Vera.ai: VERification Assisted by Artificial Intelligence

Denis Teyssou
Agence France-Presse

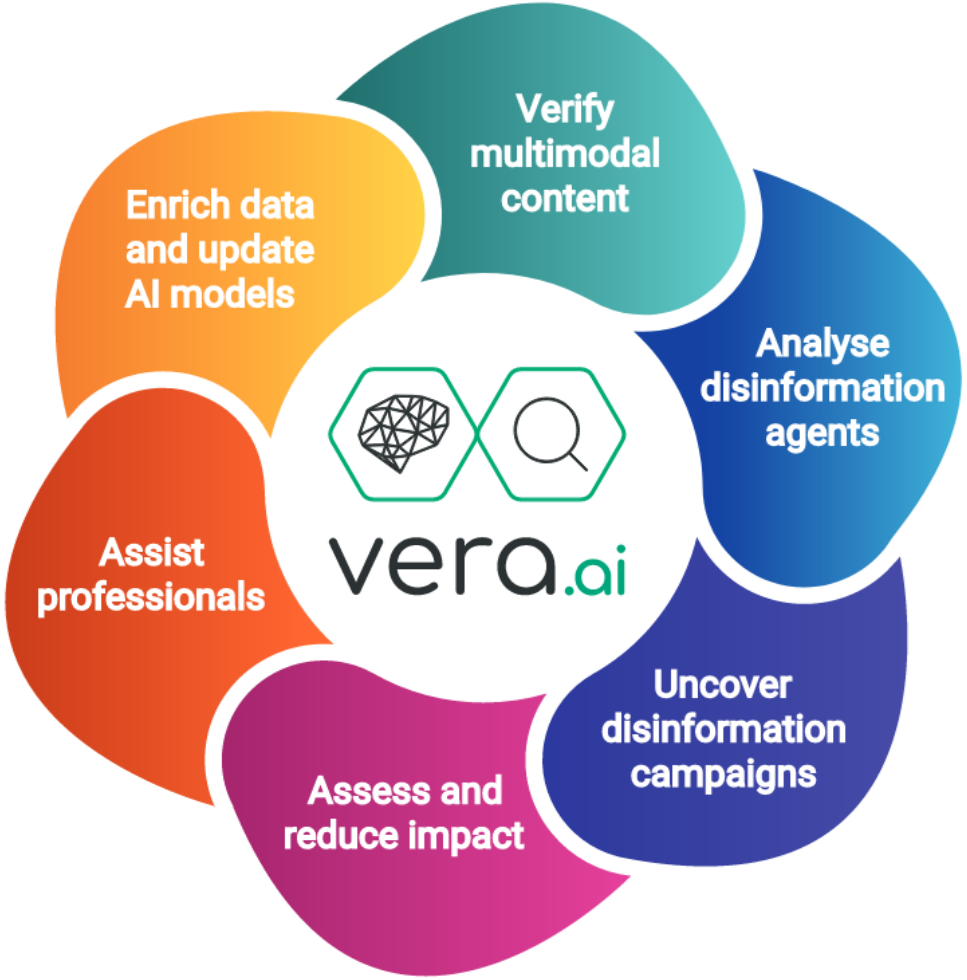
25 October 2022

#Disinfo2022 Conference

vera.ai

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

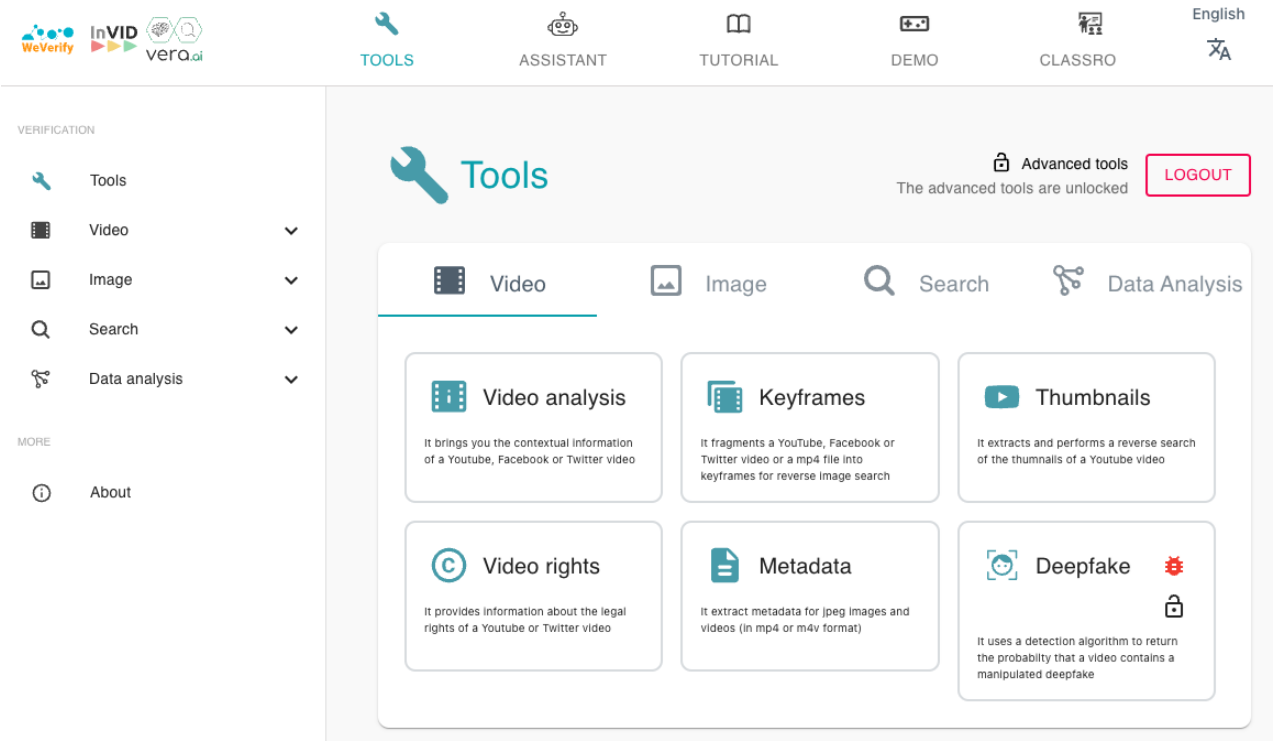
