vera.ai: VERification Assisted by Artificial Intelligence

D5.3 – AI-enhanced Verification Tools for Professionals/DBKF final release and integration with AI platforms

Project Title	vera.ai
Contract No.	101070093
Instrument	HORIZON-RIA
Thematic Priority	CL4-2021-HUMAN-01-27
Start of Project	15 September 2022
Duration	36 months

vera.ai is a Horizon Europe Research and Innovation Project co-financed by the European Union under Grant Agreement ID: 101070093, an Innovate UK grant 10039055 and the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 22.00245.

Deliverable title	AI-enhanced Verification Tools for Professionals/DBKF final release and integration with AI platforms
Deliverable number	D5.3
Deliverable version	1.0
Previous version(s)	D5.2
Contractual Date of delivery	14.09.2025
Actual Date of delivery	15.09.2025
Nature of deliverable	Demonstrator, pilot, prototype
Dissemination level	Public
Partner Responsible	ATC
Author(s)	Denis Teyssou, Valentin Porcellini, Bertrand Goupil (AFP), Kalina Bontcheva, Olesya Razuvayevskaya (USFD), Christina Vaia, Danae Tsabouraki, Charis Bouchlis (ATC), Nedelina Mitankina (ONTO)
Reviewer(s)	Jochen Spangenberg (DW), Ivan Srba (KInIT)
EC Project Officer	Peter Friess

Abstract	D5.3 reports the final releases of vera.ai's end-user tools and documents how research outcomes from WP3/WP4 were transferred into production within WP5. The report details per-tool developments, evaluation-driven improvements, and integration choices, alongside challenges encountered in real-world deployment.

Keywords	user-facing tools, verification, integration, user evaluation, user feedback

Copyright

Revision History

Version	Date	Modified by	Comments
V0.1	01/07/2025	Danae Tsabouraki (ATC)	Draft ToC
V0.2	15/07/2025	Nedelina Mitankina (ONTO)	Input to section 5
V0.3	20/07/2025	Denis Teyssou, Valentin Porcellini, Bertrand Goupil (AFP)	Input and update to section 2
V0.4	12/08/2025	Kalina Bontcheva, Olesya Razuvayevskaya (USFD)	Input to section 3
V0.5	26/08/2025	Danae Tsabouraki, Charis Bouchlis (ATC)	Input to section 4
V0.6	27/08/2025	Danae Tsabouraki (ATC)	General editing and formatting
V0.7	29/08/2025	ATC, ONTO, AFP, USFD	Final checks and partner reviews
V0.8	03-06/09/2025	Jochen Spangenberg (DW), Ivan Srba (KInIT)	Internal review comments and feedback
V0.9	10/09/2025	WP5 partners	Addressing internal review feedback
V1.0	15/09/2025	Olga Papadopoulou, Symeon Papadopoulos (CERTH)	Deliverable sent to EC

Glossary

Term	Definition
ADM	Ablated Diffusion Model
AIoD	AI-on-Demand platform (EU)
API	Application Programming Interface
ArgoCD	GitOps-based deployment automation tool
AWS	Amazon Web Services (cloud platform)
C2PA	Coalition for Content Provenance and Authenticity
CI/CD	Continuous Integration / Continuous Deployment
CloudFront	AWS Content Delivery Network
CooRTweet	Open-source library for coordinated-sharing analysis
DBKF	Database of Known Fakes
D3lta	Open-source library for duplicate-verbatim detection
ELG	European Language Grid
EXIF	Exchangeable Image File Format (metadata)
ExifTool	Tool/library to read EXIF metadata
EKS	Amazon Elastic Kubernetes Service
Heroku	Platform-as-a-Service for deploying applications
IFCN	International Fact-Checking Network
Karpenter	Kubernetes cluster autoscaling tool
KPI	Key Performance Indicator
KSE	Keyframe Selection & Enhancement (service)
Kubernetes (K8s)	Container orchestration system
LDM	Latent Diffusion Model
LLM(s)	Large Language Model(s)
LLR	Log-Likelihood Ratio (used in Synthetic Audio Detection UI)
Matomo	Open-source web analytics platform
MGT	Machine-Generated Text (service)
MongoDB Atlas	Managed MongoDB cloud service
MUI	Material UI component library
NDD	Near-Duplicate Detection (service)
OSINT	Open-Source Intelligence
ProGAN	Progressive Generative Adversarial Networks
RSS	Really Simple Syndication (web feed format)
S3	Amazon Simple Storage Service
SLA	Service Level Agreement
SNA	Social Network Analysis
Solr	Apache Solr search platform
Terraform	“Infrastructure as Code” provisioning tool
UI/UX	User Interface / User Experience
UGC	User-Generated Content

URL	Uniform Resource Locator (web address)
VK	VKontakte (social network)
WP	Work Package

Table of Contents

Index of Tables

Index of Figures

Executive Summary

This deliverable accompanies the final release of vera.ai's user-facing tools, providing details about their integration and technical developments, from M19 of the project until its conclusion, in the scope of WP5. The deliverable focuses on the following tools:

- The InVID-WeVerify-vera.ai Verification Plugin[1] (also referred to as "the Verification Plugin" or "the Plugin") – Main vera.ai partner developing the tool: AFP
- The Verification Assistant (also referred to as "the Assistant") – Main vera.ai partner developing the tool: USFD
- Truly Media[2] – Main vera.ai partner developing the tool: ATC
- The Database of Known Fakes[3] (also referred to as "DBKF" or "the database") – Main vera.ai partner developing the tool: ONTO

Over the reporting period, the end-user tools matured along three axes: (i) systematic integration of WP3/WP4 results into real workflows, (ii) explainability-first UI/UX patterns shaped by participatory evaluations, and (iii) tighter interoperability across independently deployed services. Together, these advances aim to expand modality coverage (text, image, video, audio), reduce users' cognitive load, and create smoother connections between tools.

Across tools, development was guided by continuous feedback from professional users and by active, hands-on collaboration between industry and research partners from the outset. This "lab-to-market" approach aimed to ensured that integrations targeted concrete newsroom needs - reliability, interpretability, and speed - rather than model capacity alone. The rapid emergence and mainstreaming of large language models further underscored the need for real-world data, robust governance, and agile iteration when deploying AI in production.

Concretely, the Verification Plugin broadened and stabilised cross-modal detection while introducing staged release practices (evaluation → beta → public) to manage false positives and model drift. The Verification Assistant consolidated multimodal evidence in a single guided workflow and strengthened extract-and-analyse paths for modern, JavaScript-heavy platforms. Truly Media integrated multiple AI services into collaborative verification workflows, adding resilience measures (polling, caching) and usability features that make complex outputs more readable at a glance. The DBKF expanded its role as a backbone for discovery of prior debunks, improving coverage and multilingual access and wiring its results directly into other tools.

The work also surfaced a common set of challenges and mitigations:

- **Explainability and trust as product features**: Partners tried to enhance explainability by progressive disclosure, consistent labelling, per-segment/timeline visual cues, and contextual "how-it-works" guidance, prioritising actionable interpretation over raw model additions.

[1] https://u.afp.com/plugin
[2] https://www.truly.media/
[3] https://weverify-demo.ontotext.com/

- **Process for quality control**: Staged rollouts and telemetry helped detect drift early and contain the impact of immature models; governance proved as important as modelling.
- **Platform volatility**: Frequent changes to social platforms demanded automated extractor tests, rapid hot-fix capacity, and user fallbacks (e.g., local uploads).
- **Scale and dependencies:** Latency, black-box behaviour, and downtime in partner modules required monitoring, caching, clear ownership, and tight collaboration among partners.
- **Human-AI collaboration**: Integrations were designed to augment expert judgement, not replace it, ensuring that outputs fit existing editorial workflows.
- **Data availability and maintenance**: The value of the DBKF and text-analysis features is bound by source coverage, metadata quality, and upstream API stability, reinforcing the importance of resilient data pipelines.

Looking beyond the project, partners commit to sustaining and evolving the tools. The Verification Plugin will continue as a maintained, open-source front end, with planned work on provenance/watermarking cues, richer Social Network Analysis (SNA) (including coordinated-behaviour views), narrative analysis and video highlighting, local history and shareable links, renewed Firefox support, and Telegram-to-RSS conversion. The Verification Assistant will pursue funding for continued maintenance, with near-term focus on refactoring, Python upgrades, and robust extraction from modern web apps. Truly Media, already in use by major newsrooms and EDMO partners, will integrate mature research outputs under appropriate licensing, formalise service/SLA models for partner components, and complete a platform redesign to better accommodate these integrations. The DBKF will sustain a public demonstrator and re-platform its chatbot stack to mitigate upstream LLM/API changes while seeking further growth via EC projects and commercial avenues.

Finally, the deliverable sets out cross-cutting recommendations for future work: keep explainability central; establish mechanisms for staged evaluation with feedback loops; design for dependency risk and graceful degradation; and invest in multilingual, well-governed data pipelines.

In sum, the work as part of WP5 activities has succeeded in delivering a robust, interoperable selection of tools and services that demonstrably respond to user requirements and codify practices - explainability, staged releases, and resilience - that will keep these tools reliable in production and impactful beyond the project's lifetime.

1 Introduction

The present deliverable summarises the work and outcomes achieved in WP5 over the last reporting period (M19 – M36) and continues the work that was presented in D5.2 "AI-Enhanced Verification Tools for Professionals v1 and DBKF Interim Release". The user-facing tools described in this deliverable act as bridges between the research on AI-based tools from vera.ai's WP3 and WP4 and the final output delivered to end-users. This is realised through integrating the AI components into verification platforms. In this way, target users' workflows can benefit from the novel AI methods for detection and analysis of misinformation content across modalities (text, image, video, and audio), as well as from the advanced techniques for detecting disinformation campaigns and narratives.

The **KPIs related to the user-facing tools** correspond to those defined for each individual component and are presented in detail in the respective WP3 and WP4 deliverables. In addition, one KPI is directly connected to WP5, "*New or improved AI components integrated in Truly Media and the Verification Plugin*" with target value >10. At the end of Year 2, the KPI target value had already been met, with a total of 12 components newly integrated or existing integrations enhanced in the user-facing tools. By the end of the project, the KPI target value was again exceeded by double, with a total of **23 unique components integrated and a total of 30 integrations** (some components have been integrated in multiple tools).

1.1 Verification Tools in the Big Picture of vera.ai

Considering the high-level vision of the vera.ai project, verification tools are designed primarily to **support professionals** by incorporating and providing access to a range of verification tools, helping them in different stages of their workflows (see Figure 1).

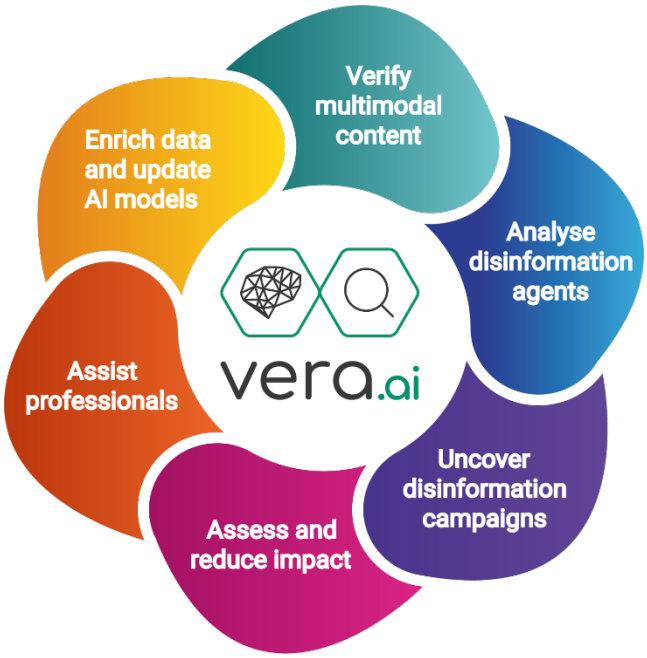

Figure 1 High-level concept of vera.ai project

In terms of multimodal verification, the **Verification Plugin** offers capabilities in different modalities. In vera.ai, the emphasis lies on integrating AI-based features aiming to detect mainly AI manipulations, as well as support social network analysis and geolocation of visual content. The **Verification Assistant,** integrated into the Plugin, has integrated the six services developed in T3.1. These services detect genre, framing, persuasion techniques, subjectivity, machine-generated text, and previous fact-checks in the content being verified. In addition, the Verification Assistant has extended support for the verification of content from YouTube and URL domain analysis. **Truly Media,** in turn, facilitates verification through content collections that include photos, videos, social media posts, and other relevant content, following specific verification workflows integrated in the platform. The **DBKF** then provides the capability to search for debunks from credible sources, a primary one being signatories of the International Fact-Checking Network (IFCN)[4]. These sources are considered trustworthy because IFCN signatory organisations are committed to maintaining fact-checking excellence. Users can conduct searches using keywords, named entities, and general concepts, as well as visual similarity searches.

