vera.ai results integrated in Truly Media
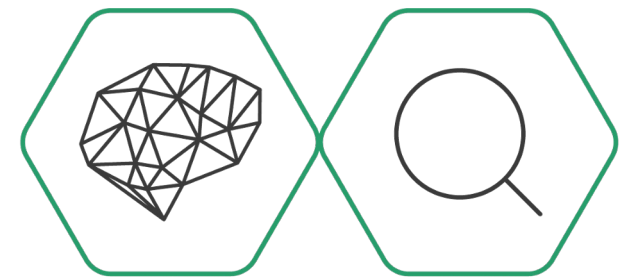
Danae Tsabouraki

ATC

17/06/2025

Final vera.ai event

Truly Media overview

Collect
Look into social media content via keyword searches to detect
interesting topics / trends.

Organize
Create collections with relevant items, assign tasks to members in a seamless work management section for fact checking.

Verify
Select most important items of collection to verify, using checklist and a variety of verification tools.

Truly Media overview

Truly Media overview

vera.ai integrated services

#1 Deepfake Video Detection
CERTH-ITI

Video

#5 Synthetic Text Detection
KInIT

Text

#2 Keyframe Selection & Enhancement
CERTH-ITI

Video

Audio

#5 Synthetic audio detection
FhG-IDMT

#3 Synthetic Image Detection
CERTH-ITI/UNINA

Image

Vera.ai demonstrator

TRULY media

Synthetic text detection

vera.ai

Synthetic video detection

Deepfake Analysis ✕

DeepFake Video Detection

This video tool aims to detect whether the faces that are present in this media have been either manipulated or newly generated by the use of AI-powered Deep Learning methods.Faces shown with intense red color indicate higher manipulation probability.

AI-Generated Face Detection ⓘ

Analyzed video contains faces that are **FAKE**

Overall probability: 99%

Analysis duration: 9.26s

Analyzed shots

Shot #1 is **FAKE**

Shot duration: 0 - 28s

1 extracted face(s)

Face #1 is **FAKE**

Probability: 98%

View this face in different shots

AI Face Reenactment Detection ⓘ

Powered by **CERTH-ITI** and **UNINA** in

User Guide

This document aims to increase the transparency and understandability of the Deepfake Detection Service for end-users by offering insights into how the AI-driven results are delivered, which are the limitations of the service, and how users should interpret the generated results. It provides answers to the questions listed below that users may have, especially when using the service for the first time. The information has been provided by ATC, the operator of the Truly Media platform, in cooperation with DW, a media organisation that co-developed Truly Media and uses the platform.

1. How should I use the service?
2. How should I interpret the results of the service?
3. Which are the limitations of the service?
4. How can I ensure that the analysis is optimal?
5. Why does the service generate false positive/negative results?
6. How much time does the service need to analyse an image/ video?
7. How accurate are the service's results?

How should I use the service?

The Deepfake Detection Service should be used to support your work and the analysis results should be taken into consideration in combination with other factors and indicators (e.g. credibility of the source and other contextual information), to come to a decision about the veracity of the analysed image/video. It is not the aim of the service to provide a yes or no answer to the question of whether a face in an image or video is deep-faked, rather to provide an indication of the probability of such face manipulations. Your own judgement should always be used to come to a final decision.

How should I interpret the results of the service?

Keyframe Selection and Enhancement
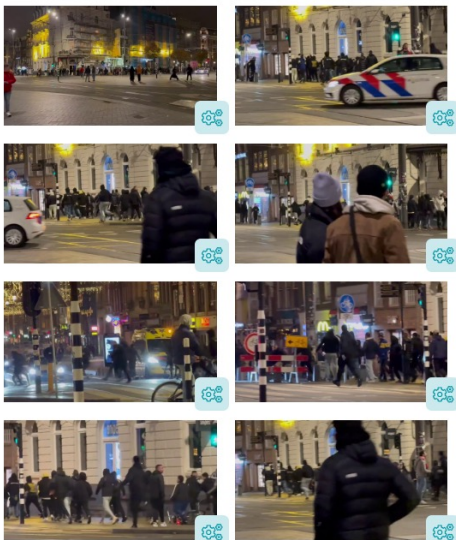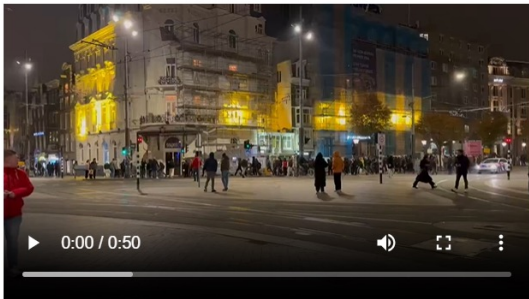
Keyframe Selection and Enhancement