Foundation: WeVerify user-facing tools

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

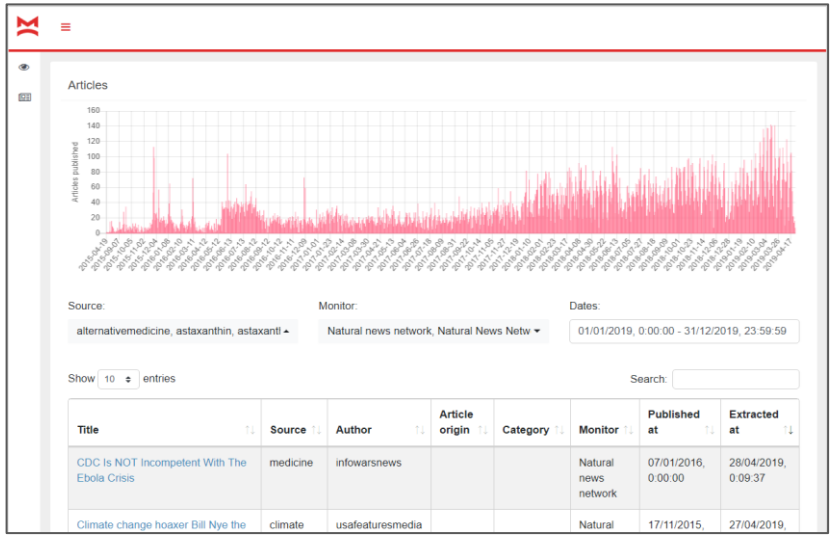
Existing verification features by vera.ai partners

Image Verification Assistant

Detect digital tampering in images.

