

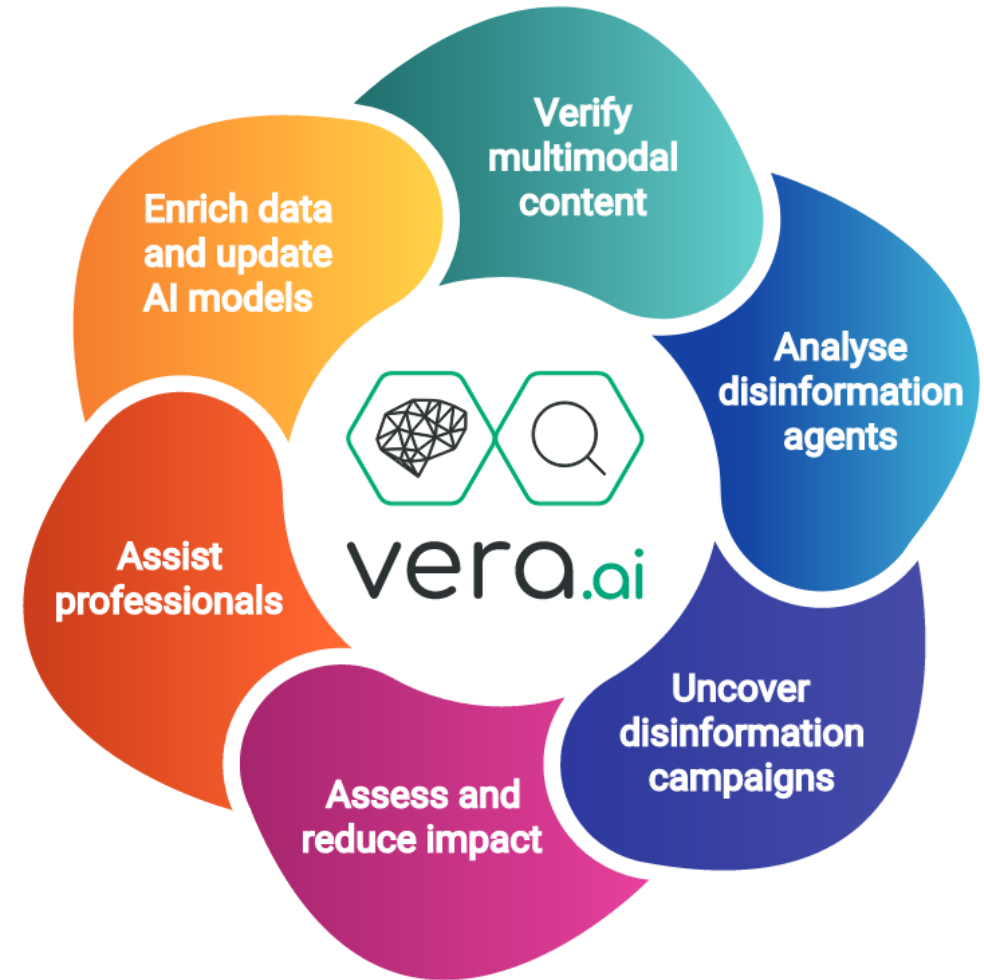
Vera.ai: VERification Assisted by Artificial Intelligence

Denis Teyssou
Agence France-Presse
25 October 2022
#Disinfo2022 Conference



Overall goal

Develop novel AI- and network science-based methods that assist verification professionals throughout the complete content verification workflow



vera.ai consortium



Participant No.	Participant organisation name	Country
1 (Coordinator)	Centre for Research and Technology Hellas	Greece
2	The University of Sheffield	UK **
3	Università di Urbino Carlo Bo *	Italy
4	Fraunhofer Institute for Digital Media Technology *	Germany
5	University of Amsterdam *	The Netherlands
6	Kempelen Institute of Intelligent Technologies *	Slovakia
7	Università degli Studi di Napoli Federico II *	Italy
8	Borelli Center, ENS Paris-Saclay *	France
9	Athens Technology Centre	Greece
10	Sirma AI EAD (trading as Ontotext)	Bulgaria
11	AFP news agency	France
12	Deutsche Welle	Germany
13	EU DisinfoLab	Belgium
14	European Broadcasting Union *	Switzerland **