During the final reporting period, the project's user-facing tools have advanced along three lines: (i) integration of WP3/WP4 research outputs[5], (ii) stability and UX upgrades driven by user evaluation, and (iii) tighter interoperability across tools. Together, these updates expand coverage across modalities (text, image, video, audio), improve explainability, and streamline verification workflows for professionals.

A summary of the main developments for each tool spanning the last reporting period follows.

The Verification Plugin: Since March 2024 the plugin has seen ten releases, culminating in v0.87 (August 15, 2025). Highlights include a full integration of the Keyframes Selection and Enhancement service (KSE)[6], a gauge-based Synthetic Image Detection UI with per-model scores, Near-Duplicate Detection (NDD) to surface look-alike images, a charted Deepfake Video analysis, Geolocation improvements (including local file upload), and a standalone Machine-Generated Text (MGT) tool for controlled testing. New detector families from CERTH/UNINA[7] have been staged through evaluation/beta roles before broader roll-out. The NDD feature stores strong detections (≥70%) for reuse against future submissions, improving recall for look-alike content. Beyond media analysis, the team rebuilt Social Network Analysis to address Twitter/X's and CrowdTangle's post-API deprecation, adding coordinated-sharing exploration via CooRTweet methods, developed in the context of WP4,[8] and duplicate-verbatim detection with D3lta; infrastructure and CI/CD pipelines were also modernised.

The Verification Assistant: The Assistant broadened its "guided verification" scope with new partner services and explainable text analytics. Additions include concept-enrichment (named-entity word clouds), YouTube top-comments ingestion paired with a multilingual stance classifier, URL domain analysis

[4] https://www.poynter.org/ifcn/

[5] These outputs are detailed and described in WP3 and WP4 deliverables, which are accessible through the project's website here: https://www.veraai.eu/deliverables

[6] For a comprehensive description of the KSE, refer to Deliverable 3.3 Cross-modal and user feedback enhanced verification tools.

[7] For a comprehensive description of these methods, refer to Deliverable 3.2 Multimodal deepfake and manipulation analysis.

[8] For a comprehensive description of these methods, refer to Deliverable 4.2 Coordinated sharing behaviour detection and disinformation campaign modelling methods.

with source warnings, and a conversational helper wired into the Synthetic Image Detection interface for documentation/feedback queries. It now aggregates six credibility signals - news framing, genre, persuasion techniques, subjectivity[9], machine-generated text[10], and previous fact-checks - with controls and highlights to support user interpretation. Integration with the DBKF enables automatic surfacing of prior fact-checks matching the analysed page.

Truly Media: The platform integrated three WP3/WP4 services: KInIT's MGT detection (whole text or excerpts), CERTH's Keyframe Selection and Enhancement (KSE) for representative frames with face/text focus, and IDMT's Synthetic Speech Detection[11] with interactive probability waveforms and in-app guidance. Iterative adjustments and improvements to the integrated services following feedback from evaluation sessions ensured that the integrated tools have a stronger alignment with user expectations and needs. Technical work included stateless back-end integration, result caching, asynchronous task polling, and a dedicated scraping microservice for clean text. These UI/UX refinements aim to improve consistency with model outputs and reduce cognitive load for users. In addition, Truly Media added a secure video retrieval/audio-extraction service for third-party hosts and collaboration upgrades such as checklist locking and multi-assignee support to better fit newsroom teamwork.

Database of Known Fakes (DBKF): The DBKF expanded its corpus and capabilities while strengthening its role as a cross-tool backbone. Content growth leveraged the Google FactCheck Explorer API and custom ingestion, bringing the searchable claims total to ~190k with broader IFCN/EDMO coverage and improved language/geography diversity. New exploration features include Narrative Cluster Search (multilingual clustering with timeline/visualisation views) and a conversational DBKF Chatbot that exposes search functions via natural language with inspectable reasoning traces. Event-type search (36 types in English) and document-type facets (claims, claim reviews, social posts, news) further deepen analysis. Infrastructure upgrades accompanied these additions. DBKF services are now consumed directly by the Assistant to flag similar debunks during webpage analysis.

1.2 Verification Tools and User Requirements

The verification tools enhanced within the scope of vera.ai closely align their technical developments with the needs, translated into requirements, of media professionals and disinformation researchers as described in deliverable D2.1 - AI against Disinformation: Use Cases and Requirements[12]. Taking a closer look into some of the prominent needs of fact-checkers and researchers in terms of instruments and functionalities as outlined in D2.1, the following aspects can be highlighted:

- high demand for tools able to detect AI-generated videos and deepfakes;
- capabilities to investigate the spread of false claims, especially where such claims can be identified with keywords;

[9] For a comprehensive description of these methods, refer to Deliverable 3.3 Cross-modal and user feedback enhanced verification tools.
[10] For a comprehensive description of the machine-generated text detection method, refer to Deliverable 3.2 Multimodal deepfake and manipulation analysis.
[11] For a comprehensive description of this methods, refer to Deliverable 3.2 Multimodal deepfake and manipulation analysis.
[12] No reference link to D2.1 could be provided, as it is a non-public deliverable.

- detection of re-emerging disinformation (also known as 'zombie stories');
- access to a comprehensive database of fact-checks.

In addition, users strongly emphasise the importance of explainability and transparency. In other words, tools should alleviate the burden of output interpretation for users by effectively "translating" technical explanations into easily understandable language. In this way, users will have a clear picture about the reliability of results, as well as about the limitations of the current systems (in case certain functionalities/models are not yet supported). Technical partners in vera.ai were dedicated to implementing and refining explainability features, guided by evaluation feedback from WP2-led activities.

Each verification tool developed within vera.ai addresses different aspects of media professionals' workflows, aligning with the above-mentioned areas. Here is a summary of the tools' capabilities, while more details for each of them can be found in the respective section of this deliverable:

- The InVID-WeVerify-vera.ai Verification Plugin[13] (also referred to as "the Verification Plugin" or "the Plugin") is a kind of "Swiss army knife" toolbox for multilingual and multimodal content analysis, covering various verification steps and services such as image and video forensics, text analysis, deepfake analysis, and more. AFP acts as the core partner that maintains and manages the tool as a whole and coordinates the integration of individual components that make up the plugin.
- The Verification Assistant (also referred to as "the Assistant") guides users of the Verification Plugin through the verification paths, suggests the appropriate tools to be used and explains their results. The main vera.ai partner developing and running the tool is USFD.
- The Truly Media[14] platform supports workflows for collaborative verification of content from social and digital media. The main vera.ai partner developing and maintaining the platform is ATC.
- The Database of Known Fakes[15] (also referred to as "DBKF" or "the database") focuses on parts of the verification workflow that tackle resurfacing or repurposed false narratives with the help of powerful searches and insights from trustworthy debunks. The main vera.ai partner developing and running the tool is ONTO.

Since those verification tools complement each other and cover different use cases, they are available separately as web applications instead of being integrated in a monolithic platform. Nevertheless, smooth interoperability and, when feasible, integration, is ensured with the assistance of the Common Data Annotation model developed in WP5 and presented in D5.1 - Annotation model, API definitions, and Database of Known Fakes first release [16].

Truly Media has addressed user needs and requirements as identified in WP2 activities mainly through the integration of research components coming from WP3, WP4, and WP5. Integration of research outputs into the platform aims additionally to demonstrate how tools developed within vera.ai can be

[13] https://u.afp.com/plugin
[14] https://www.truly.media/
[15] https://weverify-demo.ontotext.com/
[16] https://veraai-cms-files.s3.eu-central-1.amazonaws.com/D5_1_V1_0_820349369d.pdf

integrated in professional platforms and verification workflows. The integrated components are available to the EDMO member community and associated entities through Truly Media's environment.

Table 1 maps the user requirements identified and diligently listed in WP2 against the tools that already have capabilities (or are set to develop those within the scope of the project) to address the respective needs adequately.

Table 1 Mapping user requirements against verification tools

User requirement description	WP5 Tools	How the tools help	Related WPs and partners
Detect and identify artificially generated or manipulated images	Verification Plugin, Truly Media	Detect if an image has been generated through AI	WP3 (CERTH, UNINA); WP5 (AFP, ATC)
Detect and identify audio content that has been generated or manipulated using artificial intelligence techniques	Verification Plugin, Truly Media	Detect AI-generated audio content	WP3 (IDMT); WP5 (AFP, ATC)
Video processing for new platforms (TikTok, Telegram, Instagram, VK, Odysee)	Verification Plugin, Truly Media	Public-facing tools for accessing data platforms	WP3 (CERTH); WP5 (AFP, ATC)
Address new platforms for social network analysis (Telegram, YouTube, TikTok,..)	Verification Plugin, Truly Media	Tools for visualising social network analysis on new platforms	WP5 (AFP, ATC)
Multilingual search of previously debunked articles	DBKF	Multilingual search functionality that went through a first round of participatory user evaluation in February 2024	WP5 (ONTO)
Multimodal search of previously debunked articles	DBKF	Visual similarity currently available thanks to CERTH's NDD, an integrated multimodal approach is yet to be developed for the next release	WP4 (CERTH, ONTO); WP5 (ONTO)
Improve indexation of third-party databases (data collection)	DBKF	Content ingestion and its maintenance are ongoing processes that ensure the timely collection and enrichment of debunks	WP5 (ONTO)
Detect keywords used for "seeding" narratives in multiple languages	DBKF	Multilingual concept search already available and narrative detection to be enabled in a future release (based on clusters of textually similar claims and/or linked concepts)	WP4 (KInIT, ONTO); WP5 (ONTO)

The developments delivered across the end-user tools directly address the user requirements in Table 1 by expanding detector coverage, restoring analysis on new platforms, and improving explainability so results are easier to interpret in real workflows. Table 2 below maps the requirements of Table 1 to the concrete technical developments delivered across the end-user tools.

Table 2 Mapping the requirements of Table 1 to the concrete technical developments delivered across end-user tools

Requirement (Table 1)	Related technical developments
Detect & identify AI-generated/manipulated images	**Verification Plugin:** integrated successive CERTH/UNINA models; added Near-Duplicate Detection (NDD) and clearer detector UI elements (gauge/labels). **Truly Media:** integrated successive CERTH/UNINA models, aligned UI with labels returned by the service and displayed results from additional detectors for consistency and clarity.
Detect & identify AI-generated/manipulated audio	**Truly Media:** integrated IDMT Synthetic Speech Detection with interactive waveform, color-coded segment probabilities, caching/polling; improved interpretability (prominent legend, "how it works" help, enhanced player controls including slow playback and sync).
Video processing for new platforms (e.g. TikTok, Telegram, Instagram, VK, Odysee)	**Verification Plugin:** continued platform support and explicit additions (in KSE and Assistant) to keep ingestion current. **Truly Media:** added a secure video-retrieval & audio-extraction service so third party-hosted media can be downloaded and analysed locally by integrated detectors. *Due to API limitations, CERTH's related component was discontinued, and alternative ways of video processing were explored by AFP and ATC.
Address new platforms for Social Network Analysis	**Verification Plugin:** rebuilt SNA to work despite X/Twitter API deprecation; added **CooRTweet** for coordinated-sharing exploration and **D3lta** for duplicate-verbatim detection; additional infra hardening for reliability.
Multilingual search of previously debunked articles	**DBKF:** strengthened multilingual/translated search; **Assistant** calls DBKF automatically during webpage analysis to flag similar debunks.
Multimodal search of previously debunked articles	**DBKF:** DBKF Similarity Search leverages CERTH's NDD and is integrated in Truly Media and (earlier) the plugin for look-alike content discovery.

Improve indexation of third-party databases (data collection)	**DBKF:** integrated Google FactCheck Explorer API plus custom ingestion, raising the corpus by >50k to ~190k searchable claims with broader IFCN/EDMO and language/geography spread.
Detect keywords used for "seeding" narratives in multiple languages	**DBKF:** developed Cluster (Narrative) Search with multilingual clustering, timelines and concept links; concept search available and used as a facet to explore cross-language narrative evolution.

Moreover, integration efforts were informed by user evaluations from previous releases, leading to updates and changes in integrated components based on evaluation results and feedback. For more details, please refer to the individual subsections titled "*Improvements and updates following evaluation feedback*" in Sections 2 to 5 of the present deliverable, as well as to D2.2 Evaluation Report.

1.3 Structure of the Deliverable

The present deliverable follows a similar structure as D5.2 "AI-Enhanced Verification Tools for Professionals v1 and DBKF Interim Release", to ensure consistency across all the tasks of WP5 and a harmonised approach and presentation of the activities that were implemented for the final release of vera.ai's user facing tools. As such, all tool sections (Sections 2-5) follow the same structure. Each section starts with *Tool Purpose and Overview*, then identifies the *End Users* and outlines the *Technical Infrastructure*. The core part, *Integration Activities and Other Technical Developments*, is split into three consistent sub-sections: (i) *Integration with Partner Services/Tools*, (ii) *Other Technical Developments*, and (iii) *Improvements and Updates Following Evaluation Feedback*. Each section then closes with *Challenges and Lessons Learned* and *Next Steps Beyond vera.ai*.

