# Evaluation AI-based factchecking tools – a vera.ai webinar
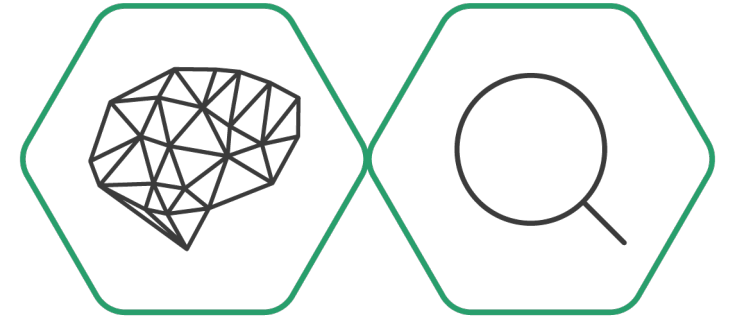
Lalya Gaye, Derek Bowler (EBU)

Anna Schild, Eva Lopez (DW)

6 November 2023

EBU Webinar
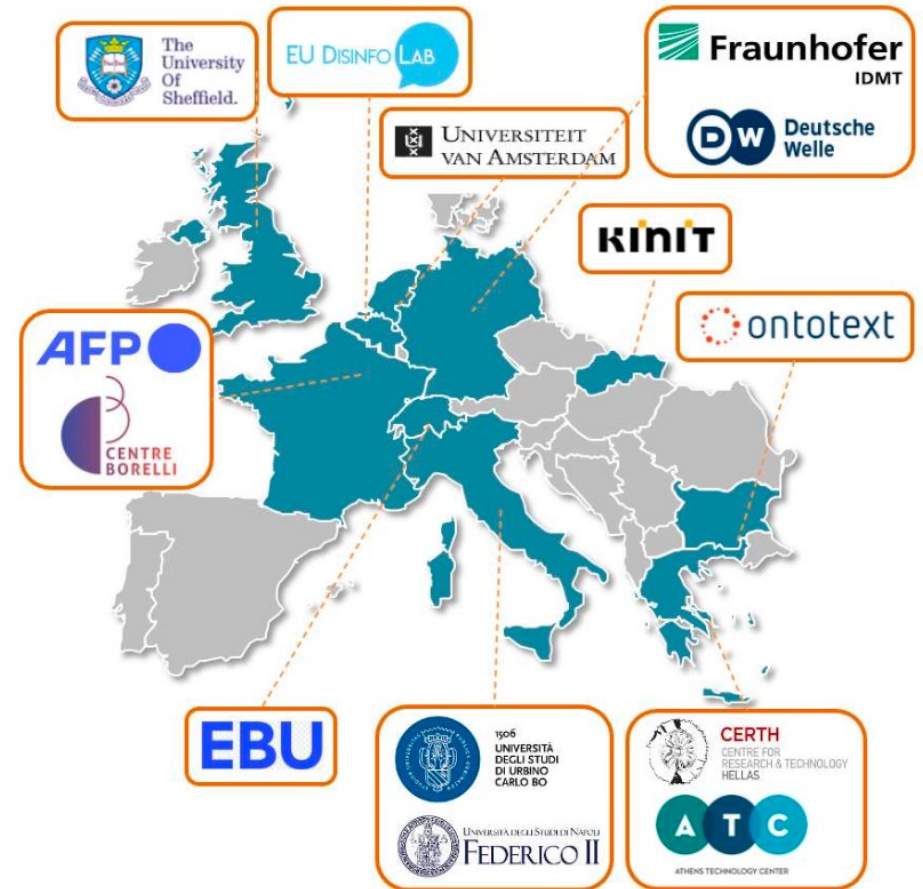
# What is vera.ai?

"VERification Assisted by Artificial Intelligence"

Develop trustworthy AI tools against disinformation, co-created with and for media professionals.
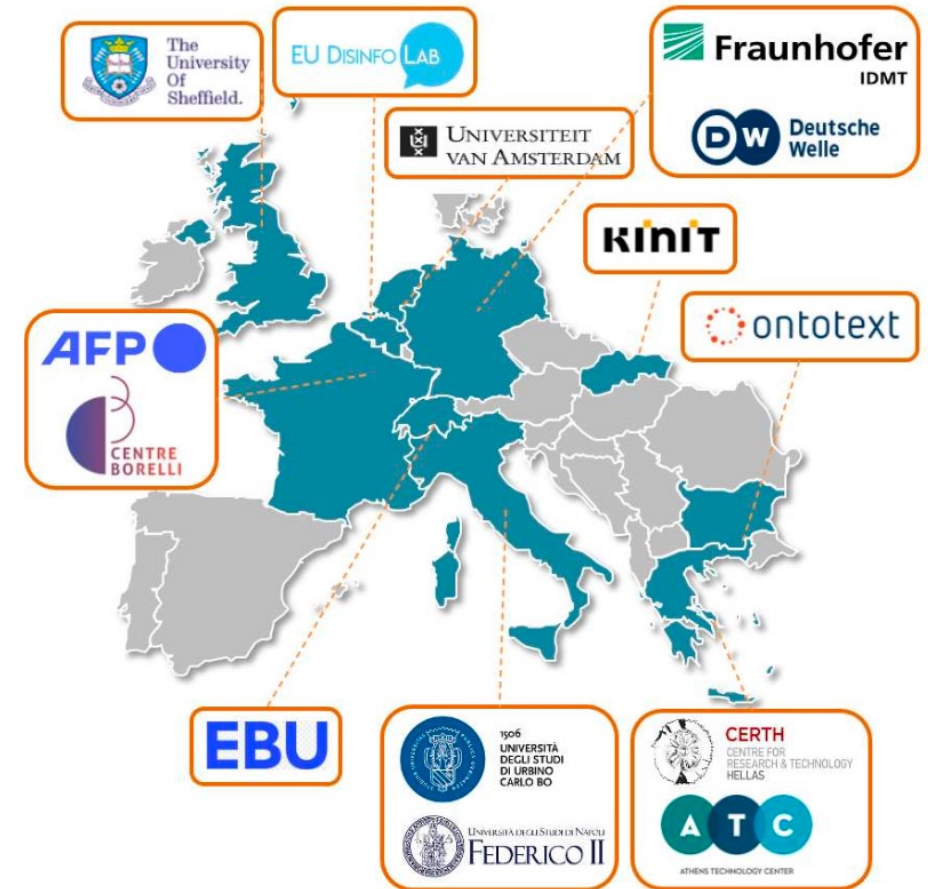
# What is vera.ai?