8 Research partners

2 Commercial organisations with relevant products

2 Global news and fact-checking providers

An NGO focusing on tackling disinformation at EU-level

The world's biggest union of public service broadcasters

* new partners after WeVerify

** associated partners

Foundation: WeVerify user-facing tools

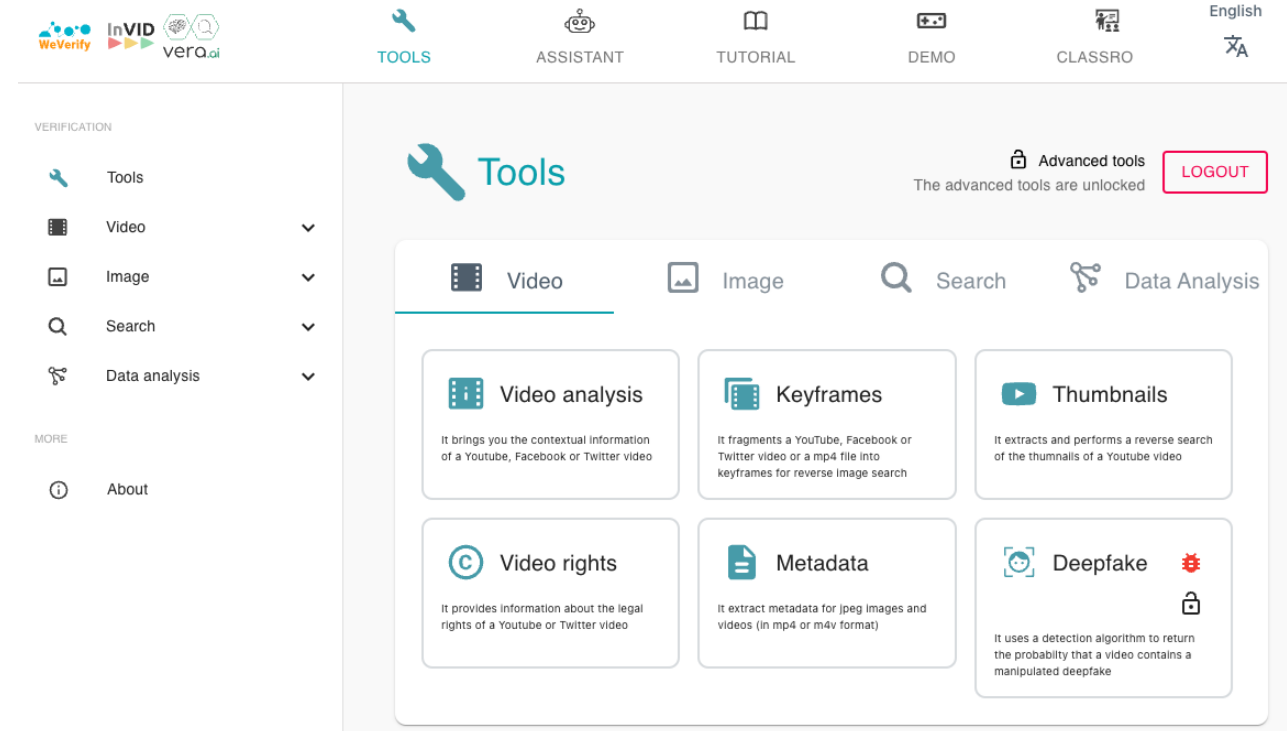


InVID-WeVerify Verification plugin (counts 80,000+ weekly active users)

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.