Section 6 summarises integration with relevant AI platforms, namely the AI-on-Demand (AIoD)[17] and European Language Grid (ELG)[18] platforms. Section 7 provides the conclusion and future steps/developments, presenting concluding remarks concerning the work under WP5. Finally, Section 7 discusses the challenges faced during the development and integration process and the actions that were taken to tackle these challenges to the extent possible. The present report ends with an outlook towards the future and considerations concerning the sustainability of the different tools beyond vera.ai's lifetime.

[17] https://www.ai4europe.eu/

[18] https://live.european-language-grid.eu/grid

2 InVID-WeVerify-vera.ai Verification Plugin – Final Release

This section provides an overview of the Verification Plugin and its technical infrastructure. It details the technical work that took place leading to the final release of the Plugin.

2.1 Tool Purpose and Overview

The InVID-WeVerify-vera.ai Verification Plugin[19] is a browser extension designed as a kind of "Swiss army knife" for information verification (indicating its multi-purpose use). The toolbox was launched in July 2017 during the InVID[20] Innovation Action (2016-2018) to help fact-checkers verify videos and images. It has been enhanced during the Horizon 2020 WeVerify[21] Innovation Action (2018-2021) to provide more capacities, such as social network analysis, an Assistant to guide end users, connection to a Database of Known Fakes (DBKF), fact-checks, and cross-network search applications.

It has been showcased at numerous events such as ICT 2018 in Vienna, at the International Fact-Checking Network's GlobalFact summit in Rome (2018), at SXSW 2019 in Austin, in Global Fact again online (2020, 2021), Global Fact 10-11 in Seoul (2023) and Sarajevo (2024) as well as in the Trusted Media Summit in Singapore (2023), to name a few high profile events.

Over the course of the vera.ai project (2022-2025), the Verification Plugin was further improved and enhanced with the integration of more AI-based tools and services. Since October 2023, the plugin has been enriched with a robust synthetic media detector, an improved Keyframes Enhancement Service and Deepfake video detection, an updated Geolocation detection service, and Credibility Signals (Text analysis: Persuasion techniques and machine generated text detector)[22].

2.2 End Users

The Verification Plugin is used by fact-checkers, journalists, researchers tackling disinformation, law and enforcement agencies, media literacy scholars, OSINT investigators and human rights defenders, among others. According to the Chrome Store data, the Verification Plugin has now reached 150,000 active users per week from 224 countries (as of early September 2025).

[19] https://u.afp.com/plugin

[20] InVID was a Horizon 2020 project, funded under grant agreement No 687786. Project website: https://www.invid-project.eu/

[21] WeVerify was a Horizon 2020 project, funded under grant agreement No 825297. Project website: https://weverify.eu/

[22] A complete description of those services is available in Deliverable D3.2.

2.3 Technical Infrastructure

The user-facing interface is a Google Chrome browser plugin, written in JavaScript using the React.js[23] framework. The plugin is published on the Google Chrome Store and reviewed by Google before any new release. The code is open-source, available on GitHub[24], and welcoming outside collaborations.

The plugin includes built-in tools written in JavaScript (e.g., Magnifier, Reverse Image Search). For more compute-intensive tools the plugin communicates with partners' APIs to utilise their servers. This includes tools like the Deepfake Analysis, Forensic Analysis, and Video Keyframe Fragmentation.

To prevent exposing the APIs' keys and authentication methods, an integration layer is used (own server) written in Java with Apache Camel[25] hosted on AWS by Sheffield University (USFD). AFP is hosting Matomo Analytics on its own AWS instance. Both the integration layer (middleware) and Matomo help the AFP team to collect raw data, monitor events and discover trends in the usage of tools including UX issues. Thanks to the cloud architecture, the infrastructure can easily be ramped up or scaled down automatically as needed. An Elasticsearch[26] server is in place to monitor the usage and the errors occurring on the integration layer. This is particularly useful to determine if a partner's service has suddenly stopped working or if users make mistakes with the tools (e.g., by submitting a post link instead of an image link).

Last but not least, a feedback button in the plugin is linked to the AFP Medialab's Slack channel to receive instant notifications from users. Below is a chart illustrating the plugin's overall architecture with all the integrated vera.ai services (Figure 2).

[23] https://react.dev/
[24] https://github.com/AFP-Medialab/verification-plugin/releases
[25] https://camel.apache.org/
[26] https://www.elastic.co/elasticsearch/

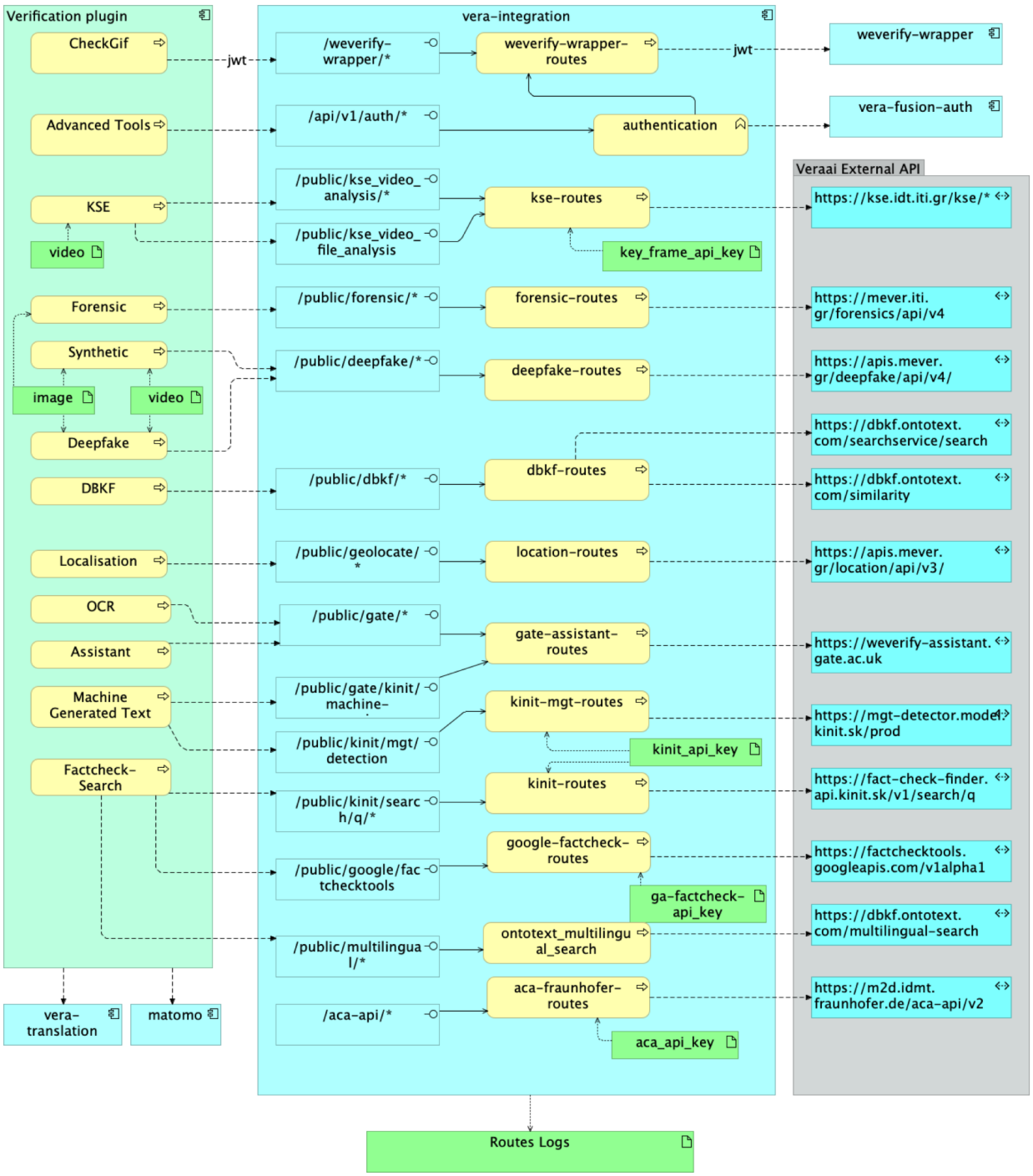

Figure 2 A view of the plugin's architecture from front-end to back-end

2.4 Integration Activities and other Technical Developments

Since March 2024, the AFP Medialab team has released ten new versions of the Plugin, with the latest being version v0.87. The Plugin is updated regularly. Table 3 presents the releases and key updates in chronological order, from the oldest to the newest, while the full release notes are available on GitHub[27].

Version number	Date	Key vera.ai components
v0.79	Apr 5, 2024	Addition of Fact Check Semantic Search engine
v0.80	Jun 5, 2024	Deepfake Image, Deepfake Video: added local file uploads Synthetic Image Detection: added gauge UI and accordion with model scores and detection labels
v0.81	Jun 26, 2024	Synthetic Image Detection, Deepfake Video, Deepfake Image: updated API calls (following API model names changes) Synthetic Image Detection: added gauge scale explanation in the UI
v0.82	Oct 25, 2024	Synthetic Image Detection: added Near-Duplicate Detection results, added local file auto-resizing feature for evaluation Added evaluation role
v0.83	Nov 13, 2024	Synthetic Image Detection: Added local file auto-resizing feature for beta testers
v0.84	Feb 12, 2025	Synthetic Image Detection: added ITW RINE algorithm Keyframes: UI improvements
v0.84.1	Apr 3, 2025	Synthetic Image Detection: added itw_spai_mever[28] algorithm
v0.85	Jun 4, 2025	Deepfake Video: new chart and UI update, added faceswap_fsfm[29] algorithm Keyframes: added new KSE service for evaluation Geolocation: added file upload Synthetic Image Detection: added new algorithms for

[27] https://github.com/AFP-Medialab/verification-plugin/releases

[28] A method for synthetic image detection, called SPAI, described in Deliverable 3.2.

[29] A method for video deepfake detection described in Deliverable 3.2.

		evaluation (sd21_bfree-dino2reg4_grip, multi_bfree-dino2reg4_grip, sd21_bfree-siglip_grip[30]) and added CERTH synthetic image detection labels for evaluation
v0.86	Jun 26, 2025	Addition of standalone Machine Generated Text tool
v0.87	Aug 15, 2025	Added KSE link mode, new C2PA version in Metadata and Synthetic Image Detection, changed deepfake video base algorithm to faceswap_fsfm.

Table 3 Verification Plugin releases and key vera.ai updates

2.4.1 Integration with Partner Services/Tools

Synthetic Media detectors

After careful performance evaluation, AFP has integrated partners' services and tools in a three steps approach when it comes to AI detectors. First, we have integrated partners' new technologies for development testing and debugging, then, we made new tools and services available to all the members of the vera.ai consortium for testing, and finally, we made these new tools available to advanced end users (beta testing users, and registered users). With this approach, we are able to make sure that we release new models only after they have been carefully tested and examined, notably for potential false positives as explained in Deliverable D2.2. Hence, we have integrated new algorithms for the synthetic image detection tool (deepfake API by CERTH, models by CERTH and UNINA) and we are still benchmarking them to select the best performing ones to roll out to users. Table 4 shows the integration date of the synthetic image detection algorithms. The algorithms have a first user role of evaluation (can be tested by the vera.ai members only), "Extra feature" (only available to the AFP team for first evaluation), or "Beta tester" (beta testing users).

Table 4 The Synthetic Image Detection algorithms added to the Verification Plugin

vera.ai models	**User Role**	**Verification Plugin version**	**Publication Date**
sd21_bfree-dino2reg4_grip, multi_bfree-dino2reg4_grip, sd21_bfree-siglip_grip	Evaluation	v0.85	June 4, 2025
itw_spai_mever	Evaluation	v0.84.1	April 3, 2025
itw_rine_mever	Evaluation	v0.84	February 12, 2025
progan-webp_r50_grip, ldm-webp_r50_grip, gigagan-webp_r50_grip	Evaluation	v0.82	October 25, 2024
gan_r50_mever, ldm_r50_mever	Extra feature	v0.81	June 26, 2024
progan_r50_grip, adm_r50_grip, progan_rine_mever, ldm_rine_mever	Extra feature	v0.78	February 9, 2024
gan (later named gan_r50_mever), unina (later named ldm_r50_grip)	Beta tester	v0.77	October 12, 2023

[30] Three variants of the B-free method for synthetic image detection presented in deliverable 3.2.

AFP has also integrated a near-duplicate detection (NDD) feature in the Synthetic Image Detection tool in the Verification Plugin using the NDD API[31] and infrastructure provided by CERTH. This allows users to see the model's detection scores for past images very similar to the image they are working with. When a detection model has a score of 70% or above, we insert the result to the near duplicate database as we consider that there is strong evidence of a detection of a synthetic image. Later, if a similar image is analysed, by the same or by a different user, and has a low detection score, we look for similar images in the near duplicate database and display them if there is a match. An example of how the detected similar images are presented is illustrated in Figure 3.