14 partners including Deutsche Welle

3 years Horizon Europe project
15 September 2022-2025

Co-funding from UK grant and Swiss SERI
(Swiss State Secretariat for Education,
Research and Innovation )

Follow-up to WeVerify and Truly Media:
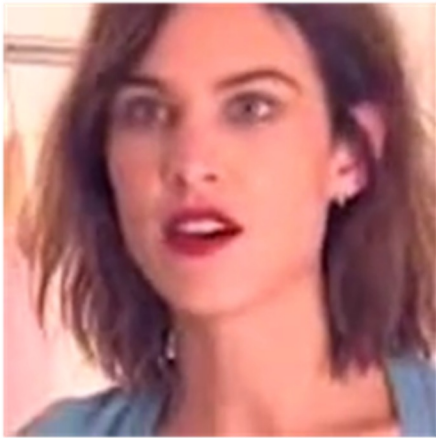compatibility with InVid.

# Objectives

Provide **novel, fair and explainable** AI tools and methods that **continuously adapt** to evolving disinformation targets, narratives, and types, based on the **latest advances** in the AI landscape.
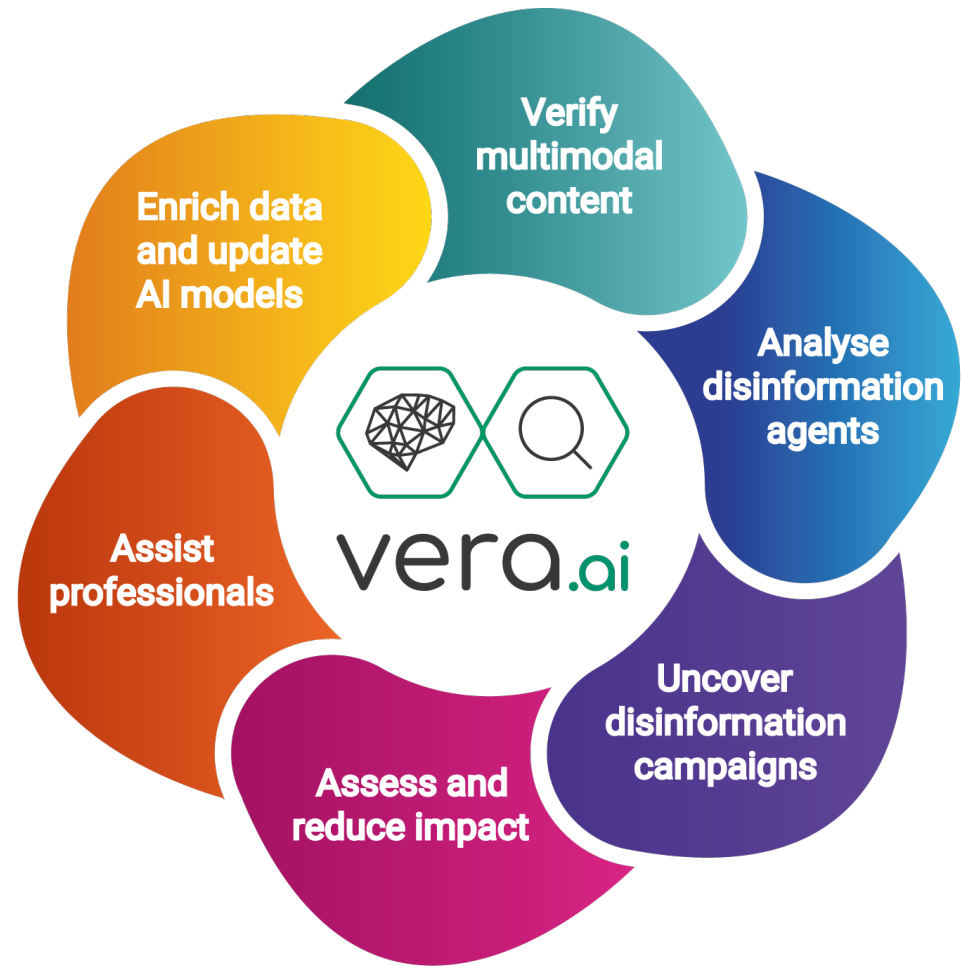
# 1. Verify multimodal content
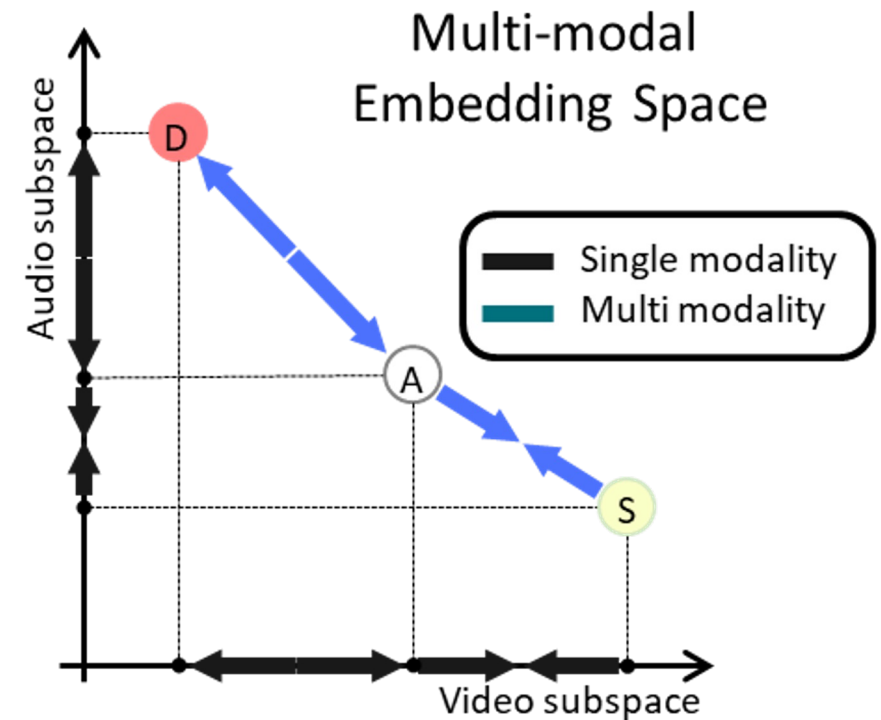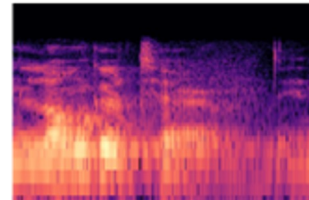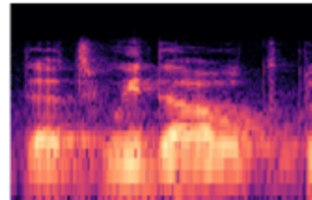
Audio-visual identity verification
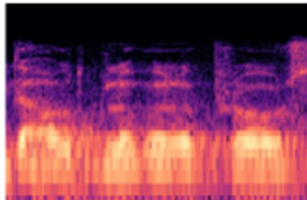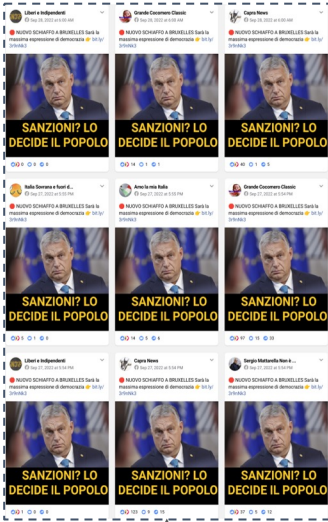
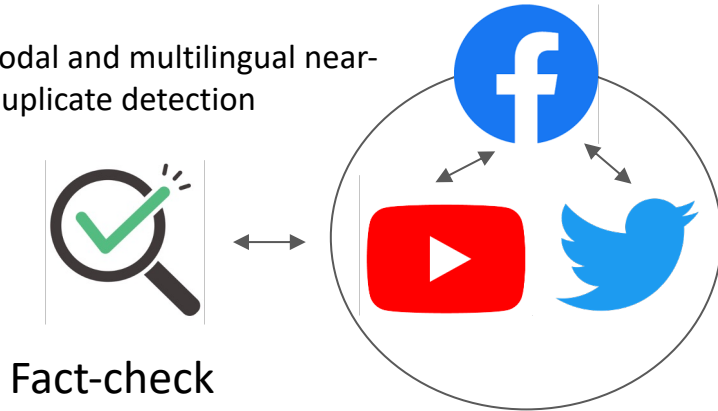# 2. Analyse the behaviour of disinformation agents

