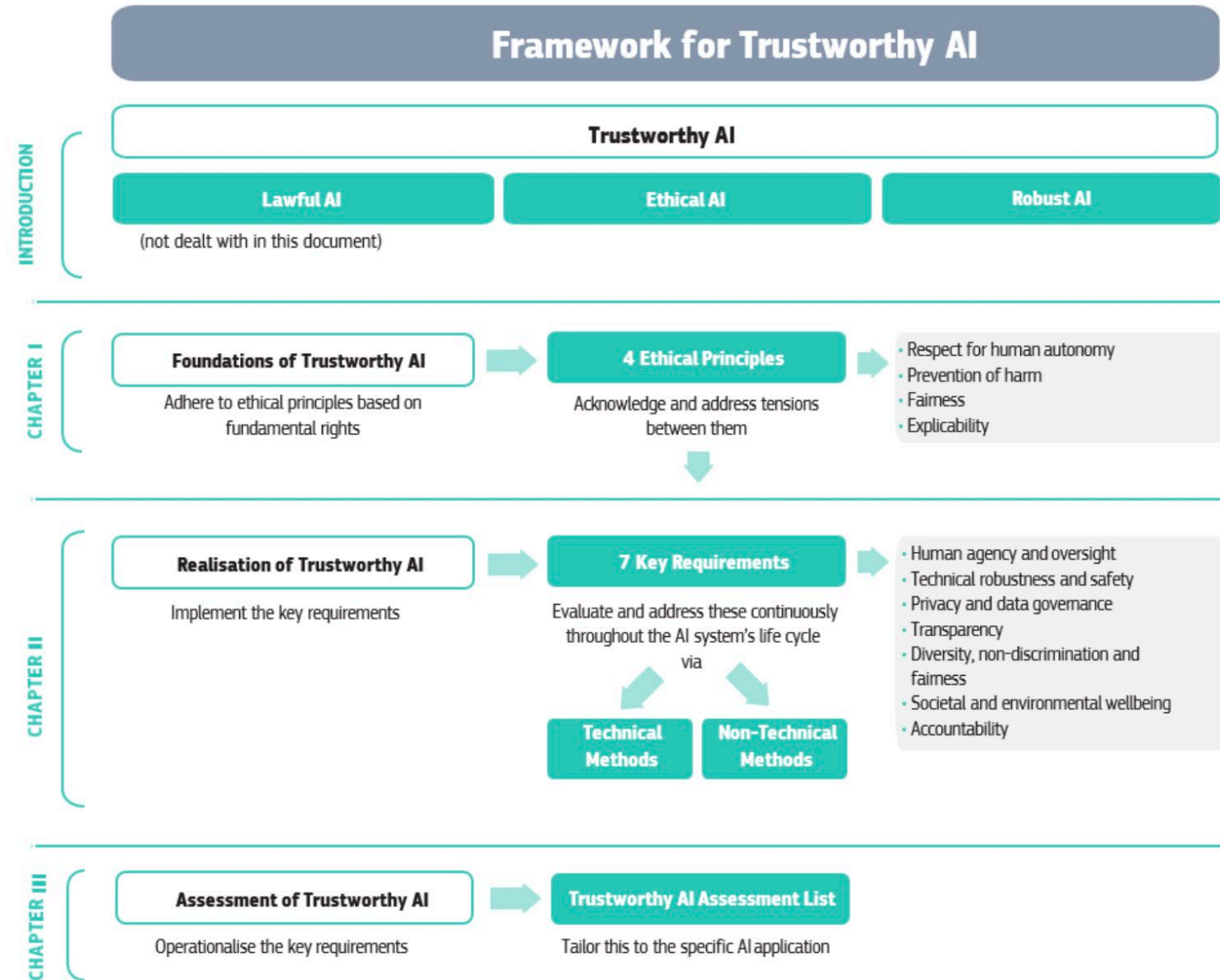
INDEPENDENT  
**HIGH-LEVEL EXPERT GROUP ON  
ARTIFICIAL INTELLIGENCE**  
SET UP BY THE EUROPEAN COMMISSION