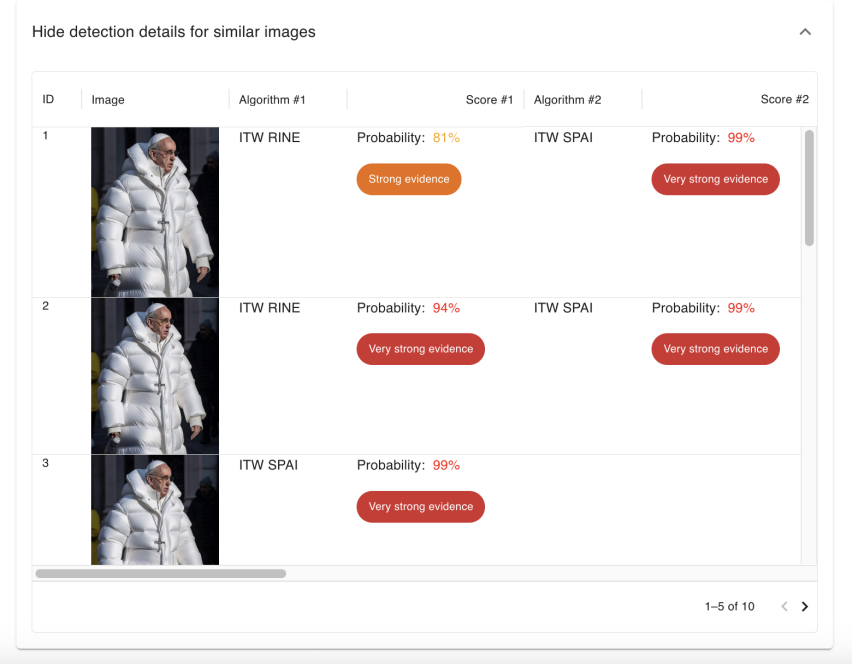

Figure 3 The near duplicate detection results

Geolocation

The improvement of the Geolocation model by CERTH[32] was integrated into the Verification Plugin as well as the support for sending local files was made available by the new API. This allows users to send images to the Geolocation API more easily, even as platforms increasingly protect their data to prevent scraping. While working on this update, we also took the chance to enhance the UI elements and responsiveness. The presentation of the results is now more straightforward to understand as showcased in Figure 4 which illustrates a query image and the detected location side-by-side with the predicted coordinates. We also display all geolocation predictions if there is more than one returned by the API. In the Figure 4, the confidence score is displayed for evaluation purposes only.

[31] A component enhanced in WP4/T4.1 and presented in Deliverable 4.1.

[32] Presented in Deliverable 3.3.

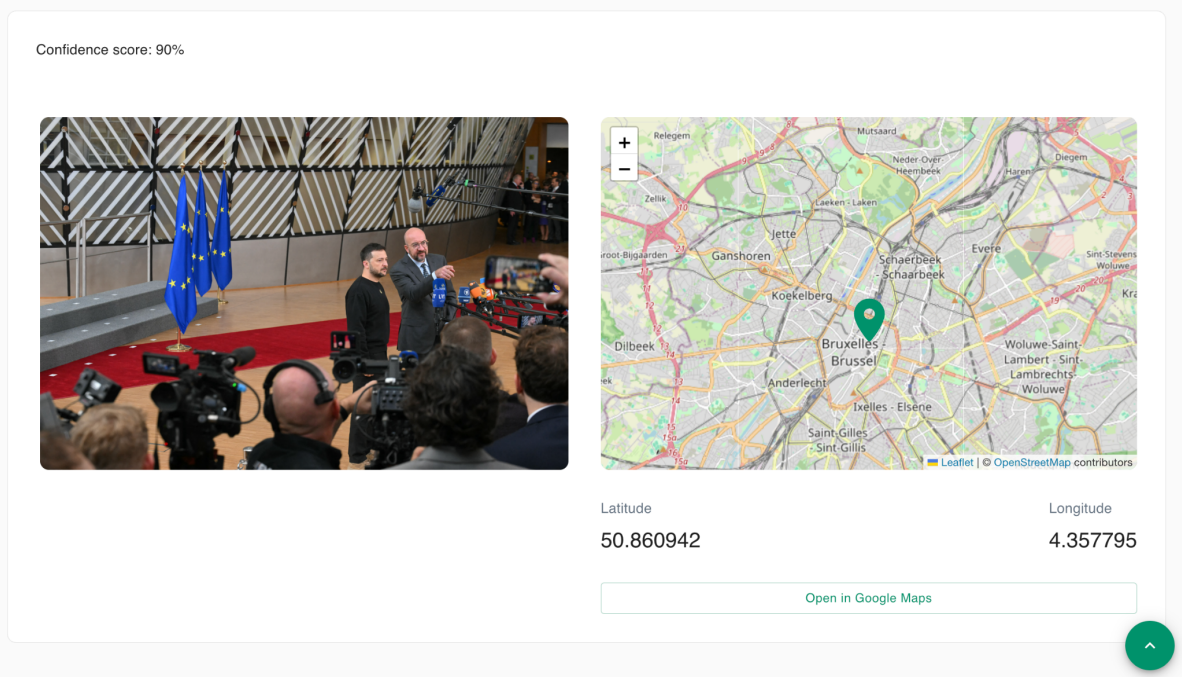

Figure 4 A screenshot of the new Geolocation tool results page

Machine Generated Text

The Machine Generated Text service provided by KInIT was implemented in the Assistant but only supports URLs. To test the performance of the service and debug potential issues, we have added a Machine Generated Text tool to the Verification Plugin to be able to provide raw text as the input. At the moment, this tool is only available for the KInIT team and the team at AFP and will become available to a broader set of users in a future plugin update. As illustrated in Figure 5, a query text is submitted for analysis, and the detection of machine generated text is presented with a graphical illustration.

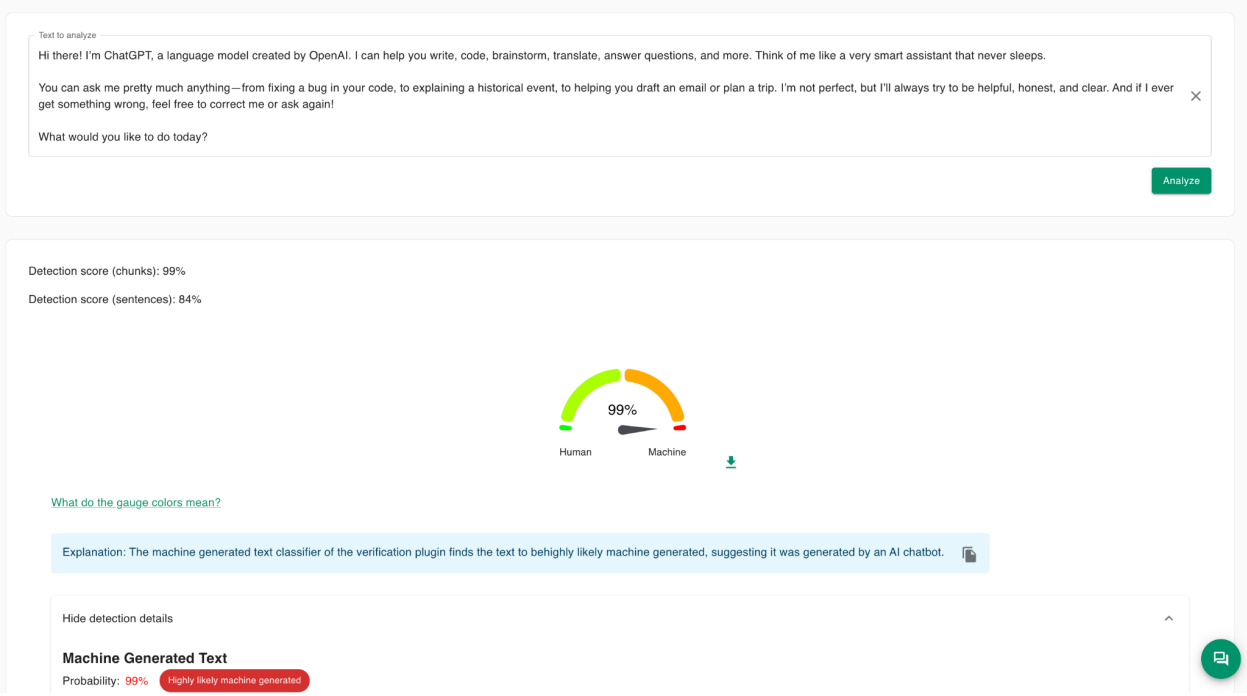

Figure 5 The standalone Machine Generated Text tool and results

Deepfake Video

We have integrated the new detection model of the Deepfake Video service by CERTH. After a period of evaluation among vera.ai consortium members with the Plugin version v0.86, the new model (faceswap_fsfm) is now accessible in the current plugin version (v0.87). We also included a chart visualising the model's detection rate over time in the video (available since v0.86) to better display the model's prediction over the duration of the video. The enhanced Deepfake Video UI is presented in Figure 6.

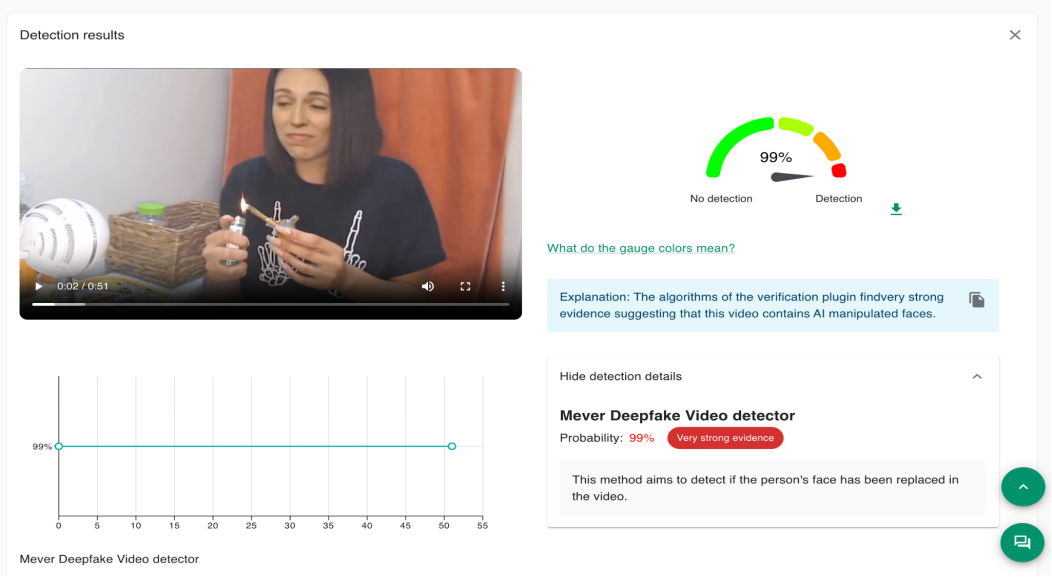

Figure 6 The enhanced Deepfake Video UI with the chart of detection percentage over time

2.4.2 Other Technical Developments

SNA (Social Network analysis)

Following the deprecation of the X/Twitter data access in July 2023, the social network analysis tool in the Verification Plugin became obsolete. To provide some new tools in the social media analysis space, we have rebuilt a SNA component aiming to resolve the data access problem, drawing inspiration from University of Amsterdam's Digital Methods Initiative tool Zeeschuimer[33]. The implementation is compatible with the ingestion and processing of data captured by Zeeschuimer on X and on TikTok as well as datasets previously captured on CrowdTangle, a tool deprecated by Meta in August 2024, for Facebook and Instagram metrics.

Beyond the analytic options available in our previous tools (such as distribution timeline, most frequent posters, most frequent hashtags) we have enhanced the capabilities by integrating in the SNA tool built into the Verification Plugin, the CooRTweet library developed by the team at the University of Urbino, to allow users to investigate links, hashtags, keywords and repost coordinated sharings.

We have also coupled the SNA tool with D3lta[34] – an open-source library made by Viginum (the French public FIMI agency) for detecting duplicate verbatim contents within a dataset, including the output of the CooRTweet component. This tool will be available in a future plugin update. An example of the SNA tool is shown in Figure 7.