Detect coordinated link sharing behaviour and outputs networks of coordinated social media agents.
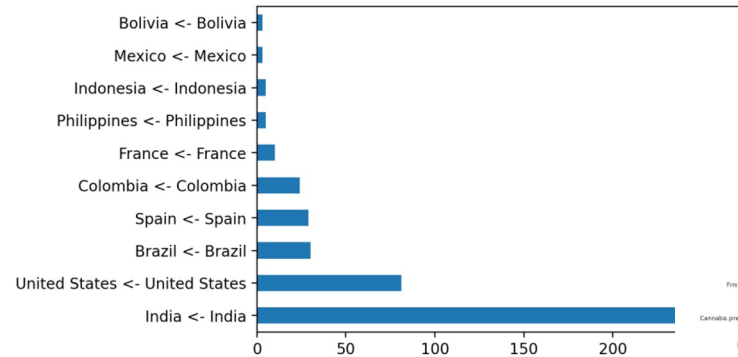
# 3. Uncover disinformation campaigns
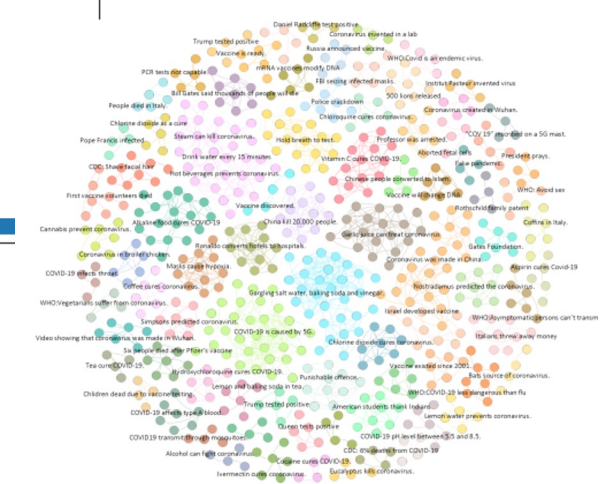