<http://u.afp.com/iZDr>

Existing verification features by vera.ai partners

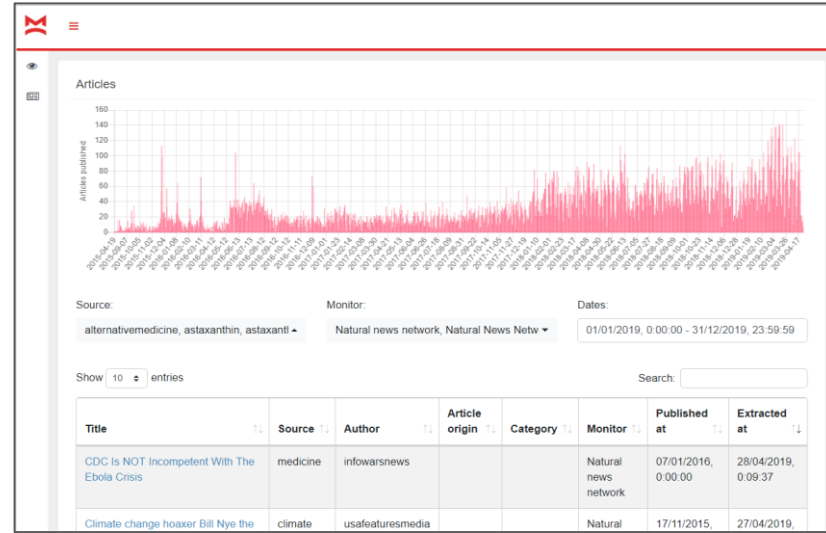


Image Verification Assistant

Detect digital tampering in images.

DeepFake Detection

Check if media contains deepfake manipulated faces



4CAT: Capture and Analysis Toolkit

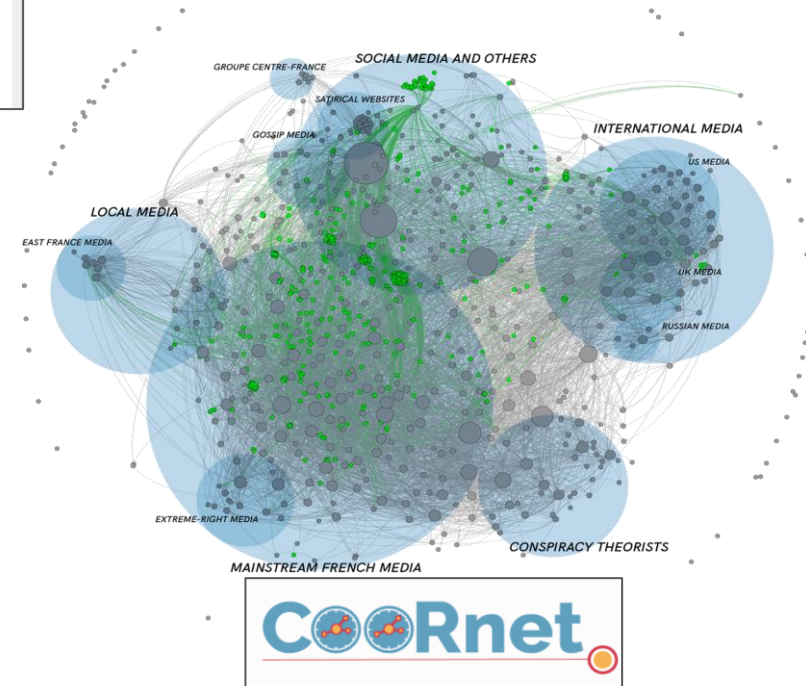
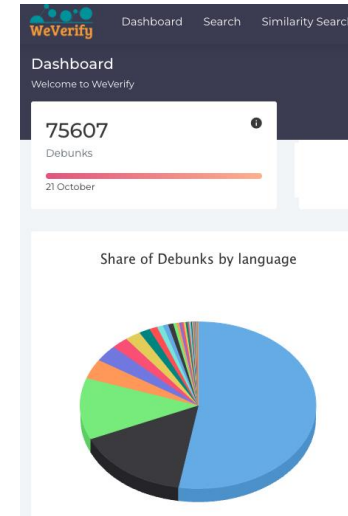
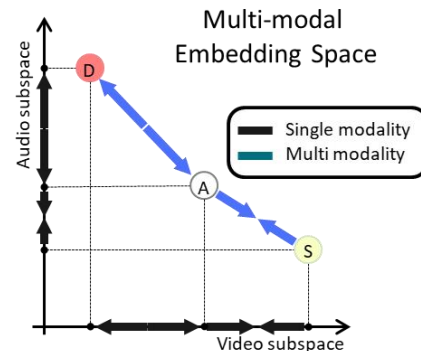
Launch tool
Create datasets