Keyframe selection and enhancement

This experimental AI-powered service automatically extracts representative keyframes from a video. The service also identifies and improves the resolution of faces and text detected within the keyframes for better clarity.

Selected keyframes

0:00 / 0:50

Detected text regions

THE WHITE HOUSE
4:16 PM

> BREAKING NEWS
PRESIDENT BIDEN TELLS STORY ABOUT MAGICAL PISTACHIO
MSNBC
DEADLINE WHITE HO

Synthetic image detection

Synthetic audio detection

vera.ai

The detection output supports you in the assessment of **which segments of the audio file of interest are more likely to be synthetic speech** and **which are more likely to to be natural speech.**

Please keep in mind that **there is always the chance of false positives. Results provide suggestions,** not definite answers!

Do you need help to interpret the detection output? You can find a full explanation in our support page.

LLR: 9.58   LLR: 26.85   LLR: 25.11   LLR: 36.68   LLR: -0.26   LLR: 6.30

0   5

Color legend:

The more intense the blue color, the more likely the audio is natural.

If the color is white or very light, the analysis is inconclusive.

The more intense the orange color, the audio is synthetic.

Detection Output Interpretation ✕

What are LLRs?

The model produced one log-likelihood ratio (LLR) for each 2 seconds of input audio, represented by the colored squares in the waveform. These describe **if the detected traces are more likely to appear in synthetic speech, or in natural speech.**

How can I interpret the LLR values, and the colors provided in the waveform?

In our model, **positive LLRs** support the hypothesis that the speech in the audio **is likely to be created artificially**.