**ETHICS GUIDELINES  
FOR TRUSTWORTHY AI**


















**TROIS COMPOSANTES** qui doivent être respectées tout au long du cycle de vie du système :

- 1. il doit être légal, conforme à toutes les lois et réglementations applicables**
- 2. il doit être éthique, garantissant l'adhésion aux principes et valeurs éthiques**
- 3. il doit être robuste, tant d'un point de vue technique que social, car même avec de bonnes intentions les systèmes d'IA peuvent causer des dommages involontaires.**







No	Participant Organisation Name	Short Name	Country
1	Fondazione Bruno Kessler	FBK	Italy 
2	ETHNIKO KENTRO EREVNAS KAI TECHNOLOGIKIS ANAPTYXIS	CERTH	Greece 
3	Università degli Studi di Trento	UNITN	Italy 
4	National Centre for Scientific Research Demokritos	NCSR-D	Greece 
5	University of Cambridge	UCAM	UK 
6	Centre National de la Recherche Scientifique	CNRS	France 
7	University Politehnica of Bucharest	UPB	Romania 
8	Fincons Group AG	FINC	Switzerland 
9	SAHER (Europe) OÜ	SAHER	Estonia 
10	GDI Global Disinformation Index gUG	GDI	Germany 
11	Demagog Association	DEMAGOG	Poland 
12	Fundación Maldita.es contra la desinformación: periodismo, educación, investigación y datos en nuevos formatos	MALDITA	Spain 
13	Ellinika Hoaxes	ELLINIKA	Greece 
14	EURACTIV Media Network	EURACTIV	Netherlands 
15	SKY Italia srl	SKYTG24	Italy 
16	ASOCIATIA DIGITAL BRIDGE	ADB	Romania 
17	Europejskie Media Sp. z o.o.	EMS	Poland 



**Human in the loop** : Le projet AI4TRUST vise à renforcer la **réponse humaine** pour lutter contre la désinformation dans l'UE en dotant les **chercheurs scientifiques et les professionnels des médias (vérificateurs de faits)** de technologies avancées basées sur l'IA permettant:

- 1) la surveillance, la détection et l'enregistrement de la désinformation (dis-), des informations erronées (mis-) et la combinaison des deux (mal-) sur les médias sociaux en ligne et les médias traditionnels
  - **multicanal** (médias sociaux en ligne et fils d'actualité distincts)
  - **multilingue** (couverture de 70 % des médias/utilisateurs de l'UE)
  - **multimodale** (contenu textuel, visuel et audio).
- 2) l'estimation du risque de consommation d'informations non fiables ;
- 3) la création d'un environnement en ligne fiable impliquant des chercheurs, **des professionnels des médias et des décideurs politiques** pour faciliter la création et la distribution d'informations fiables, tout en étiquetant et en contrant les fausses informations et la désinformation.
- 4) l'interdisciplinarité (informatique, physique statistique, sciences sociales, etc.).