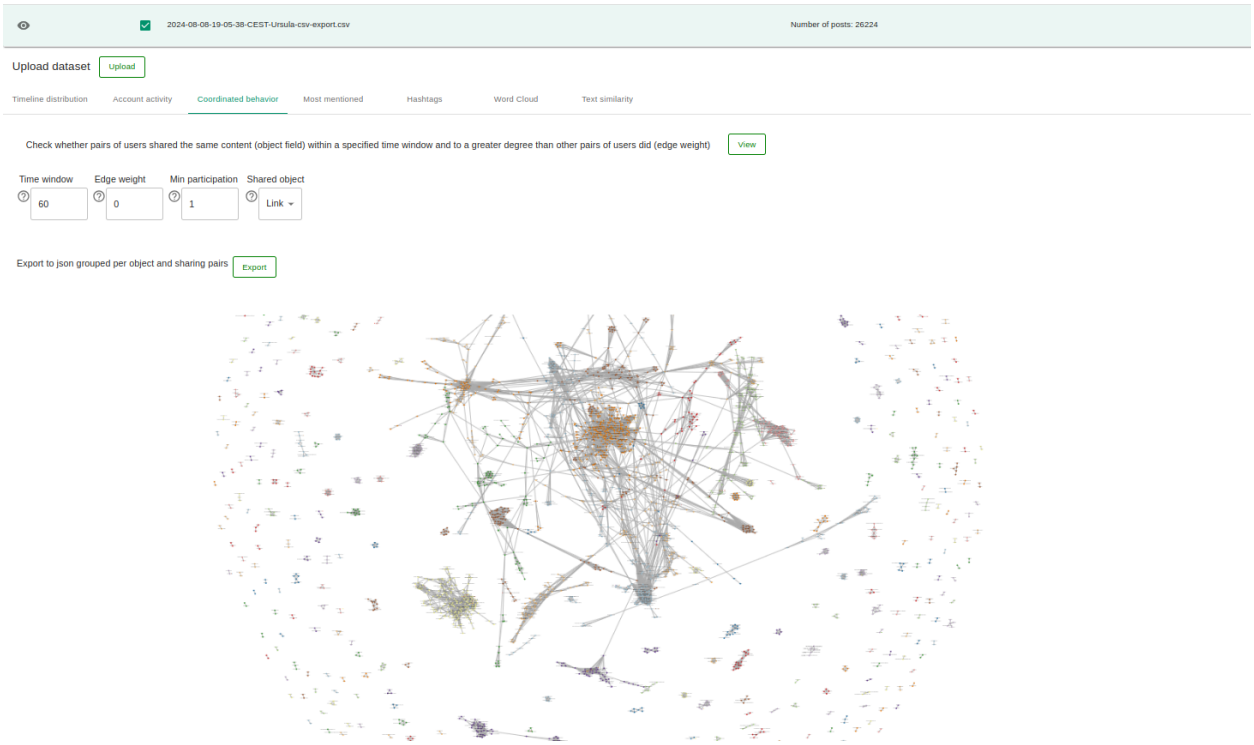

Figure 7 The Coordinated Inauthentic Behaviour visualisation in the new SNA tool

[33] https://github.com/digitalmethodsinitiative/zeeschuimer

[34] https://github.com/VIGINUM-FR/D3lta

Staging Back-End migration and Automation deployment

The plugin's middleware of our development server was moved to a more reliable AWS Cloud to secure better server reliability. We now use Amazon CodePipeline and S3 as a CI/CD pipeline to create new builds of the Verification Plugin. Additional services, such as the plugin translation service, have also been migrated.

All applications running on the back-end are built with Jenkins and deployed using ArgoCD on Amazon EKS. Karpenter automates the scaling of these applications, while Terraform facilitates the deployment of our infrastructure by having infrastructure as code. This combination enables us to minimize manual intervention in the deployment process by automating as much as feasible our back-end infrastructure.

Disinformation Monitoring Dashboard

Disinfo Deck, our dashboard for monitoring fact checks and RSS feeds, is now available in the Verification Plugin for our beta testing users (v0.86). This will improve the tool's discoverability for end users and enable us to gather feedback on desired features and overall user experience.

Metadata

We have updated the Metadata tool to use ExifTool for the videos. We use an implementation of ExifTool in WebAssembly. The library that was previously used did not support many file formats and was providing less EXIF data. For the images, we have also changed the library to use exifr, a library more performant than ExifTool in the browser, which also provides more metadata and supports more image file formats than the previous library. We also added C2PA metadata readings to the metadata tool and to Synthetic Image Detection since plugin version v0.87.

Front-End Caching

We have started introducing front-end caching to the Keyframes tool (v0.87 release of the Verification Plugin) with the intent of having a better user experience by decreasing the wait time if the user already sent the same request URL within a defined time. The front-end application then does not send the data to the server. This allows to have results faster as the user had already loaded them sometime beforehand. We are planning to extend the front-end cache to the other tools available in the plugin and are currently considering adding a local history of the past requests executed by the user so that they can go back to a previous result very fast. This will be an optional feature, and users could choose to disable it. Partners' current APIs are sensitive to load, and front-end cache allows to have fewer elements in the servers processing queues.

Refactoring

The plugin has had several years of active development, and legacy code, dependencies, and structures have been refactored. We introduced a significant refactoring to ensure that adding new tools is much more straightforward by having unique references. We have taken a similar approach to list other important variables and settings across the codebase, for the users' roles and permissions, and for the tools available at the top of the Verification Plugin's UI. This has allowed us to prevent errors and avoid hard-coded strings by referencing the constants defined only once in the code.

2.4.3 Improvements and Updates Following Evaluation Feedback

Synthetic Image Detection

We introduced a gauge chart for AI detectors with traffic light colours to align with other similar tools available in the industry. This helps end users to understand the results of the detection by keeping familiar colours. We also added key findings takeaway as a short text to help interpret the results. The model's detection results are structured from a concise overview to in-depth detail, enabling users to investigate further as needed. We set a 70% threshold, above which the models generally provide a reliable assessment of whether an image is synthetic. This is visually represented by the orange and red zones on the gauge chart.

C2PA

During the evaluation, we had multiple requests of fact checkers to get C2PA information for synthetic images (Adobe Firefly, ChatGPT) when they are available. It helps to improve the detection as C2PA provides provenance information for unmodified images. Enterprise AI models sign the image file with C2PA metadata and Generative AI information.

We have integrated a C2PA metadata reading and a visual GenAI alert in the Synthetic Image Detection tool. We have also added a C2PA component in the metadata tool to display all the C2PA provenance information. This is available in version v0.87. End-users will be able to read the C2PA Generative AI fields directly in the Synthetic Image Detection tool. An example of an image with C2PA metadata is shown in Figure 8 and Figure 9.

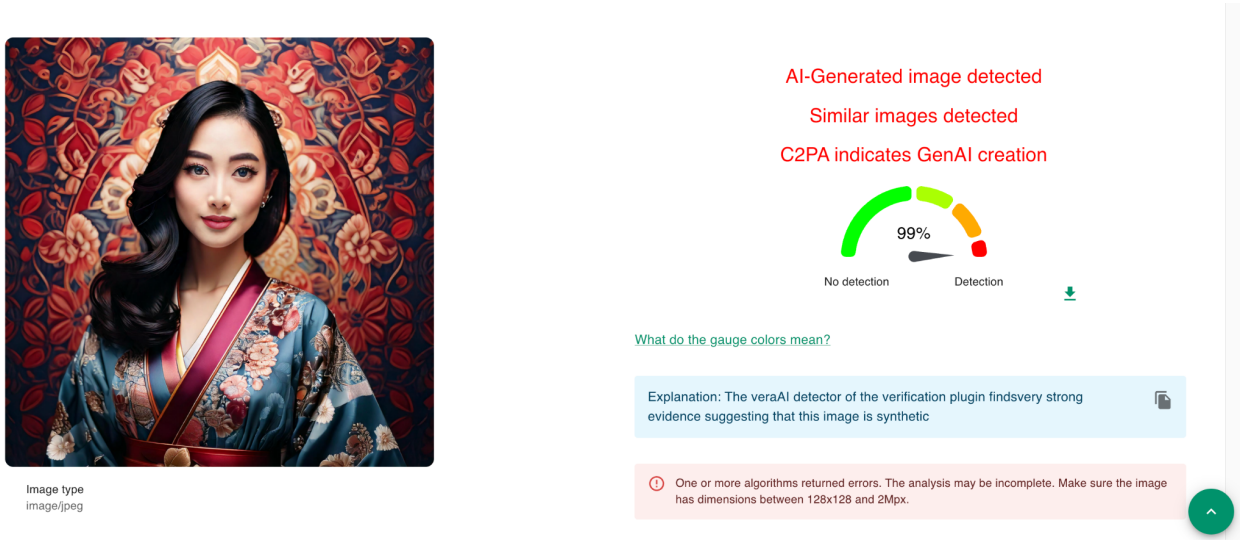

Figure 8 An Adobe Firefly Image with a C2PA warning

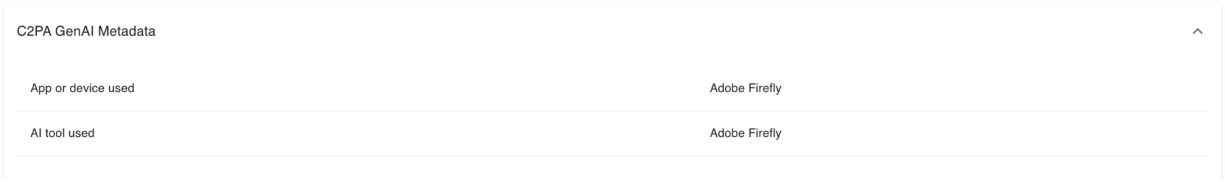

Figure 9 Presentation of the C2PA GenAI metadata in the Synthetic Image Detection tool

General User Workflow and User Interface

Thanks to the six evaluation cycles' feedback, we have significantly enhanced the usability and the layout responsiveness of the plugin. By analysing the errors of the services in the logs, we mitigated end users' mistakes by consolidating how to send URLs and files to the different services.

We have added missing common usage patterns: users can now save time by dropping files in a drag and drop zone, start a request by pressing enter on their keyboard, clear the URL typed or the results in one click. We have also introduced a setting to change the font size.

Lastly, we have improved the presentation of the UI elements when the window is small. The display of the UI elements is now more responsive, and the UI adapts to various display scenarios.

2.5 Challenges and Lessons Learned

The explainability of the models' results posed a significant challenge. Since end users do not always understand the complexities of machine learning, designing for clarity and usability required multiple iterations. We developed the Synthetic Image Detection tool in close collaboration with users, gathering feedback frequently. One key insight was that users often skim or overlook on-screen information when interpreting model outputs, which can lead to misjudgements. To address this, we structured the interface to clearly highlight the main conclusion, while still allowing users to explore more detailed explanations if needed.

False positive rates present another major challenge. When models are too unreliable, end users quickly lose trust in the detection tool and may choose not to use it again. For this reason, we have introduced an "Evaluation role", for the consortium members to test and evaluate all the algorithms without the need to release them publicly to avoid misuse.

Additionally, false positives can be mistakenly added to the near-duplicate detection service, potentially leading to flawed conclusions. This is especially problematic when previous false positives are shown to new users investigating similar content, compounding the impact of earlier errors. This issue remains only partially resolved. For now, we only deploy algorithms that, based on our evaluations, demonstrate low false positive rates. We are also exploring ways to collect user feedback when model predictions appear incorrect. This would allow for further refinement of the models on the backend or the removal of potentially problematic data from the near-duplicate detection service.

There are multiple image types and sizes that digital investigators have to examine. While running tests, we encountered multiple issues based on these image properties. For some images, the models will not provide a rightful assessment, or even not provide a result at all because they are not compatible with the image submitted. On the front-end side, this is a challenge, because performing image alterations such as image resizing could alter the detection rate. The older models developed by CERTH and the UNINA only supported JPG and PNG formats and were not trained to identify WebP images, which are commonly used across the web. As a result, any output from these models on WebP files was not meaningful. We flagged this limitation to our partners for further investigation. The newer models from UNINA and CERTH have also been trained on WebP images so that they can support multiple image types. With CERTH, we are

looking into ways to implement a more reliable way to make sure that all image sizes can be ingested by all the models, with a crop when needed. This is still a work in progress.

Finally, thanks to the knowledge acquired in the vera.ai project, the Verification Plugin coding and refactoring is now boosted by our new capacity to develop with LLMs. We have put in place a framework (using AWS and Bedrock) that allows us to find and correct bugs and inconsistencies more rapidly into the code and to quickly develop new prototypes and services.

2.6 Next Steps beyond vera.ai

After the project ends, AFP Medialab will continue to support the development and the maintenance of the Verification Plugin. We intend to continue to develop new features based on users' needs and requests. The plugin will also remain free of charge and open source (for the front-end application).

We have identified several enhancements based on user feedback. While this list remains subject to ongoing prioritisation and refinement, it currently includes the following:

1. Watermarking detection: this builds on previous work we have done at AFP to be able to identify watermarked AFP images. We would like to enhance image-based tools with watermarking detection to provide more insights on the provenance of the images when applicable.
2. Enhance the SNA further, with more analysis features and visualisations.
3. A narrative analysis tool to better understand the confusion created by multiple viewpoints on the same event and misleading alternative interpretations.
4. A video analysis tool to create image clustering (frames) and identify video highlights and how long they last.
5. Local history for all the tools, with optional shareable links. We have been looking into features that could enhance the workflows of end users and would like to add such features to enhance the speed of the digital investigators.
6. Re-add support for Mozilla Firefox: the extension used to be compatible with Firefox, but it is not the case anymore due to the complexity of maintaining two code repositories.
7. Convert Telegram Channel posts into RSS feeds to add them to the Disinformation Monitoring Dashboard.

For more details we refer the reader to D6.3 Final Exploitation and Sustainability Report and to AFP's exploitation plan under Annex I of D6.3.

3 Verification Assistant – Final Release

This section provides an overview of the Verification Assistant and its technical infrastructure. It details the technical work that took place leading to the final release of the Verification Assistant.

3.1 Tool Purpose and Overview

The Verification Assistant is a tool within the Verification Plugin and as such, the information already provided in relation to the plugin holds also for the Assistant. For further information, see section 2.1.