DeepFake Detection

Check if media contains deepfake manipulated faces

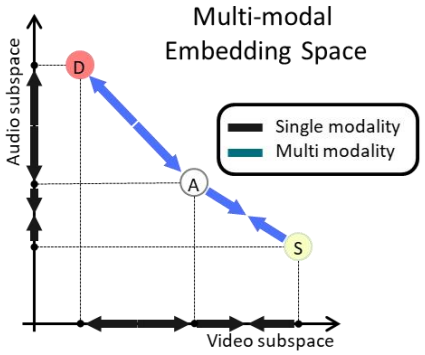


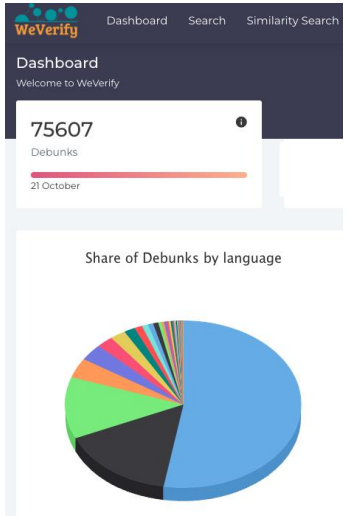
WeVerify OCR Service

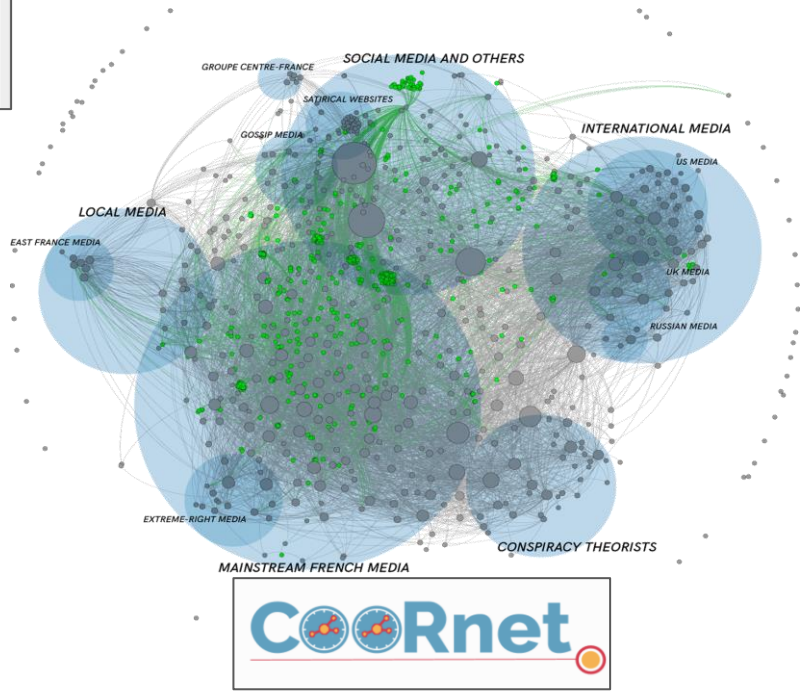