1. Multimodal and multilingual near-duplicate detection

Fact-check

2. Clustering of topically and semantically similar messages

3. Spatio-temporal analysis of disinformation campaigns

4. Hierarchical semantic model of disinformation

Medical domain

Antivax narrative

Claim 1

Claim 2
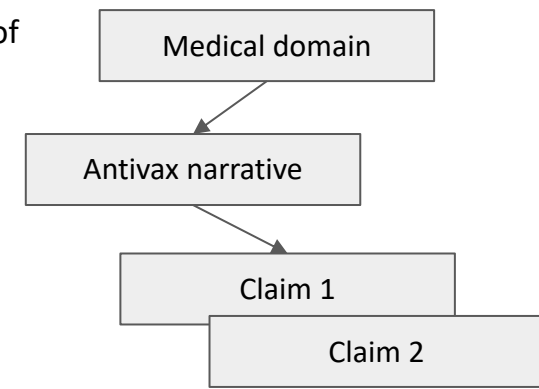
# 4. Assess and reduce impact
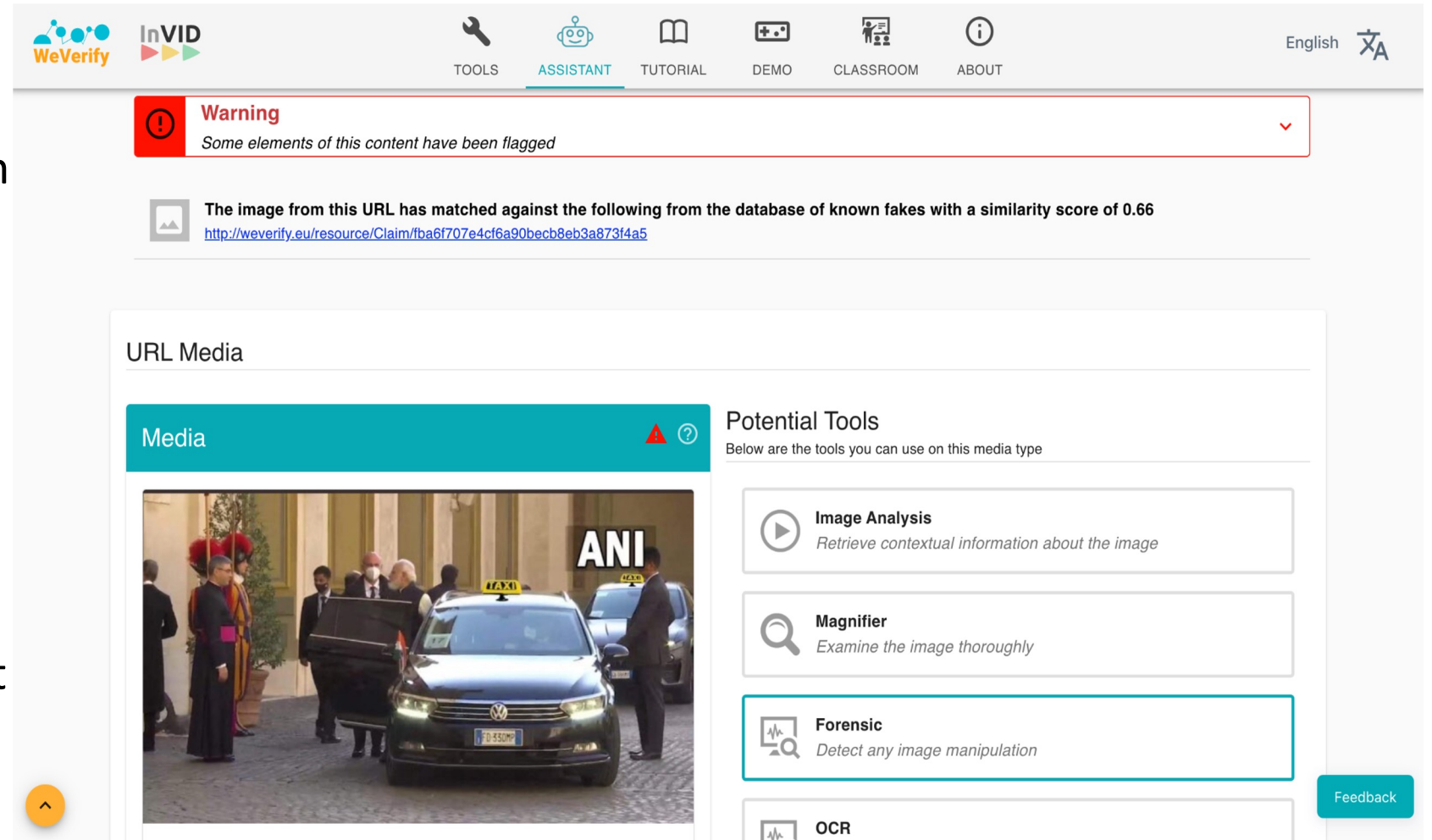
Monitoring the amplification role of platform algorithms
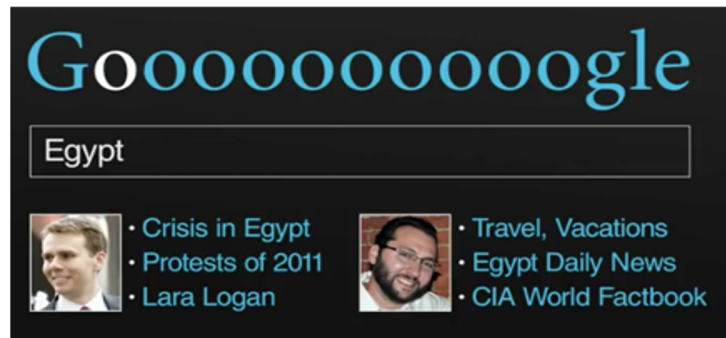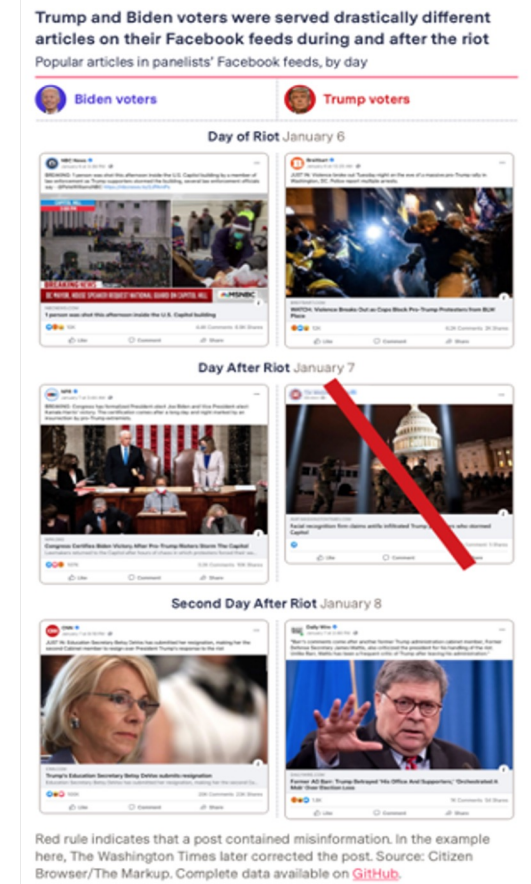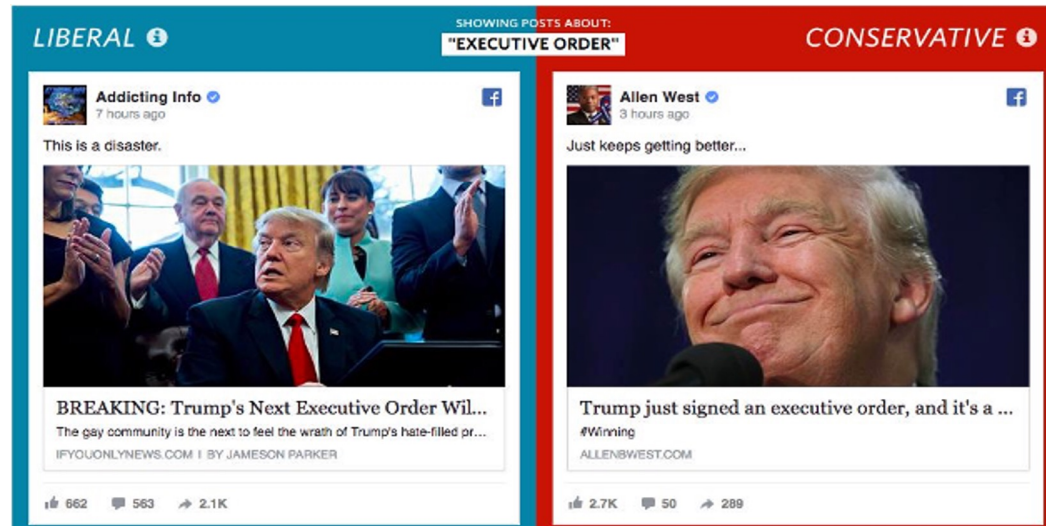
# 5. Assist professionals