The Verification Assistant is designed to guide users to the right tools in the Verification Plugin depending on the contents they have submitted. Many tools and services are available within the Verification Plugin, and the Verification Assistant will select those most appropriate for the user. For images and videos, suitable tools will be suggested. For text, external services (such as APIs to NLP classifiers) will be called, and their results will be presented to the user.

3.2 End Users

The Verification Assistant is used by fact-checkers, journalists, researchers tackling disinformation, law enforcement agencies, and human rights defenders, among others.

3.3 Technical Infrastructure

Since the Verification Assistant is an integral component of the Verification Plugin, please refer to the technical information on the latter, in Section 2.3.

The frontend is the user-facing Chrome plugin extension previously described and the backend for the Verification Assistant is separated from the backend to the main Verification Plugin. The code repository for the Verification Assistant backend is stored on GitHub as a private repository, and the deployed production version is hosted at the University of Sheffield (USFD) in the GATE cloud cluster. This is a Kubernetes cluster which requires the backend to be containerised using Docker. The backend handles the image, video, text and URL extraction from websites (typical use is social media and news articles) as well as some proxied GATE cloud calls to NLP classifiers. The Python web framework utilised is Quart, an async version of Flask.

3.4 Integration Activities and other Technical Developments

Section 3.4 documents how the Verification Assistant was extended through targeted integrations and UI work so that more signals are surfaced directly in a single pass. The Assistant now includes concept enrichment, YouTube top-comments ingestion paired with a multilingual stance classifier (support/deny/query/comment) and accordion views by stance, a URL domain analysis service that flags risky or notable domains with drill-down details, and tight integration with the DBKF so similar debunks are surfaced during webpage analysis. For image analysis, the Assistant integrates the Synthetic Image Detection tool and adds a built-in chatbot to help users ask questions and leave feedback on results. Text

analysis is unified under six credibility signals. Together, these developments strengthen multimodal evidence retrieval and provide clearer, explainable outputs within the Assistant's workflow. More details about this work can be found in the following subsections.

3.4.1 Integration with Partner Services/Tools

USFD has integrated the concept enrichment service developed by vera.ai. The service extracts named entities from the text — specifically people, locations, and organizations — and displays this as a word cloud, with the size of the name being proportional to the number of mentions. This is illustrated in Figure 10 below.

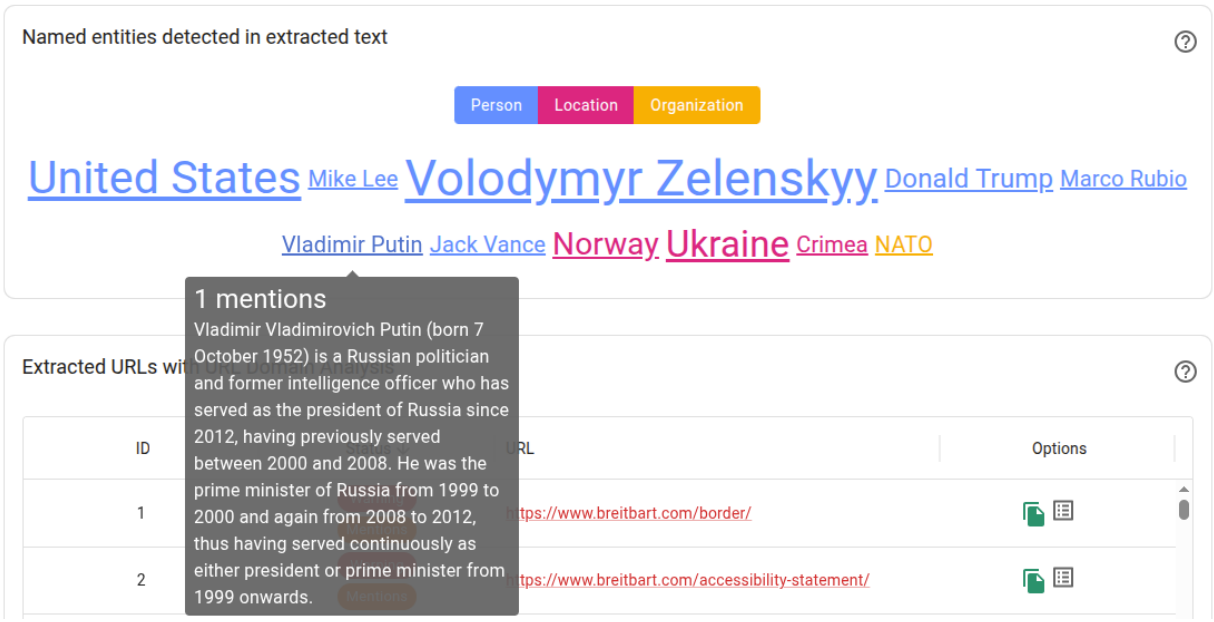

Figure 10 The named entity wordcloud with people, locations, and organisations

Users can submit URLs of YouTube videos to the Verification Assistant and the video is displayed in the "Extracted media files" box along with "Recommended tools" from the Verification Plugin. USFD has implemented fetching the video's top comments using YouTube Data API v3 and presenting these comments with their replies to the user. Additionally, the comments and their replies are sent via an API call to GATE's Multilingual Stance Classifier. Previously focussing on tweets, a comment is classified based on its stance or agreement to the original tweet. In the case of YouTube comments, this classifies the comments in regard to the video's title, and the replies in regard to its top-level comment. The four possible stances are: support, deny, query and comment. The UI displays an accordion with a list of all the top comments as well as accordions for the list of comments labelled for each stance class detected. The UI can be seen in Figure 11.

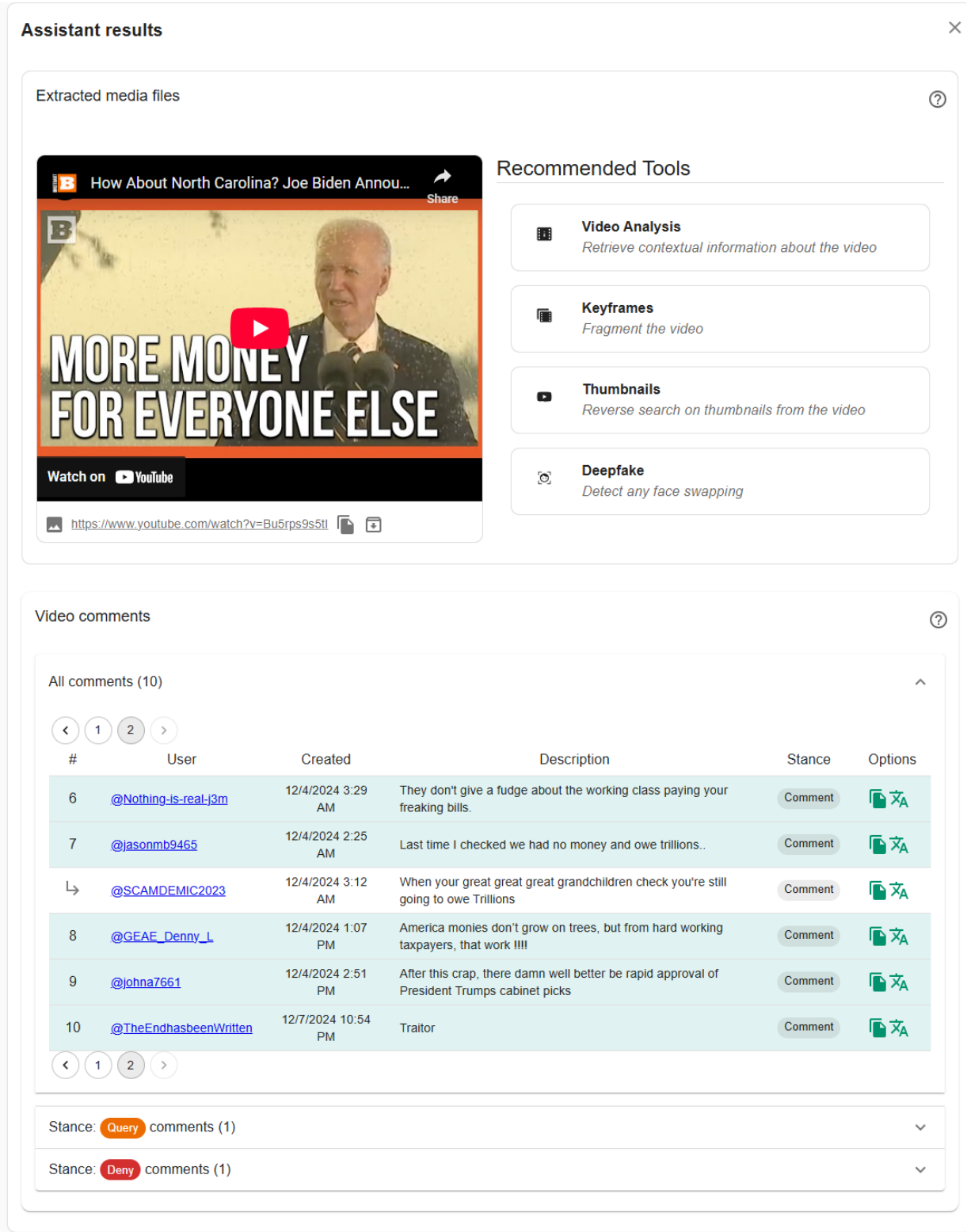

Figure 11 Extracted YouTube comments

The URL domain analysis service developed by USFD researchers has also been integrated. Extracted URLs from the analysed page are passed to the domain analysis service, which highlights warnings, mentions in

databases such as the DBKF, and known fact checking services. This is shown in Figure 12. For more details, users can click on the "Details" icon to display what has been highlighted, as shown in Figure 13.

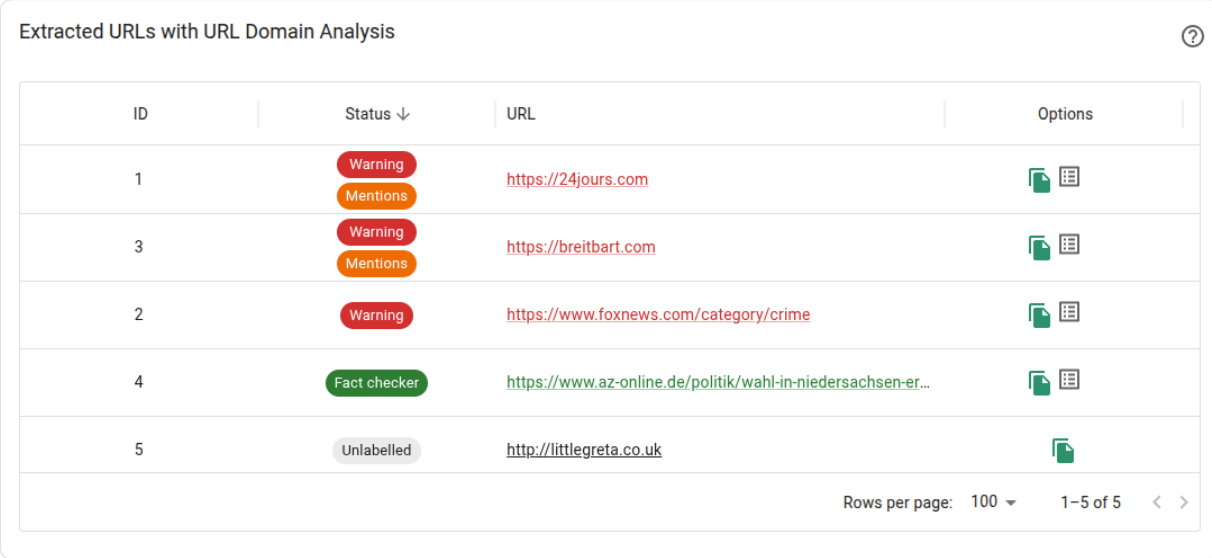

Figure 12 Domain analysis results

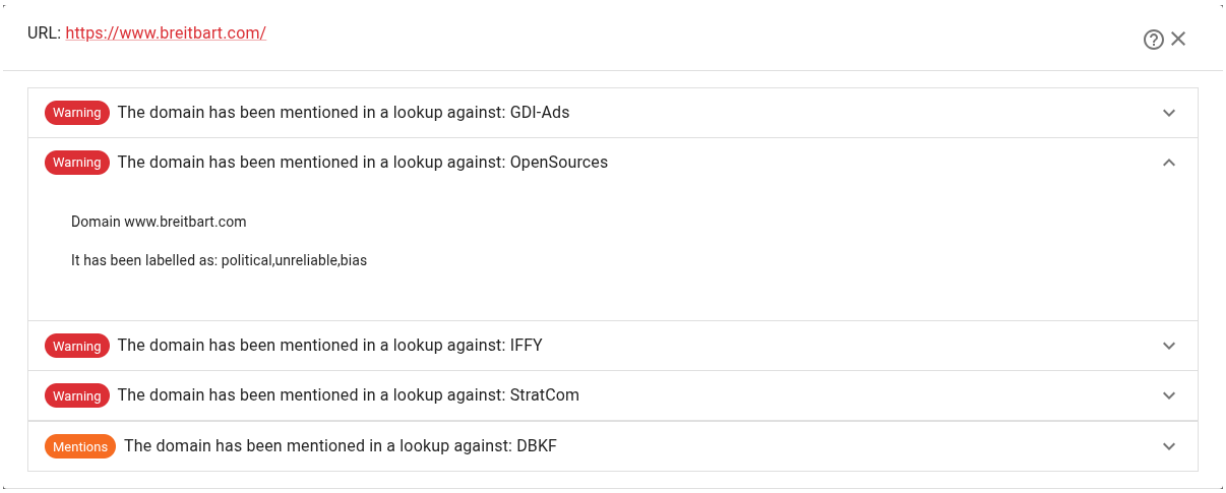

Figure 13 Domain analysis details

The vera.ai synthetic image detection tool is also integrated, which is one of several image-analysis tools available in the Verification Plugin. Users provide either a URL or upload a local file. The tool then employs several machine learning models, each of which provides a score representing how likely it is that the image was synthetically generated by AI. The best score is displayed to the user as a gauge, as illustrated in Figure 14.