This pipeline identifies URLs in the input text and tries to process them with optical character recognition to identify text in images.

WeVerify

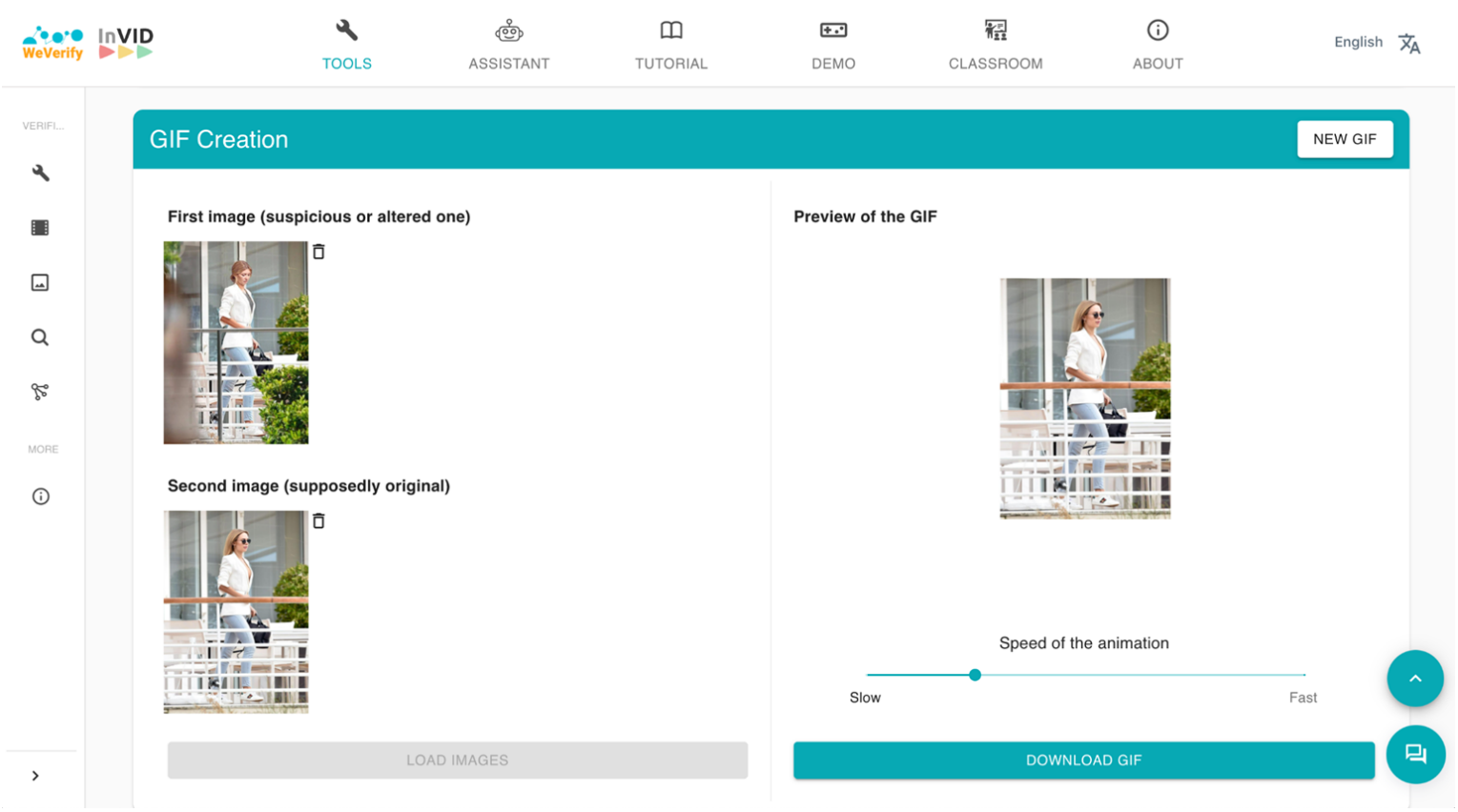
150 free requests / day
Batch processing not available