- Enhance the web-based assistant from the WeVerify plugin

- Chatbot explaining the AI tool outputs

- Assistance with debunk authoring

- Low overhead flagging of AI output errors

# 6. Enrich data and update AI models

- Indexing and curation of newly debunked content
- Expert-in-the-loop approach to retrain AI models
- U=Use of expert-sourced lists of suspicious content

# Foundation: WeVerify user-facing tools

**InVID-WeVerify Verification plugin** (counts 75,000+ weekly users)

**Truly Media** (EDMO technical platform)

**Database of Known Fakes** (DBKF)

Already gathering a large number of users across Europe and worldwide, mostly media professionals, human rights activists, etc.
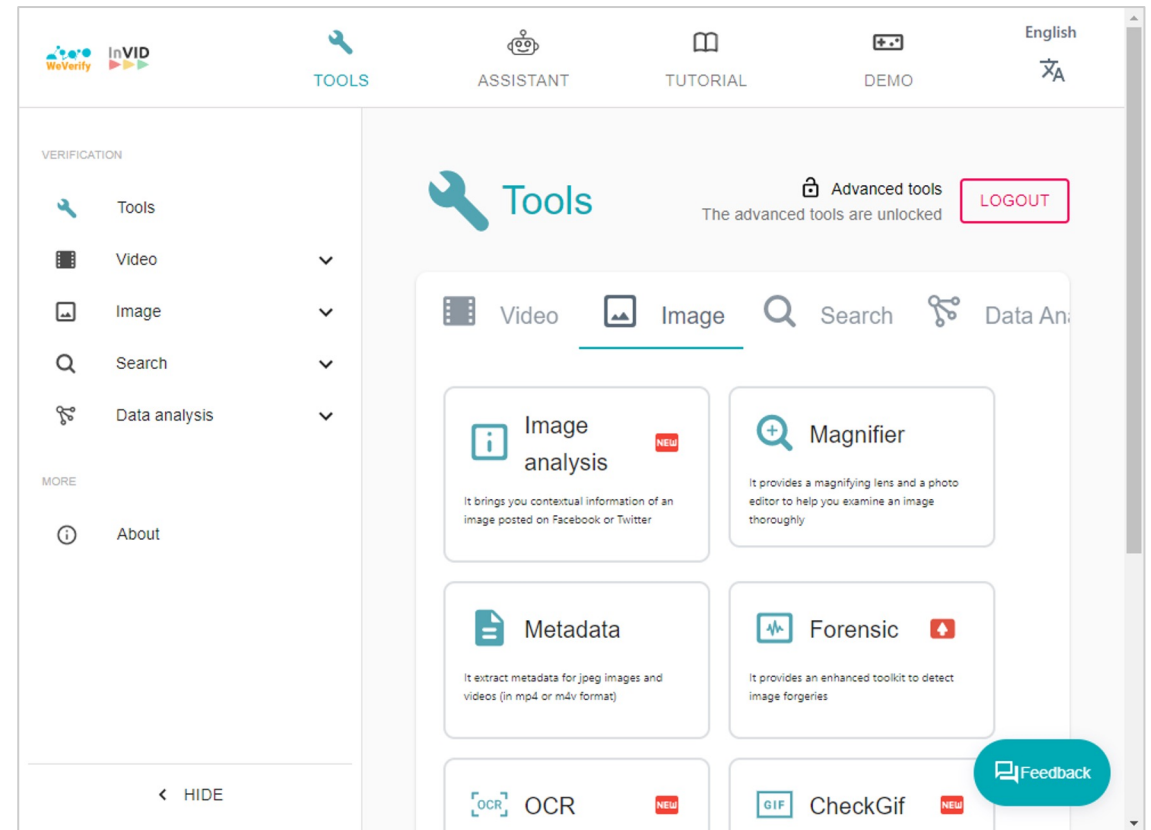
**But how do we make sure that these tools work for you and fit in your workflow?**

# User-Centred approach



The designer's share

The *real* problem
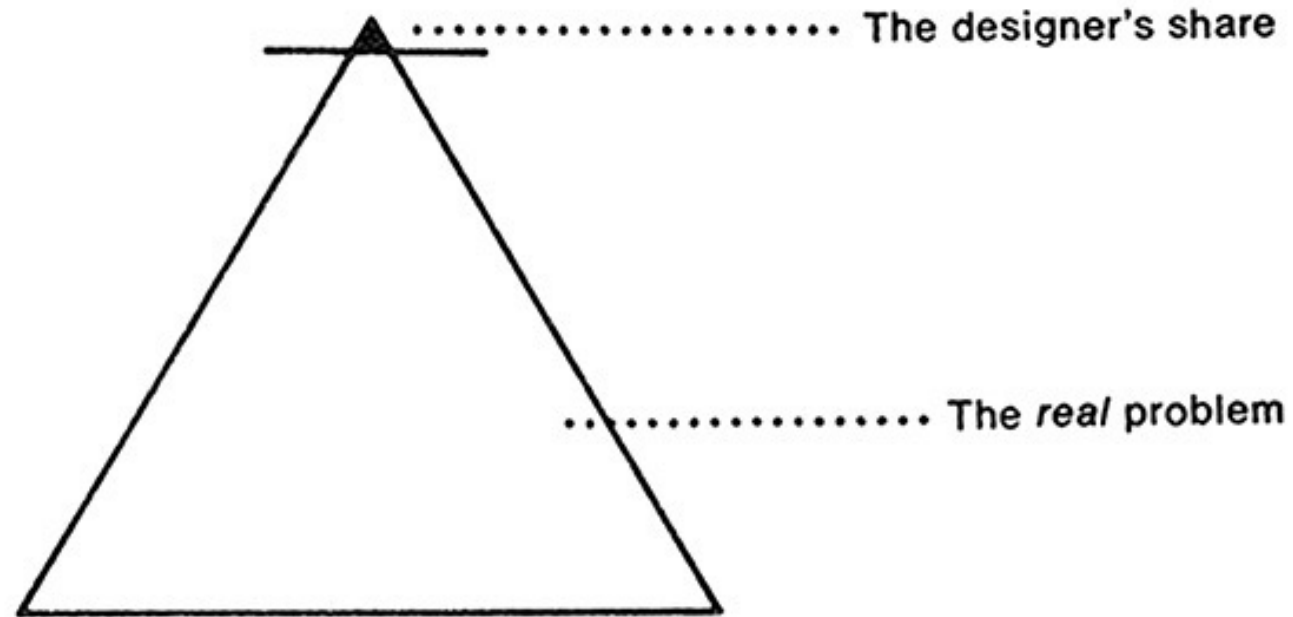
Diagram about social design by Victor Papanek.
'Design for the Real World: Human Ecology and Social Change' (London: Thames and Hudson, 2011)

# EBU + DW partnership

EBU:

AI and Data Initiative
News Department
EBU Academy
EBU community

+

Deutsche Welle

→

Previously: Helped define **user needs** and **design requirements** through **participatory design** workshops – complementary to ethnographic studies, data sprints and survey run by other project partners