WeVerify OCR Service

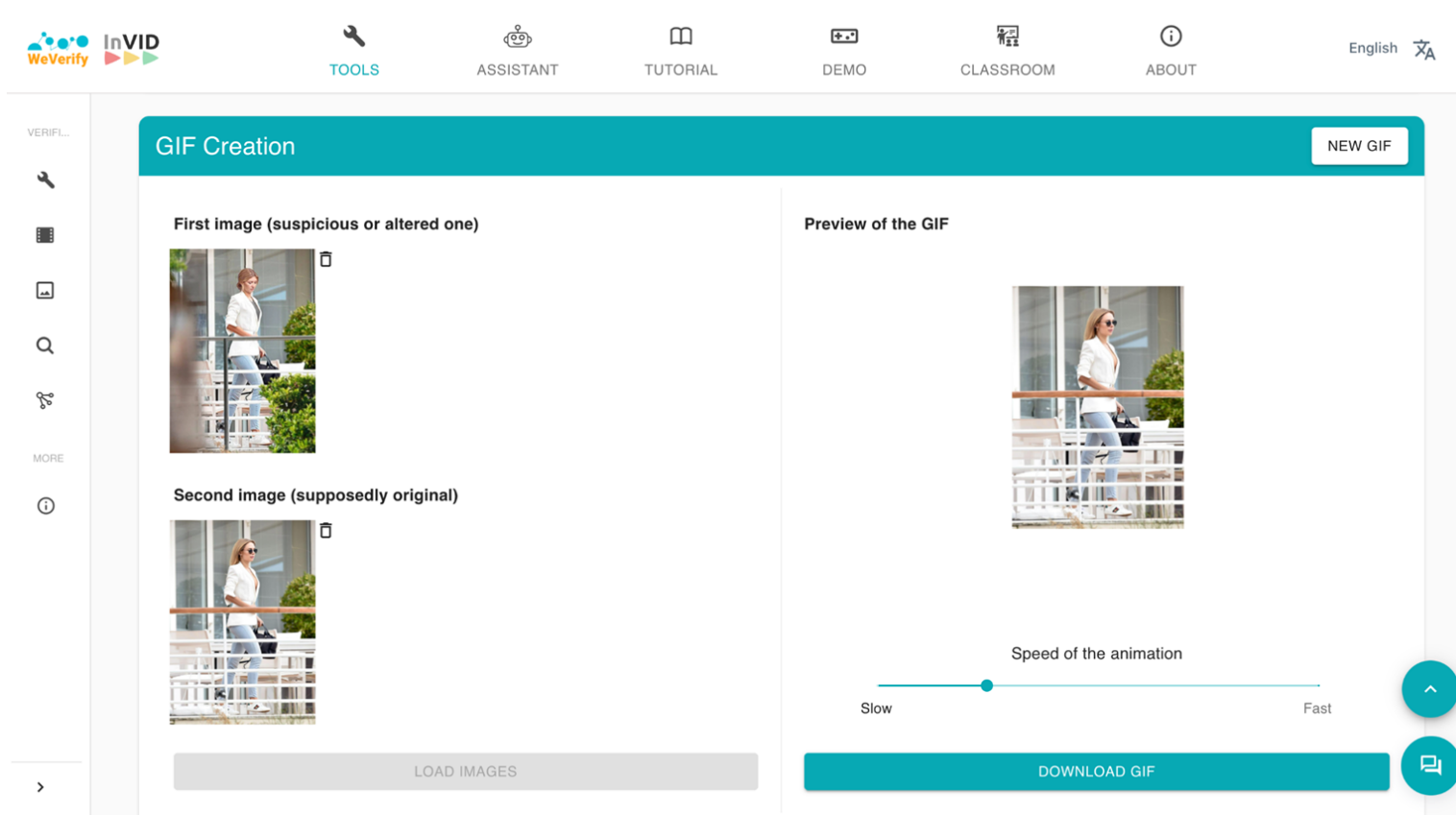
This pipeline identifies URLs in the input text and tries to process them with optical character recognition to identify text in images.