As part of this, a chatbot assistant developed by USFD researchers has also been integrated in the assistant. This feature is currently not available in the public-facing version of the plugin, but upon successful completion of the user evaluation will be included in future releases of the assistant. The chatbot aims to help users provide feedback on the output of the tool or ask it clarification questions

about the tool's documentation. User inputs and chatbot responses are displayed as “message bubbles”, giving the interface a familiar look reminiscent of instant messaging.

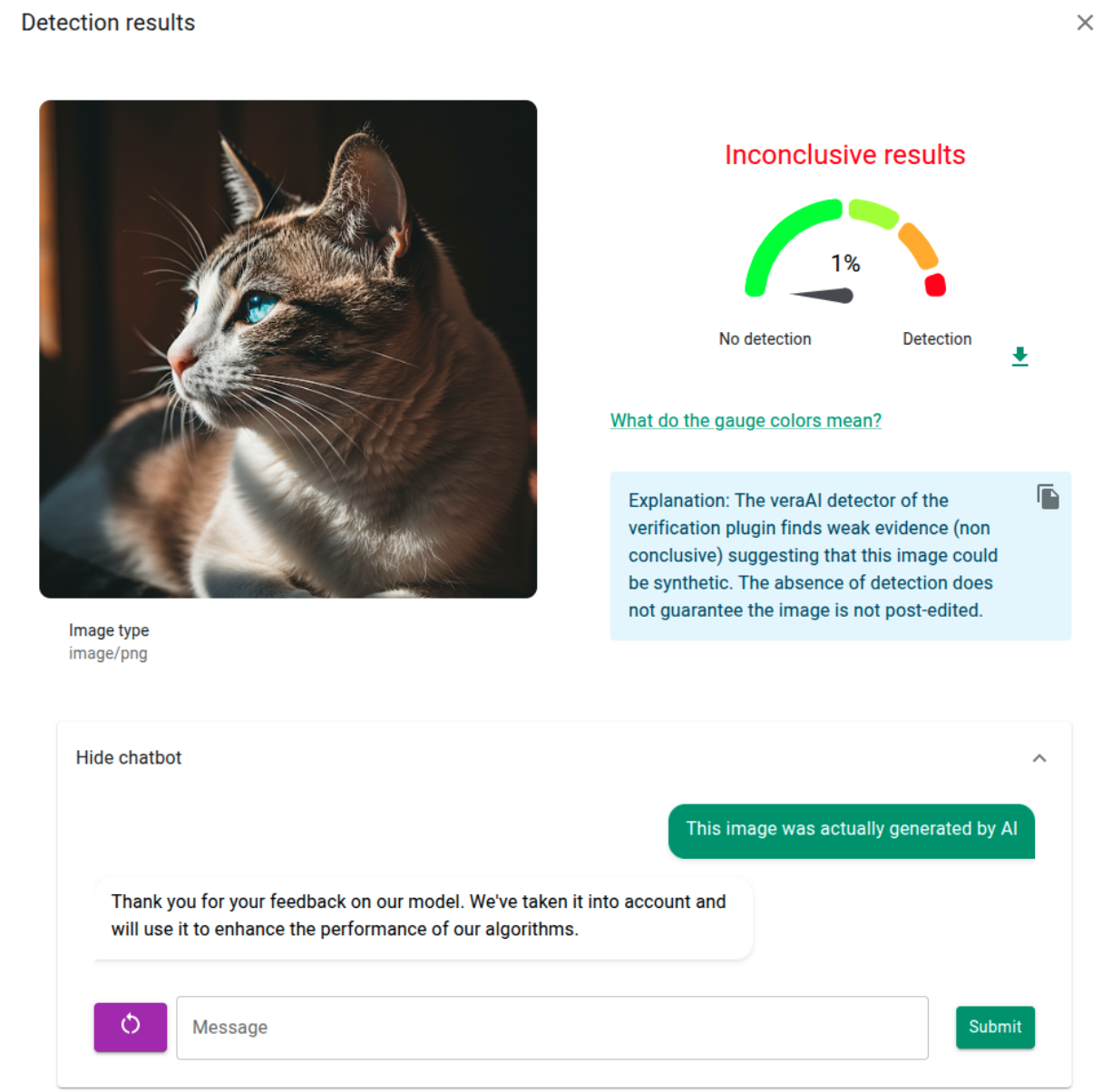

Figure 14 Synthetic image detection results with chatbot feedback

The assistant has also integrated the DBKF and Previous Fact Checks services, developed by KInIT. Text from the submitted URL is compared to previous fact-checks and texts in the DBKF for matches. If any matches are found, these are displayed to the user as illustrated in Figure 15.

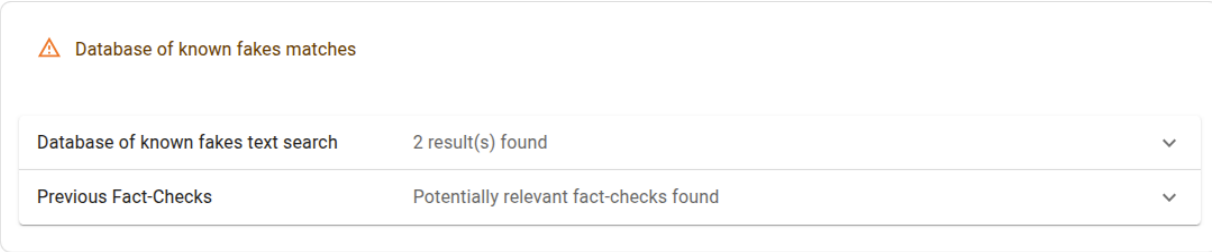

Figure 15 DBKF and previous fact checks results

The assistant has also integrated six credibility signals for text analysis. The services are: USFD's topic (news framing), USFD's news genre, and USFD's persuasion techniques; DW's subjectivity service; and KInIT's machine generated text and previous fact-checks services. The "Extracted text" box contains the text analysis results from these credibility signals. There are 6 tabs for 5 of the signals and the plain "Raw Text" extracted from the submitted URL. As previously mentioned, the previous fact-checks are contained in the "DBKF" section, letting the user know that the text, or similar text, has appeared in a fact-checker database already. Figure 16 depicts the format of the "Extracted text" box and shows the results for the news framing (topic) classifier. The left-hand side displays the text with sentences highlighted that are deemed important in the classifier's decision-making process. The right-hand side has a slider for changing the sentence importance score threshold showing the user which are more or less relevant to the classifier's decision. There is also a list of the topics detected. The same layout is shown for the news genre service. Persuasion techniques are detected at the sentence level. Therefore, the right-hand-side displays the list of techniques along with the total number of sentences annotated with that specific persuasion technique. Since multiple persuasion techniques can co-occur within the same sentence, the interface allows users to click on a specific technique in the right panel to highlight only the sentences corresponding to that technique of interest. For the subjectivity classifier, a slider varies the threshold for the subjective sentence scores and a gauge, akin to the one in Figure 17, provides the user with a full text score with a lever to a range which helps support the user's decision. The machine generated text service results are displayed in Figure 17. The gauge shows a full text score for how machine generated, or human written the text is. Underneath is a list of detected categories in traffic light colours and the sentences in the text on the left-hand side are highlighted corresponding to the class detected. Finally, the footer contains the detected language and three buttons: copy text, translate and a minimise/maximise option.