Now: Take part in evaluating tools implementation: **participatory evaluation**

**Dissemination** and **training**

Lalya Gaye

Derek Bowler

Nathalie Labourdette

Anna Schild

Eva Lopez

Jochen Spangenberg

# vera.ai: an iterative process
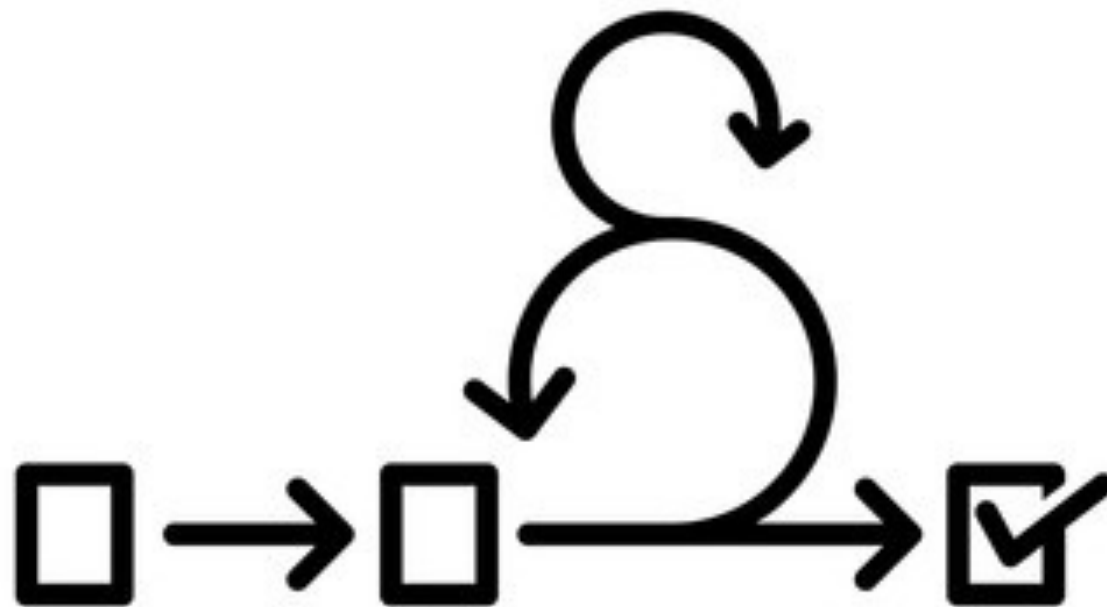
# Participatory Evaluation

Scandinavian tradition
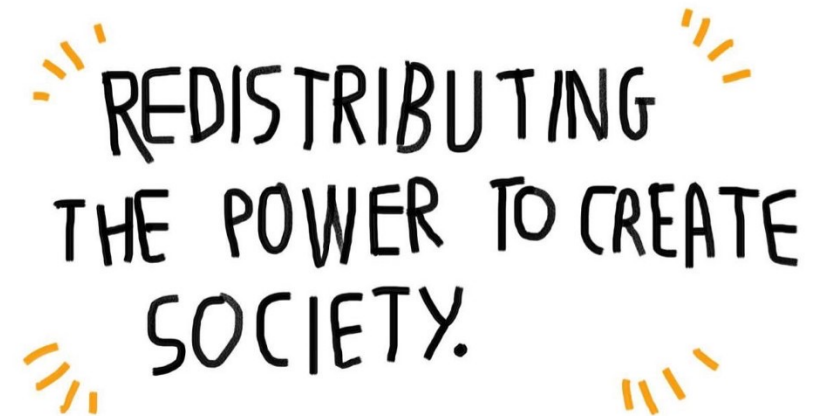
Based on participatory design, where actively involving all stakeholders in a design process > evaluation process

Value each stakeholder's perspective and input

Make sure that implementations

- meet the needs and expectations of the users
- are relevant to them
- are meaningful
- make them gain a sense of ownership of the solutions

Interdisciplinary and democratic process

REDISTRIBUTING THE POWER TO CREATE SOCIETY.

Anu Paajanen

# What is PE in vera.ai?

- A qualitative, user-centred approach, focused on **end-user practices and workflows** (vs. evaluating a prototype's technical performance): usability, error transparency, explainability, cross-examination etc

- **Stakeholders involvement** (end-users and tech partners) in the process definition and decision-making

- Applicable to prototypes at **different levels of readiness**

- Methods **tailored to evaluation needs**: interviews, online workshops, video ethnography, diaries, in-the-wild testing, winter school etc

- **Advandages**: user-centred evaluation results, fitting into user workflows, agile and iterative

# Participatory evaluation in vera.ai

- **Following up** on the user requirements defined in Participatory Design sessions

- Agile timeline

- Evaluating **features** rather than entire tools, focusing on **workflow-specific features**

- **Complementary** to the technical evaluation process (more on this later)

- **All together:** Collectively define evaluation goals and criteria: what does a design need to achieve in order to be considered successful?
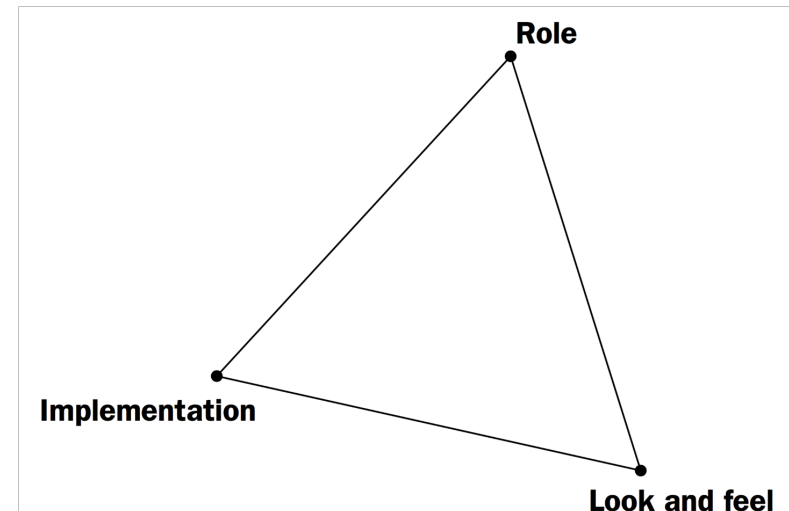
# Example process – Initial situation

## Technical partner perspective

- I know that (my) **AI systems face uncertainties**
- I want to **be transparent** about that

→ **How do I successfully inform users about uncertainties of my tool, so that they can make well-grounded decisions?**