- If 0<LLR≤1, the findings provide weak support for the speech in the audio to be created artificially,
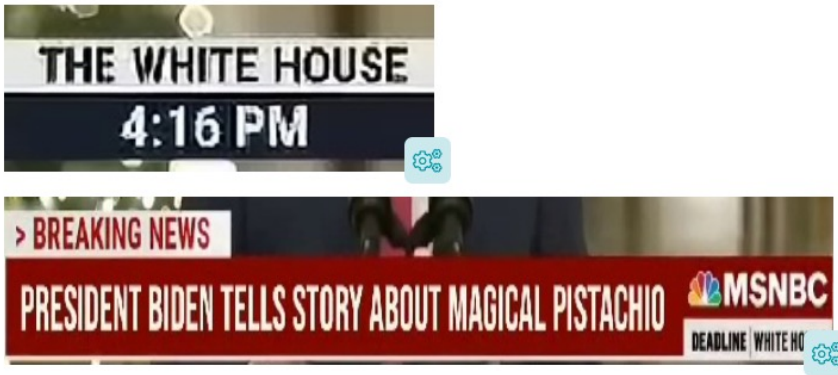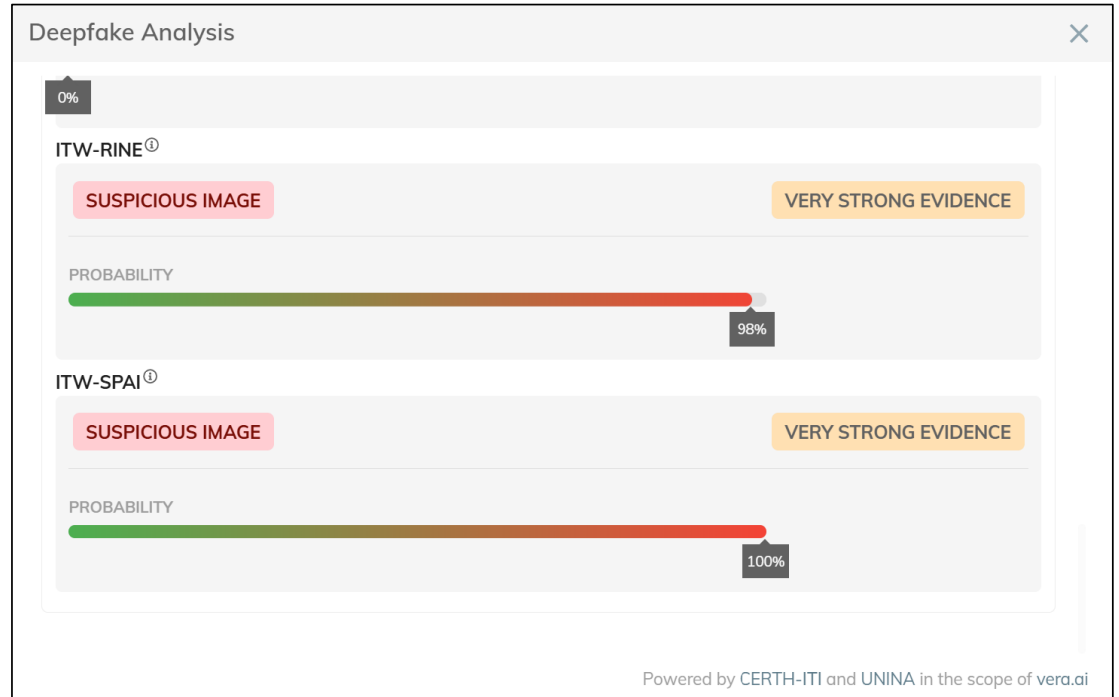- If 1<LLR≤2, the findings provide moderate support for the speech in the audio to be created artificially,
- If 2<LLR≤3, the findings provide moderately strong support for the speech in the audio to be created artificially,
- If 3<LLR≤4, the findings provide strong support for the speech in the audio to be created artificially,
- If LLR>4, the findings provide very strong support for the speech in the audio to be created artificially,

Conversely, **negative LLRs** support the hypothesis that the speech in the audio **is likely to be natural**.

- If -1≤LLR<0, the findings provide weak support for the speech in the audio to be natural,
- If -2≤LLR<-1, the findings provide moderate support for the speech in the audio to be natural,
- If -3≤LLR<-2, the findings provide moderately strong support for the speech in the audio to be natural,
- If -4≤LLR<-3, the findings provide strong support for the speech in the audio to be natural,
- If LLR<-4, the findings provide very strong support for the speech in the audio to be natural,

Finally, **if LLR=0 the analysis is inconclusive**, and the findings do not support any hypothesis.

Close

Insights from user evaluation sessions

Trust, Transparency, and Explainability are crucial

- Users repeatedly emphasized the importance of understanding *why* a result was produced - whether for synthetic text detection or keyframe selection.

- Lack of explanations or traceability of results leads to distrust, even if the outputs are technically correct.

Ease of use and clarity matter

- Tools should work *out-of-the-box* for journalists and media professionals.

- Complex terms, poor naming, hidden character limits, or vague UI cues significantly reduce user confidence and usability.

Users expect contextualized results

- Output must be actionable and integrated with relevant information: timestamps, visual cues, contextual annotations, or model uncertainty.

Lessons learned

Explainability drives trust

- Simply showing a percentage or a tag ("AI-generated") is insufficient. Users want explanations like: "This sentence is likely synthetic because it contains repetitive phrasing or lacks temporal grounding."
- Trust was lost when results contradicted user expectations.

Need for granularity and control

- Users want to control which parts of the data are analyzed or enhanced to save time and processing power.

Partial implementation undermines value

- Incomplete workflows lead to confusion and underuse of otherwise useful tools.

Internal testing adds value before broad user rollout

- Internal tests focusing on UI/UX and trustworthiness helped anticipate external feedback and fine-tune critical features.

Danae Tsabouraki/ ATC
Contact: d.tsabouraki@atc.gr

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/