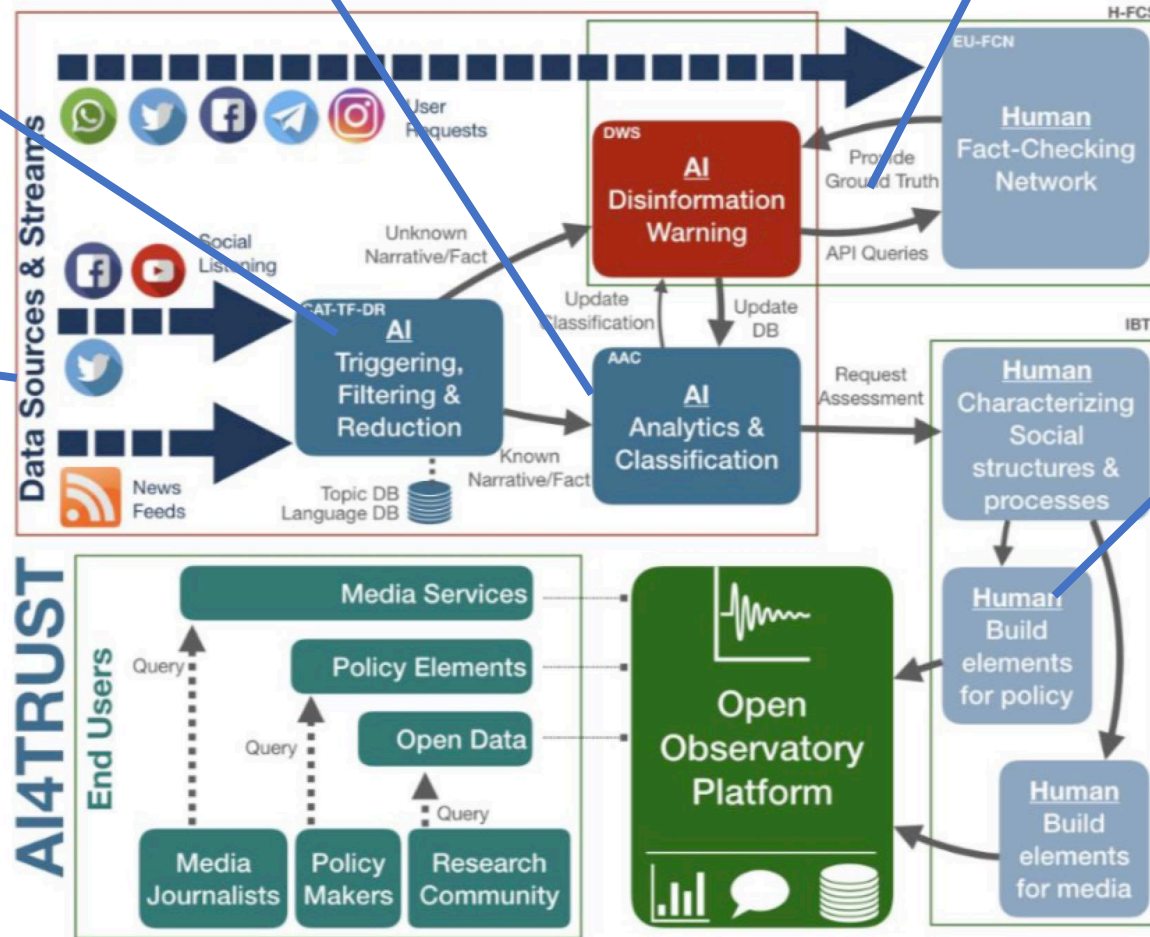


3) Le contenu associé à des récits connus est automatiquement analysé par des outils d'IA et classé comme fiable ou non fiable

4) Faits associés aux nouveaux récits filtrés en fonction de leur pertinence et du risque potentiel, puis vérifiés par des fact checkers