## End-user perspective

- I want to **integrate AI based systems** into my journalistic workflow
- For my **decision-making process**, I need to know whether/to what degree I can **rely on the results** of an AI system

→ **Does the AI system really help me in my decision-making?**

# How to go from here?
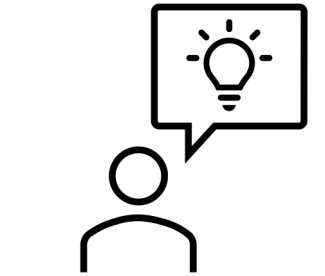
Technical partner has implemented a prototype based on user requirements:

Displaying uncertainties when presenting the results

We break down the prototype into „evaluatable" bits

We contact end-users to take part
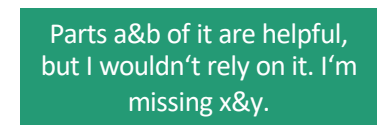
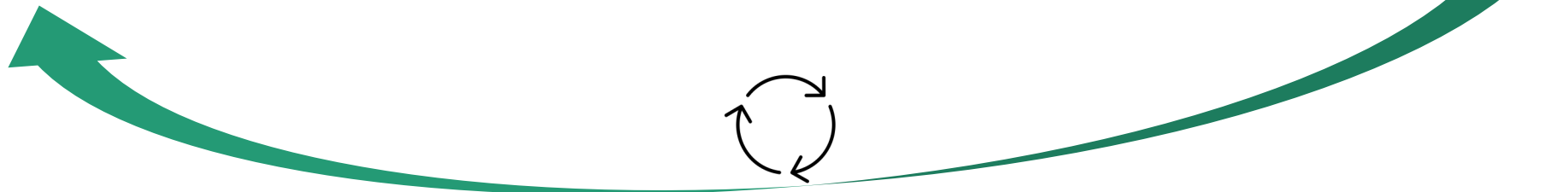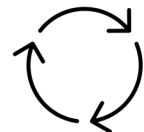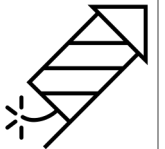We decide all together what evaluation criteria to set
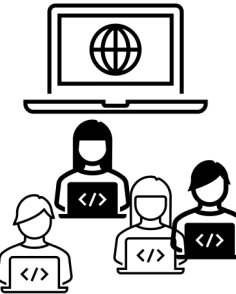
Evaluation!

End-users answer 'Does this way of displaying uncertainties help me in my decision-making?'

Uncertainties are displayed in a way that I can rely on it.

Parts a&b of it are helpful, but I wouldn't rely on it. I'm missing x&y.

# Search

Search for Debunks by filters and/or keyword

## Filters

Clear filters

### Language

All Languages ▼

### Organization

All Organizations ▼

### Author

All Authors ▼

### Locations

All Locations ▼

### Concepts

All Concepts ▼

### Date

01/01/2000 📅        01/01/2023 📅

Filter

Search...    🔍

**Search Result For:**

LOCATION:    CONCEPT:    AUTHOR:    ORGANIZATION:

Search tips

**About:** 127312 Debunks    **Order By:** Relevance | Date

**DEBUNK**                                      📅 19 hours ago

### Fotografia da atriz que interpretou a "Mulher Maravilha" viraliza nas redes sociais: alegadamente, Gal Gadot ter-se-á juntado aos militares israelitas para servir a sua pátria. Mas será mesmo assim?

Link to Debunking Article

👤 Marta Ferreira 📊 3.00

**DEBUNK**                                      📅 23 hours ago

### Conseguir emprego depois dos 50 anos pode ser, muitas vezes, uma tarefa difícil. Será que este é o limite máximo de idade no que diz respeito à admissão de novos trabalhadores na Função Pública?

Link to Debunking Article

# Get involved!

**We're looking for professional journalists, factcheckers and media researchers**

**Contact:** Lalya Gaye <gaye@ebu.ch>

Would your colleagues be interested too?

Follow our website / social channels for more activities!

https://www.veraai.eu/

# Technical Evaluation (AFP)

Complementary to the participatory evaluation:

Tools performance vs. workflow

Contact: Denis Teyssou (AFP) denis.teyssou@afp.com

# Testing vera.ai synthetic beta image detector

Testthe new synthetic image detector:

Onstall on your Chrome-based browser the [InVID-WeVerify verification plugin](#)



Home > Extensions > Fake news debunker by InVID & WeVerify

**InVID WeVerify** Fake news debunker by InVID & WeVerify

Featured

★★★★☆ 25 ⓘ | News & Weather | 100,000+ users

# Testing vera.ai synthetic beta image detector

**Register for the advanced tools** (if you do not have them yet). If you're a new registrant, please indicate "**betatester**" as a Role.

# Testing vera.ai synthetic beta image detector

If **you are already an advanced tools user** and do not see the synthetic media detector in the image tab, then please **send AFP a message through the feedback button**. They will upgrade your profile asap so that you can get access.



**Synthetic image detector**

**Feedback button**

# Questions and comments?

# Lalya Gaye
## Contact: gaye@ebu.ch

vera.ai

CERTH — CENTRE FOR RESEARCH & TECHNOLOGY HELLAS

The University Of Sheffield.

ATC — ATHENS TECHNOLOGY CENTER

KINIT

ontotext

DW Deutsche Welle

UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II

1506 UNIVERSITÀ DEGLI STUDI DI URBINO CARLO BO

AFP

UNIVERSITEIT VAN AMSTERDAM

école normale supérieure paris—saclay — université PARIS-SACLAY

EBU — OPERATING EUROVISION AND EURORADIO

EU DISINFO LAB

Fraunhofer IDMT

Follow us on Twitter: @veraai_eu
Website: https://www.veraai.eu/