150 free requests / day
Batch processing not available

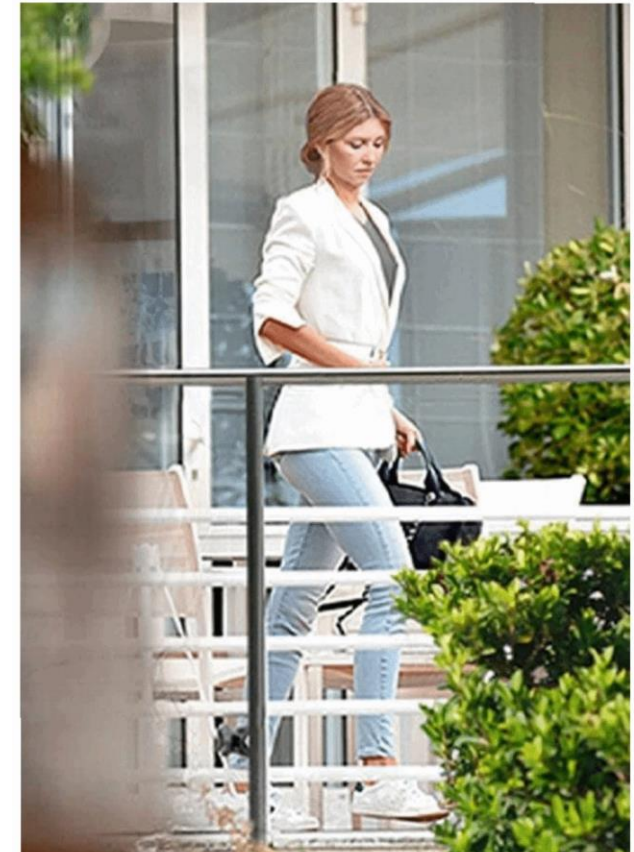
A) Anchor video S) Same Subject D) Different Subject



“Fact-checker in the loop” AI design approach



The screenshot shows the InVID web interface for GIF creation. At the top, there is a navigation bar with icons for TOOLS, ASSISTANT, TUTORIAL, DEMO, CLASSROOM, and ABOUT, along with a language selector set to English. The main area is titled "GIF Creation" and features a "NEW GIF" button. It is divided into two columns: "First image (suspicious or altered one)" and "Second image (supposedly original)". Both columns contain a video frame of a woman in a white blazer walking on a balcony. Below the images are "LOAD IMAGES" and "DOWNLOAD GIF" buttons. A "Speed of the animation" slider is positioned between the columns, ranging from "Slow" to "Fast".



“Debunker in the loop” → <http://u.afp.com/vera-survey>

Hide technical complexity and minimise errors

Results

Output images

Panorama

Reg. image1

Reg. image2

inliers (129)

outliers (5)

input img1

input img2

Compare



Zoom 1x

DE FACTO
Des clés pour mieux s'informer

À LA UNE EXPLORER COMPRENDRE À PROPOS



Copyright AFP 2017-2022. Droits de reproduction réservés.