2) Sujets et récits existants automatiquement identifiés

1) Données saisies en continu des médias sociaux et des médias d'information



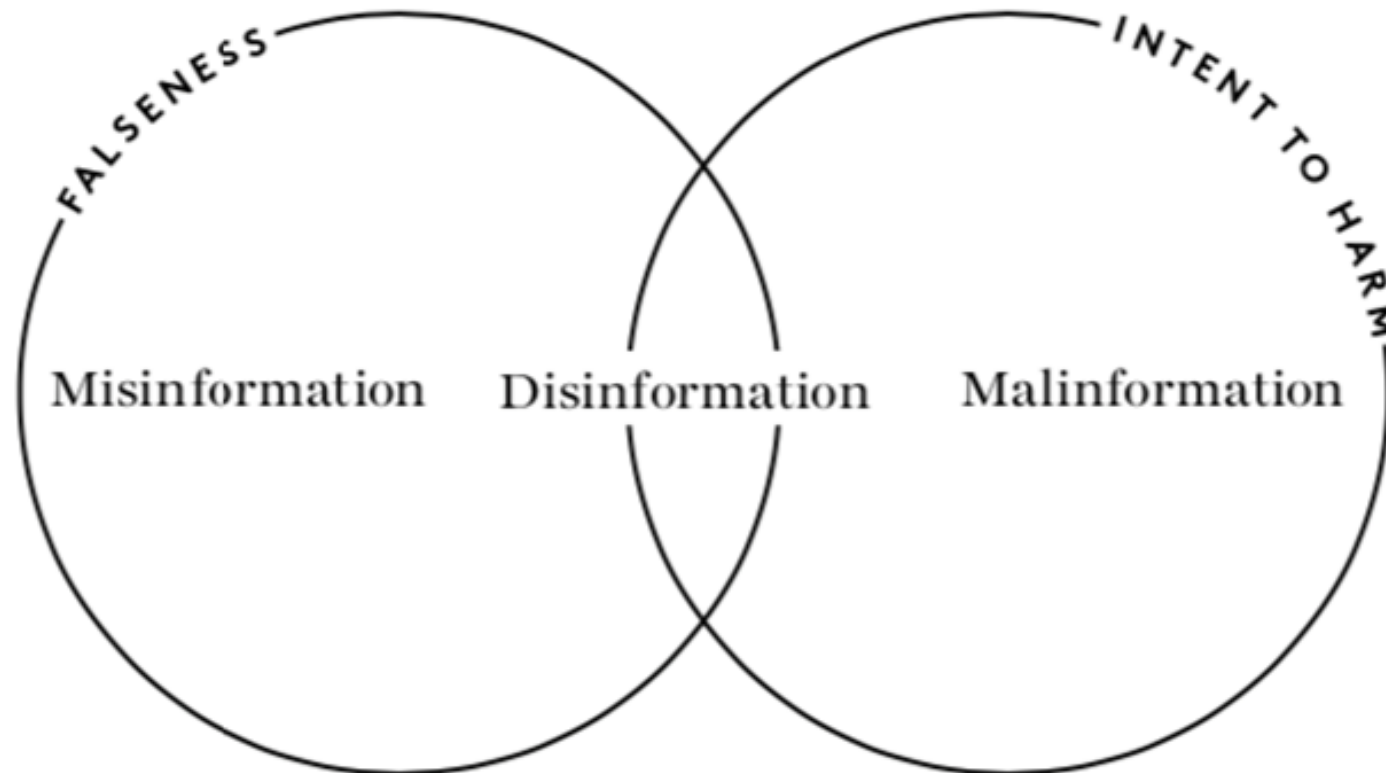
5) le résultat agrégé de l'analyse est contextualisé, notamment l'analyse de réseaux sociaux et la cartographie de la production sociale de l'information et désinformation.

7) Les utilisateurs finaux des plateformes sont les journalistes des médias, les décideurs politiques et les chercheurs, qui profitent tous des informations fournies d'une manière adaptée à leurs besoins.

6) La plateforme Open Observatory est un outil communautaire conçu dans le but de créer un environnement de confiance où l'on peut naviguer et discuter des nouveaux sujets "brûlants".



La sécurité n'est pas seulement une propriété du modèle. La production sociale de l'information par le jugement de pertinence des acteurs (identifications, autorités et normes) compte







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

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



# AI4TRUST



**CERTH**  
CENTRE FOR  
RESEARCH & TECHNOLOGY  
HELLAS



UNIVERSITÀ  
DI TRENTO



NATIONAL CENTRE FOR  
SCIENTIFIC RESEARCH "DEMOKRITOS"



CNRS CENTRE NATIONAL  
DE LA RECHERCHE  
SCIENTIFIQUE



**GDI** Global  
Disinformation  
Index



**DEMAGOG**

**M** **MALDITA.ES**  
PERIODISMO PARA QUE NO TE LA CUELEN

ASTIKI MI KERDOSKOPIKI ETAIRIA KENTRO  
KATAPOLEMISIS TIS PARAPLIROFORISIS /  
CIVIL NON-PROFIT COMPANY KENTRO  
KATAPOLEMISIS TIS PARAPLIROFORISIS

**EURACTIV**

ASOCIATIA  
DIGITAL  
BRIDGE

EUROPEJSKIE  
MEDIA SP ZOO

**FINCONS**  
**FINCONS**  
GROUP

UNIVERSITY OF  
CAMBRIDGE