“Fact-checker in the loop” AI design approach

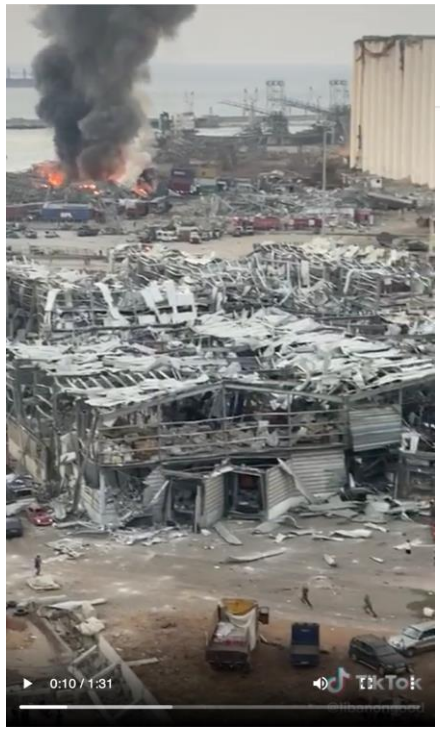


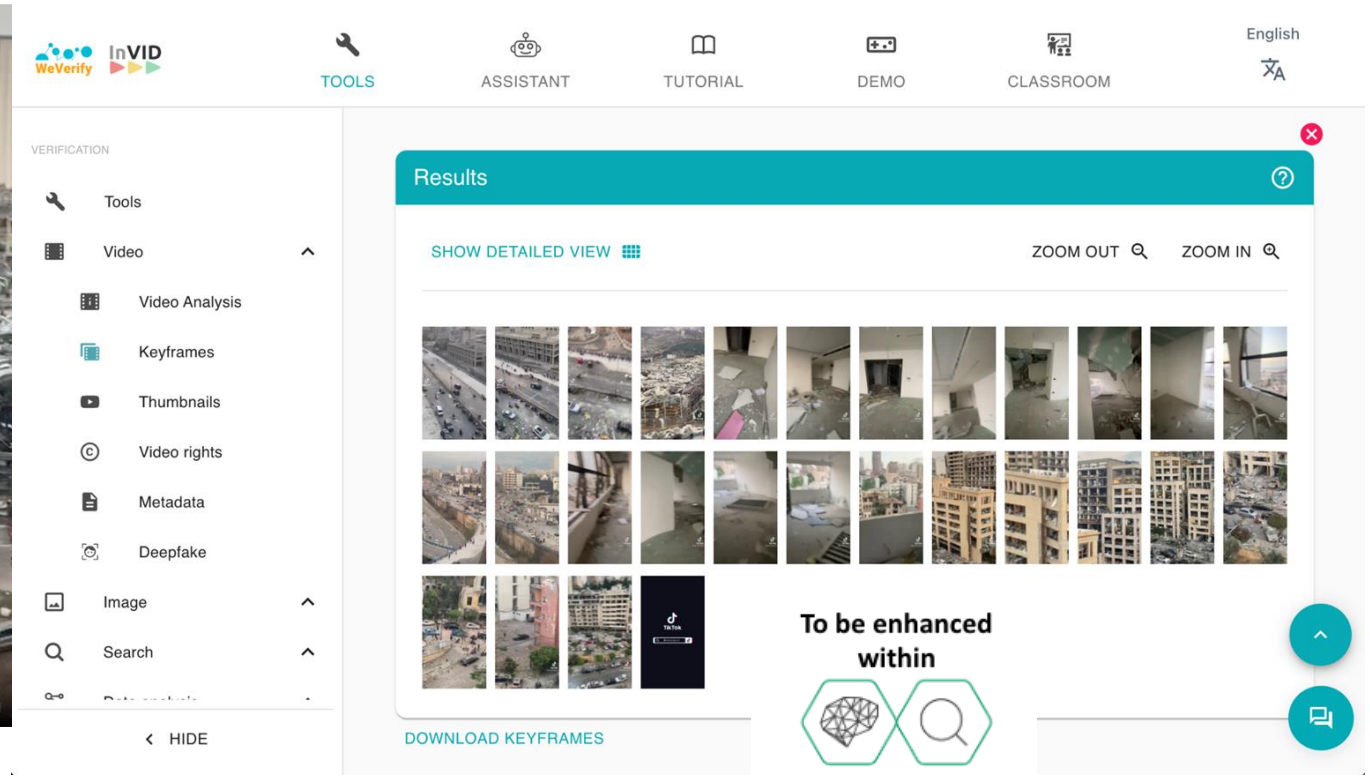
“Debunker in the loop” ⇨ http://u.afp.com/vera-survey

Hide technical complexity and minimise errors

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

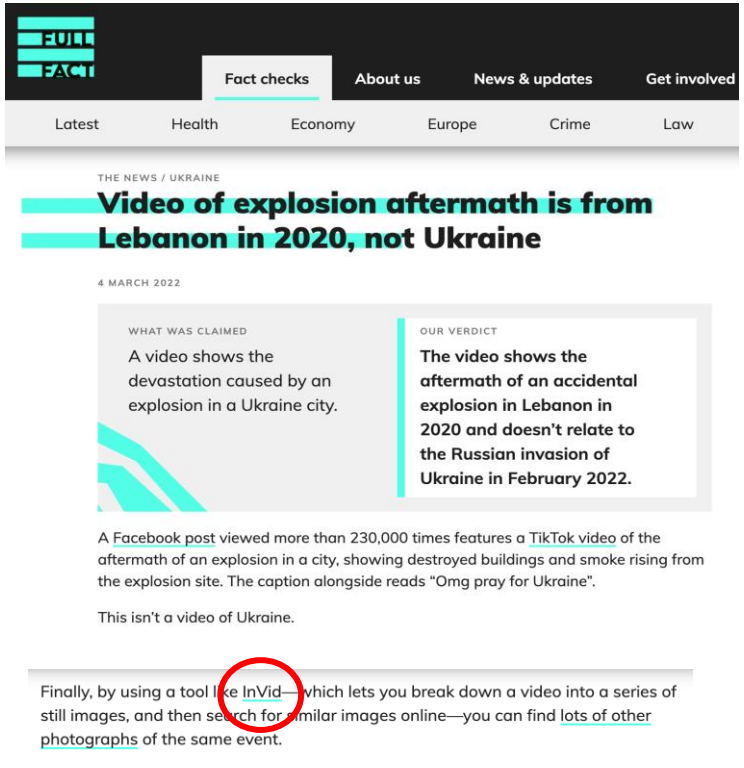
Fact-checkers are using AI computer vision tools