Auteur(s)
AFP France

Using AI to automate, simplify and improve fact-checkers UX

Fact-checkers are using AI computer vision tools




A screenshot of the InVID verification tool interface. The top navigation bar includes "WeVerify InVID", "TOOLS", "ASSISTANT", "TUTORIAL", "DEMO", "CLASSROOM", and "English". A left sidebar lists verification options: Tools, Video, Video Analysis, Keyframes, Thumbnails, Video rights, Metadata, Deepfake, Image, Search, and a "HIDE" button. The main area shows "Results" with a "SHOW DETAILED VIEW" button and "ZOOM OUT" / "ZOOM IN" controls. A grid of keyframes is displayed, showing various scenes of destruction. At the bottom, there is a "To be enhanced within" section with the vera.ai logo and a "DOWNLOAD KEYFRAMES" button.

A screenshot of a Full Fact article titled "Video of explosion aftermath is from Lebanon in 2020, not Ukraine". The article is dated 4 MARCH 2022. It includes a "WHAT WAS CLAIMED" section stating "A video shows the devastation caused by an explosion in a Ukraine city." and an "OUR VERDICT" section stating "The video shows the aftermath of an accidental explosion in Lebanon in 2020 and doesn't relate to the Russian invasion of Ukraine in February 2022." Below the verdict, it says "This isn't a video of Ukraine." At the bottom, a red circle highlights the text "Finally, by using a tool like InVID—which lets you break down a video into a series of still images, and then search for similar images online—you can find lots of other photographs of the same event." The word "InVID" in this text is circled in red.

<https://fullfact.org/online/video-of-explosion-aftermath-is-from-lebanon-in-2020-not-ukraine/>

And also AI multilingual tools




TOOLS ASSISTANT TUTORIAL DEMO CLASSROOM ABOUT English

VERIFICATION

- Tools
- Video
- Image
- Image Analysis
- Magnifier
- Metadata
- Forensic
- OCR
- CheckGif
- Deepfake
- Geolocalizer
- Search
- Data analysis