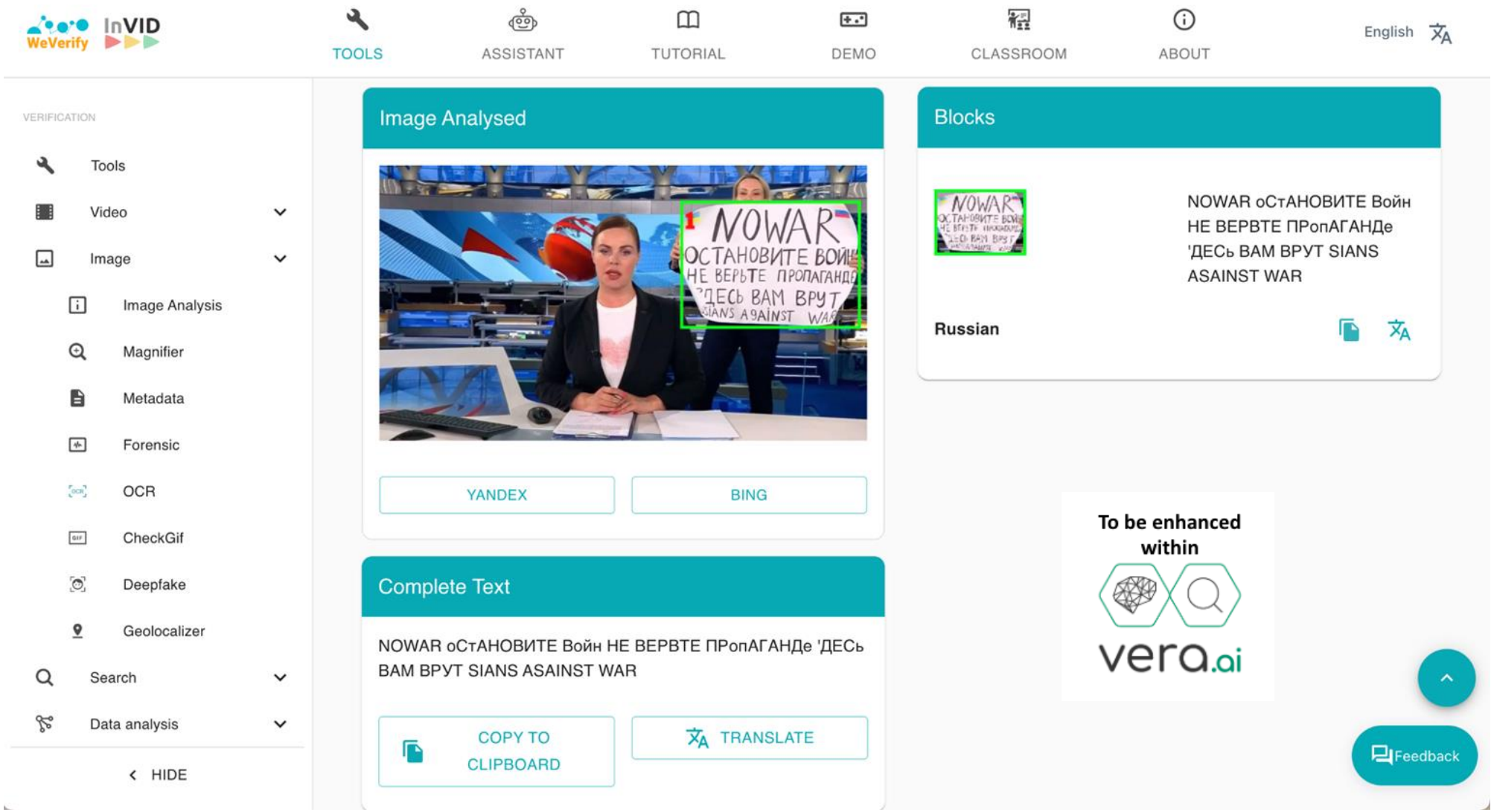
To be enhanced within

vera.ai



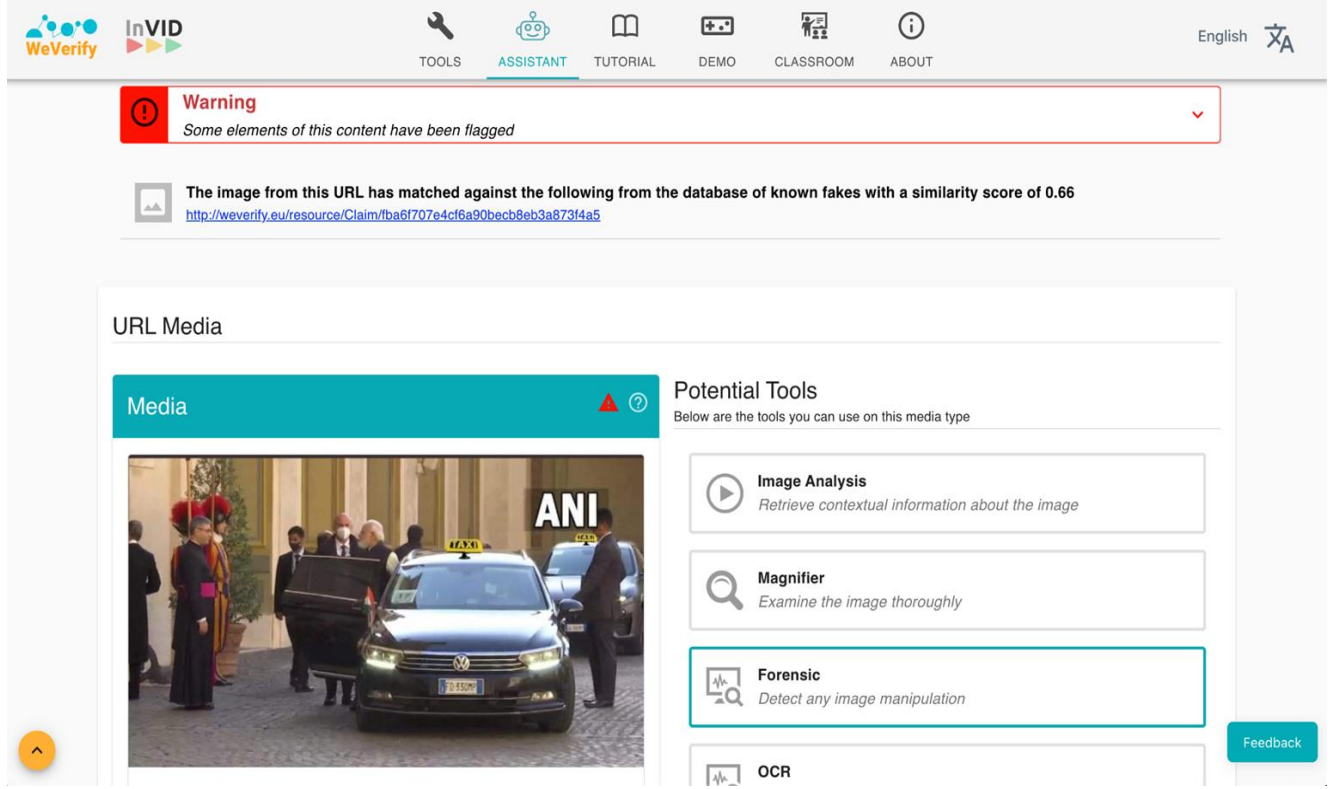
https://fullfact.org/online/video-of-explosion-aftermath-is-from-lebanon-in-2020-not-ukraine/

And also AI multilingual tools



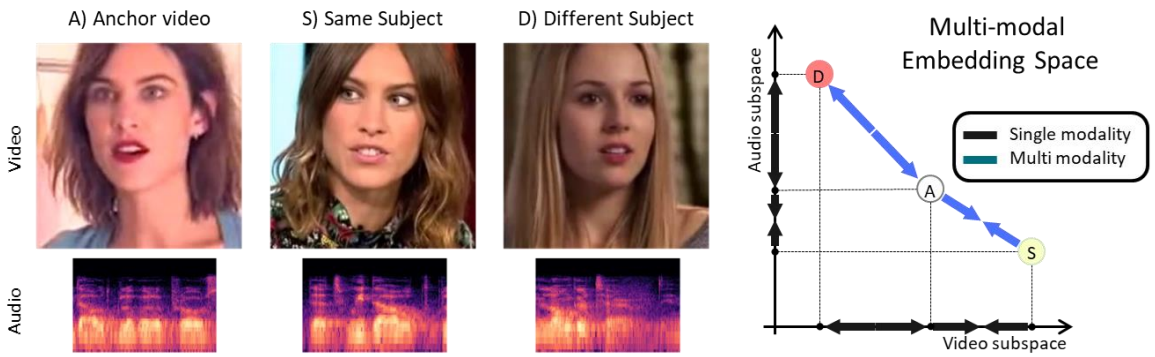
Providing more guidance to professionals

- 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

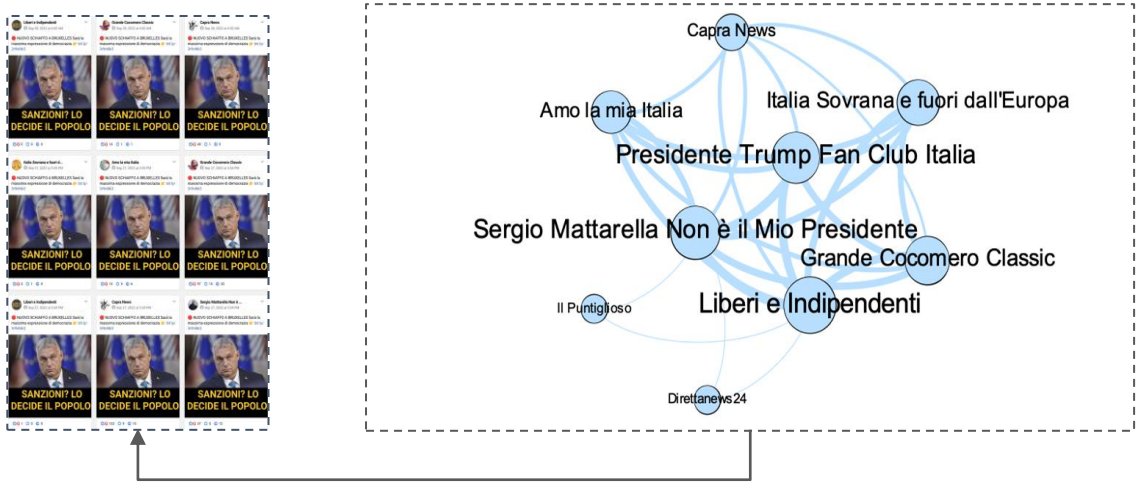


- **Generalization:** much higher error rates on unseen data and new kinds of disinformation
- **Robustness to new data:** platform-specific formats, APIs, metadata
- **Multi-modal and cross-modal detection** (e.g. out-of-context)
- **High-level transparency including explainability and interpretability**

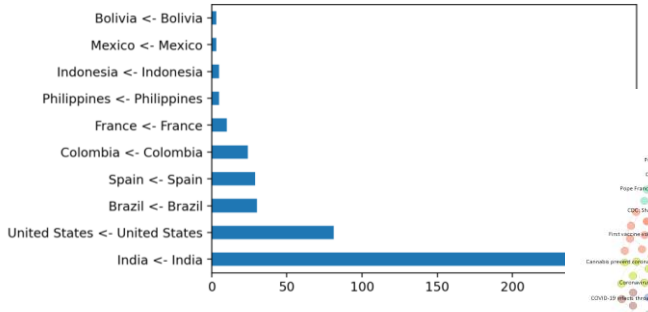
Unsolved challenges ahead

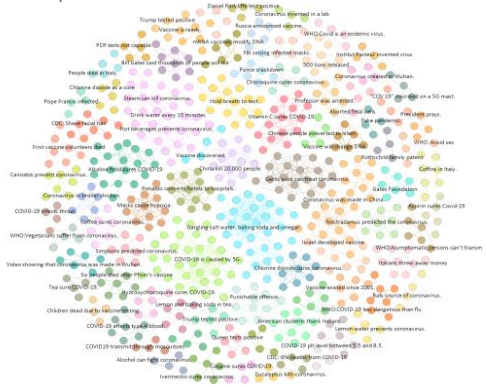


Deepfakes, hybrid deepfakes, audio-visual manipulation

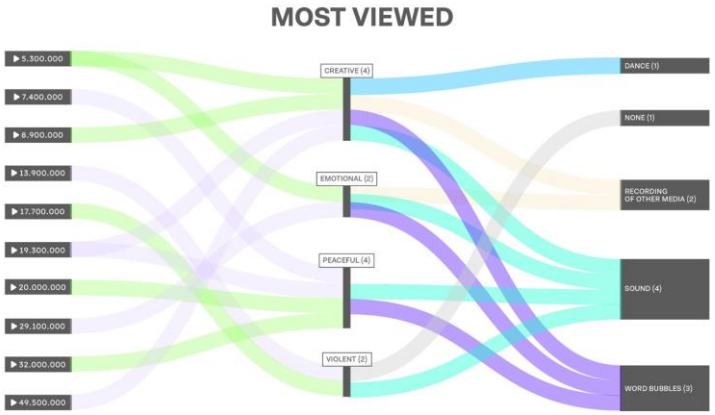


Link sharing behaviour analysis, graph neural networks, node classification, knowledge inference





Spatio-temporal analysis of disinformation campaigns



amplification role of platform algorithms

Thank you! Questions?

vera.ai

Contact: @dteyssou

CERTH CENTRE FOR RESEARCH & TECHNOLOGY HELLAS

The University Of Sheffield.

ATC ATHENS TECHNOLOGY CENTER

KINIT

ontotext

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 DisinfoLab

Fraunhofer IDMT

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