HIDE

Image Analysed




YANDEX BING

Complete Text

NOWAR оСтАНОВИТЕ Войн НЕ ВЕРВТЕ ПРопаГАНДе 'ДЕСь ВАМ ВРУТ SIANS ASAINST WAR

COPY TO CLIPBOARD TRANSLATE

Blocks



Russian

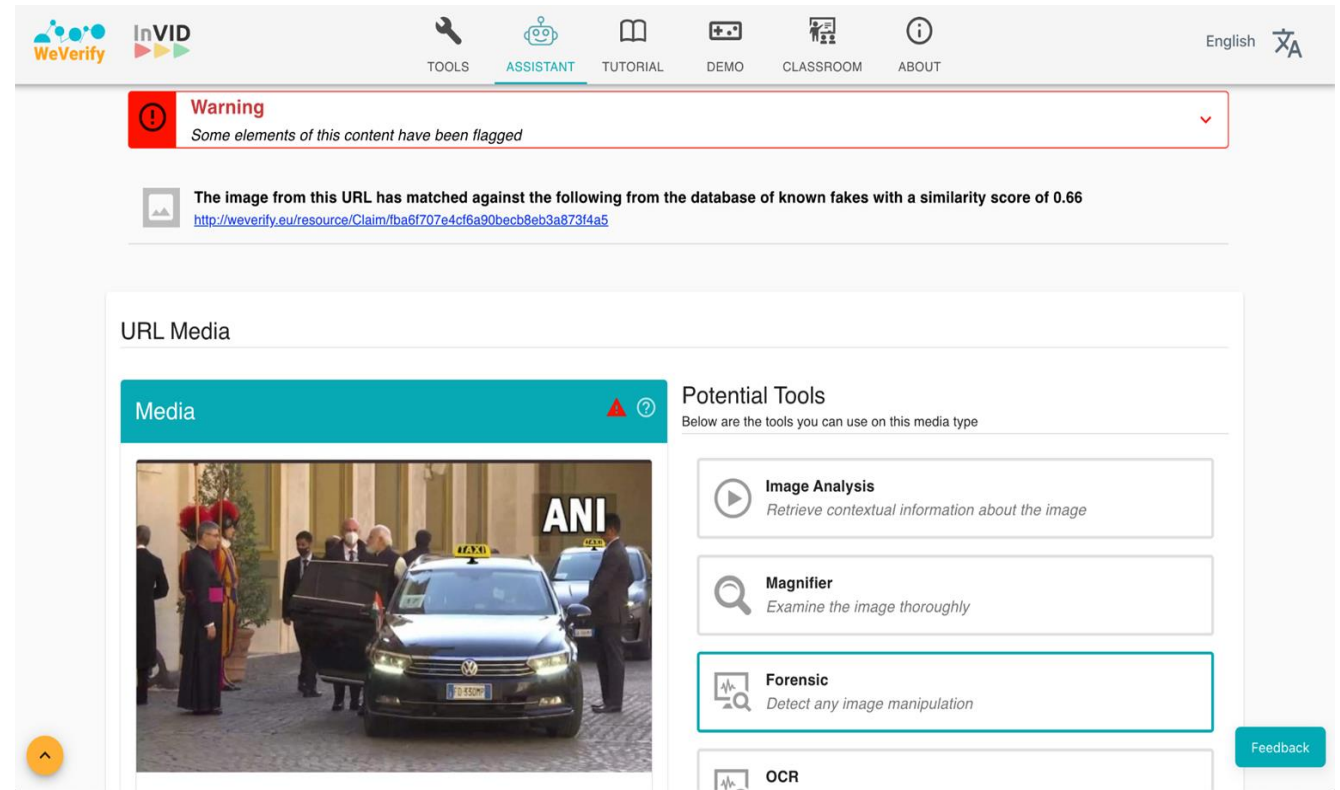
NOWAR оСтАНОВИТЕ Войн
НЕ ВЕРВТЕ ПРопаГАНДе
'ДЕСь ВАМ ВРУТ SIANS
ASAINST WAR

To be enhanced within
vera.ai

Feedback

Providing more guidance to professionals

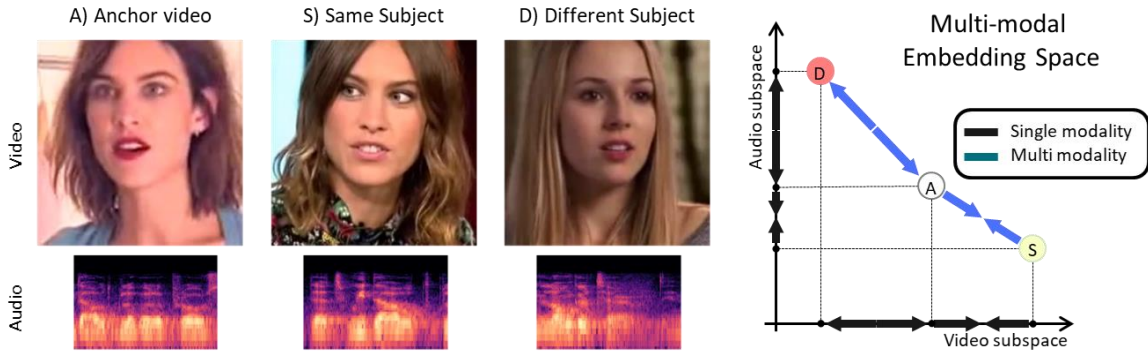
- Enhance the web-based assistant from the InVID-WeVerify plugin
- Chatbot explaining the AI tool outputs
- Assistance with debunk authoring
- Low overhead flagging of AI output errors
- Enrich data on-the-fly and update AI models



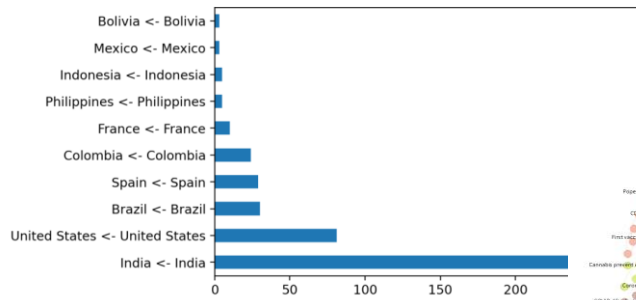
The screenshot shows the InVID-WeVerify web interface. At the top, there is a navigation bar with icons for TOOLS, ASSISTANT, TUTORIAL, DEMO, CLASSROOM, and ABOUT. A warning message is displayed: "Warning: Some elements of this content have been flagged". Below this, a message states: "The image from this URL has matched against the following from the database of known fakes with a similarity score of 0.66" and provides a URL: <http://weverify.eu/resource/Claim/fba6f707e4cf6a90becb8eb3a873f4a5>. The main content area is titled "URL Media" and shows a media player with a video thumbnail of a car. To the right of the media player, there is a section titled "Potential Tools" with the following options: "Image Analysis" (Retrieve contextual information about the image), "Magnifier" (Examine the image thoroughly), "Forensic" (Detect any image manipulation), and "OCR". A "Feedback" button is located at the bottom right of the interface.

- **Generalization:** much higher error rates on unseen data and new kinds of disinformation
- **Robustness to new data:** platform-specific formats, APIs, metadata
- **Multi-modal and cross-modal detection** (e.g. out-of-context)
- **High-level transparency** including explainability and interpretability

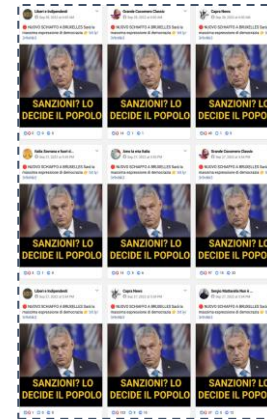
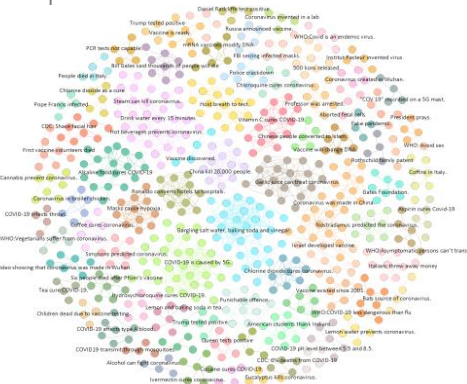
Unsolved challenges ahead



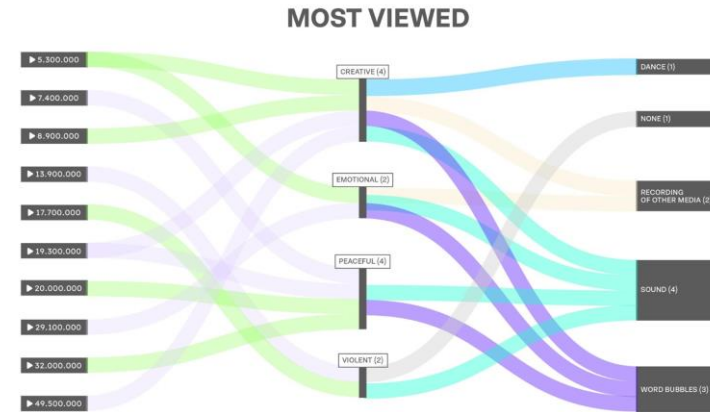
Deepfakes, hybrid deepfakes, audio-visual manipulation



Spatio-temporal analysis of disinformation campaigns



Link sharing behaviour analysis, graph neural networks, node classification, knowledge inference



amplification role of platform algorithms

Thank you! Questions?



Contact: @dteyssou



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

Co-financed by the European Union, Horizon Europe programme,
Grant Agreement No 101070093.

Additional funding from Innovate UK grant No 10039055 and the
Swiss State Secretariat for Education, Research and Innovation (SERI)
under contract No 22.00245