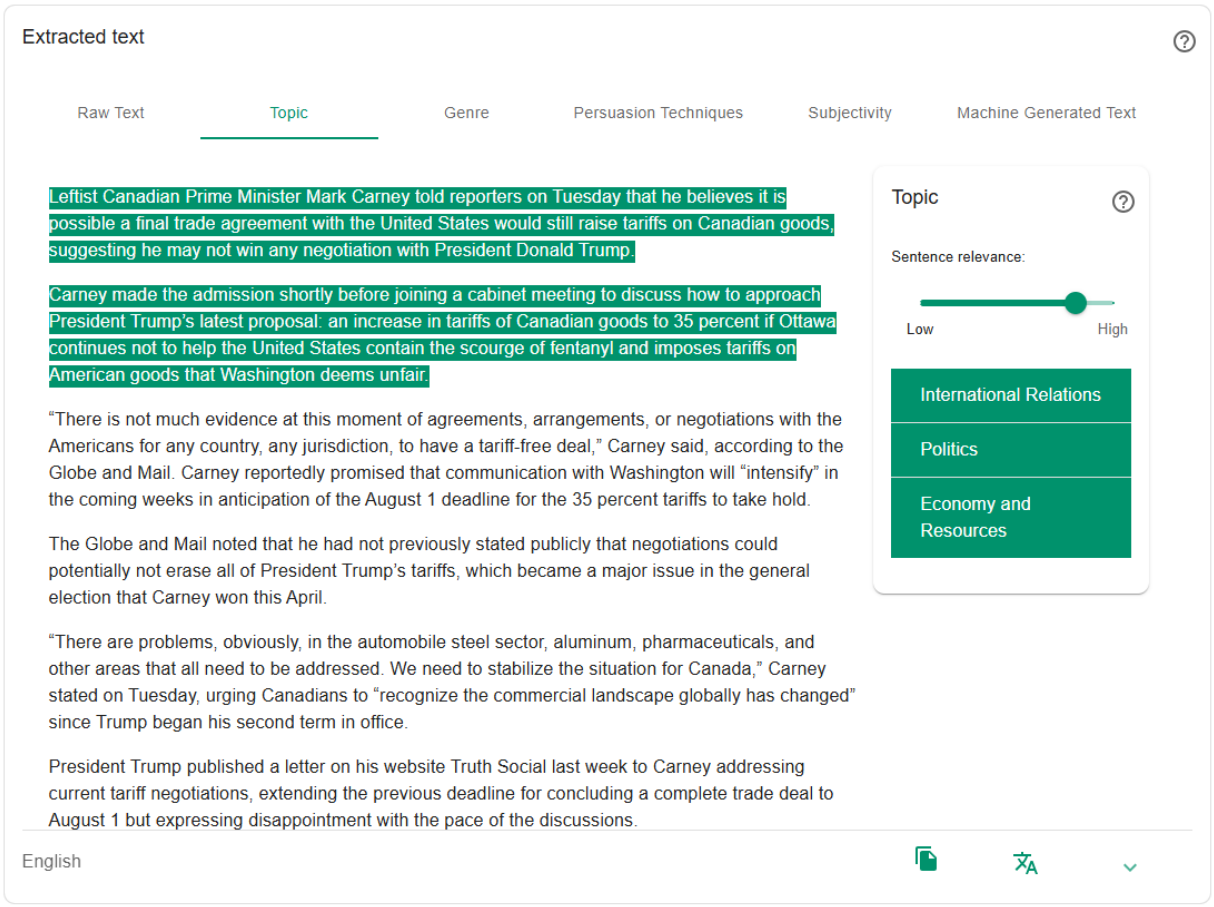

Figure 16 CREDSIGNALS-TOPIC: Credibility signals - news framing (topic) service

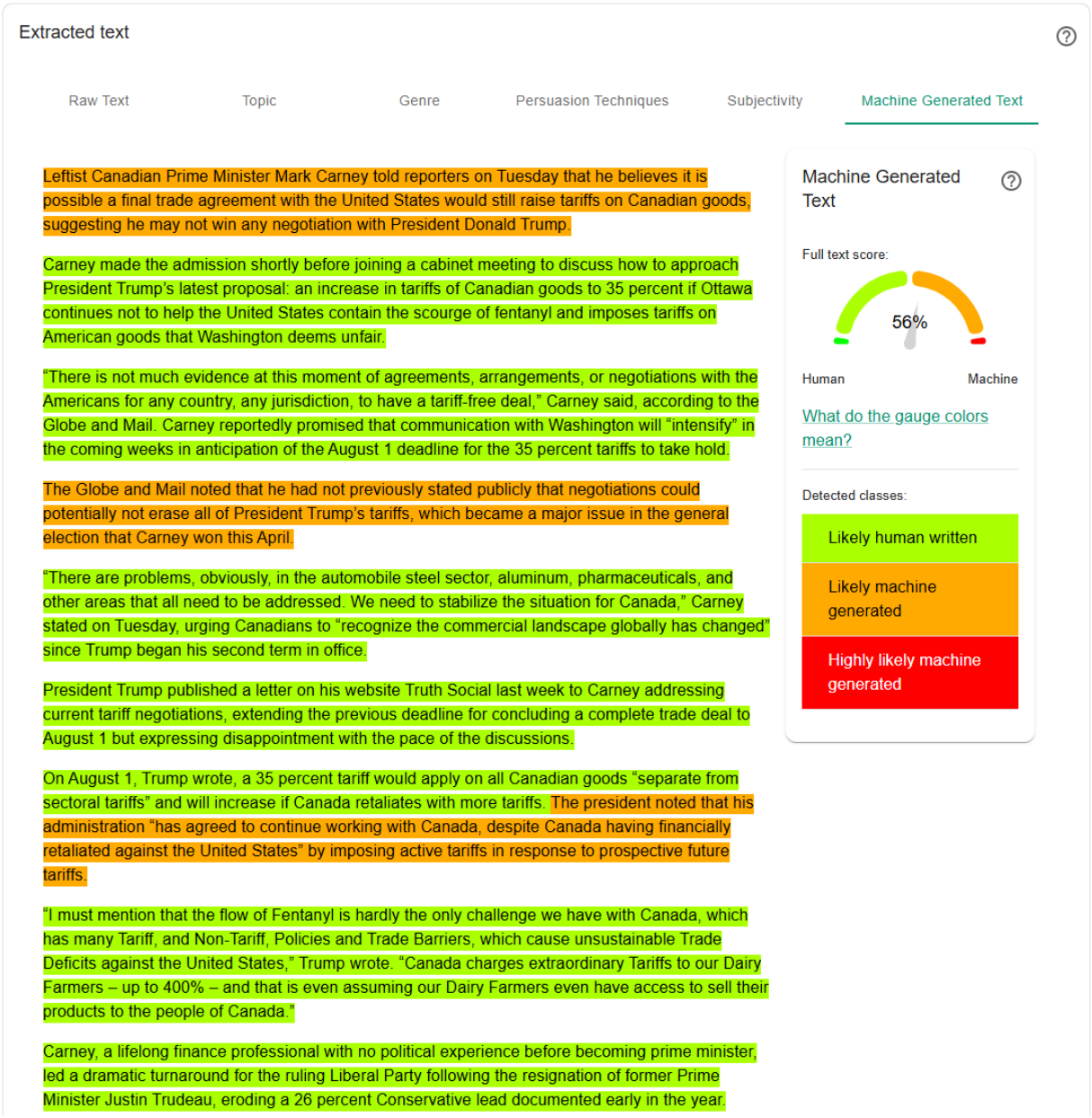

Figure 17 CREDSIGNALS-MGT: Credibility signals - machine generated text service

3.4.2 Other Technical Developments

Aside from the main features discussed above, USFD has also updated the plugin interface from MUIv6 to MUIv7. USFD has also continued to maintain the plugin backend to ensure that it remains able to scrape all the main social media sites as their site layouts evolve and change; and have also added explicit support for several new sites, including VK, Snapchat and BlueSky.

3.4.3 Improvements and Updates Following Evaluation Feedback

Comments from user evaluation regarding the look and feel of the UI have also been addressed. The colours, components, and spacing are now more unified, and the positioning of the loading icons is unified to make the UI more intuitive. In particular, the extracted images and videos are now separated into their own sections on the results page and present the results page as a single card which can be closed to reset the plugin.

The integration of the credibility signals has required a fair amount of reworking to get the user interface right for the users. The classifiers produce a lot of information, so decisions were required to decide how much of this information is necessary to provide for the users to make their opinions and decisions. Multiple feedback rounds have reduced the complexity of the colours to simplify and not overwhelm users, as well as added a threshold slider on some of the credibility signals as an alternative way to present the important sentences or classifier confidence. Initially, the credibility signals were contained in their own box with an accordion for each one, but given feedback this was reworked into the tabs inside the "Extracted text" box in the latest version.

Furthermore, extracted named entities link to their corresponding wikidata entry, with the exact mention count and a brief abstract being visible on hover. Users are able to filter which classes of entity (person, location, and organisation) are visible by clicking on the appropriate buttons. The word cloud also appears in a consistent order when the page is reloaded, with the entities appearing in order of their first mention.

3.5 Challenges and Lessons Learned

Social media websites regularly update their code which means that the Assistant has at times failed to extract the text or media because one of these websites was updated. Scraper tests have been added which act as an early warning system for when this happens. If the tests report a failure, or a user does, it becomes a priority to fix the functions to extract the correct information again when possible. There have been occasions when a social media website update prevents any extraction by the Assistant. For media, guidance has been added which explains to the user how to download the media and upload a local file instead. For text, the error is correctly caught and displayed to the user.

Upgrading libraries has been a necessary part of the Assistant. In July 2024, the Python version being used for the Assistant backend was upgraded from 3.8 to 3.11, as 3.8 was reaching its end of life in October 2024. Further work has gone into future proofing the code repository by using GitHub actions to run the tests on Python versions 3.12 and 3.13 to discover any issues that might arise in the future.

3.6 Next Steps beyond vera.ai

The verification assistant is integrated in the Verification plugin and as such will continue to be supported in its present form and provided to users free of charge for at least 3 years after the end of the project. As the project ends, the Assistant will be maintained and updated for the users.

There is further need for some improvements such as: code refactoring for readability, update to a more recent version of Python in the backend and maintenance of the social media content analysis. For these

reasons, and also to enable us to develop new features of the assistant, we are now applying for follow up UK and EU funding.

For more details we refer the reader to D6.3 Final Exploitation and Sustainability Report and to USFD's exploitation plan under Annex I of D6.3.

4 Truly Media – Final Release

This section provides an overview of Truly Media and its technical infrastructure. It details the technical work that took place leading to the final release of the AI-enhanced Truly Media version.

4.1 Tool Purpose and Overview

Truly Media[35] is a web-based journalism platform focused on the collaborative verification of content from social and digital media, co-developed by ATC and DW. It is used by DW and other organisations and companies, such as fact-checking organisations, broadcasters and news agencies[36]. The upgrade of this platform with AI-driven functions is expected to create additional business benefits for both DW and ATC. The main goal is to make a complex and difficult process easier for any kind of journalist by using external tools and bringing them together in one place, be they own developments or coming from third parties.

By using Truly Media, journalists can first collect and archive content around a topic they are investigating from different digital sources in a Collection. Collections are like thematic folders where relevant content is added and organised. Content inside a Collection can be further annotated and filtered. Figure 18 below illustrates a Truly Media Collection with related items.

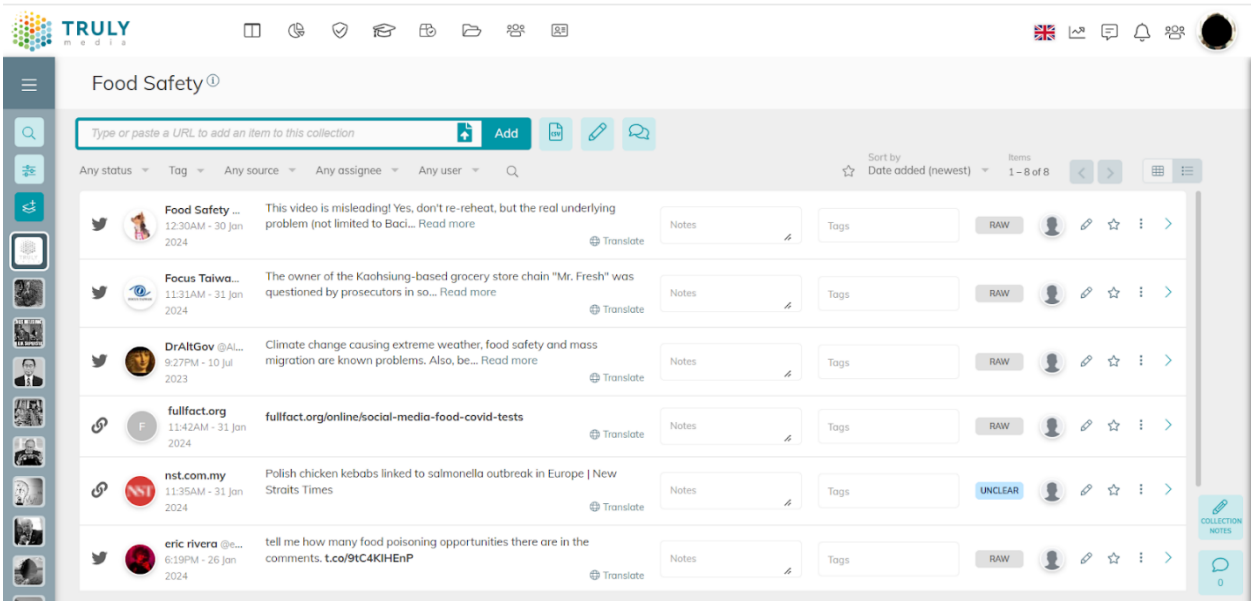

Figure 18 View of a Truly Media Collection

Single content items in a Collection, such as photos, videos, and social media posts, can be verified in detail through each item's verification page (see Figure 19 below). For this purpose, the platform provides

[35] https://www.truly.media/

[36] In case you would like to test the platform and the integrated vera.ai functionalities, please contact d.tsabouraki@atc.gr.

important third-party tools, which offer both high-end functions such as “reverse image search” as well as technologically basic but important services like “image magnification”. Following their investigation, journalists can then mark a content item as “pending”, “unclear”, “verified” or “fake”. Content Collections, the verification process, as well as single results can be easily shared and discussed with other users through a set of collaboration tools, like shared notes, chat, and direct messages.

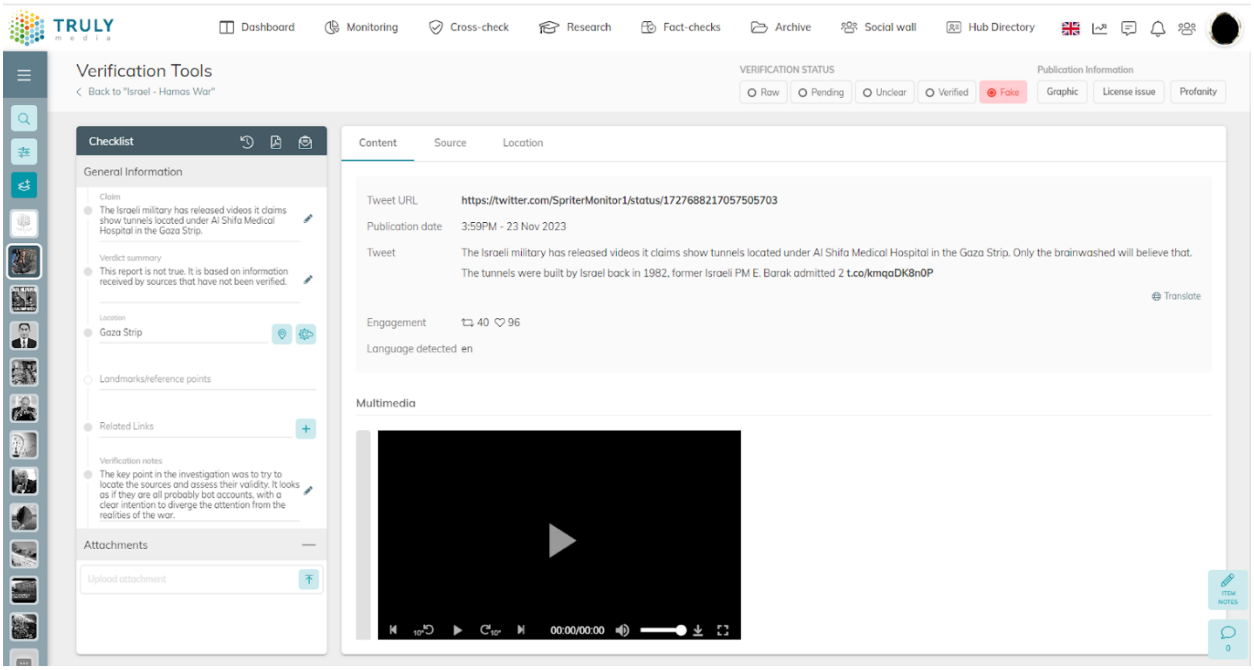

Figure 19 View of Truly Media’s verification page

Although individual journalists can use the platform, the focus of the approach is on remote collaboration across teams. In summary, the existing functions are:

- Monitoring social networks and web sources;
- Organising work/findings in Collections;
- Easily importing content;
- Collaborating in real-time;
- Extending verification networks;
- Managing content;
- Extracting and visualising useful information;
- Reporting and exporting results;
- Using effective verification tools and functions.

4.2 End Users

Truly Media is primarily targeted at journalists, as well as fact-checking and verification experts in media organisations. The “journalist” end-user target group can be broken down into four types of journalists: news journalist, verification specialist, social media, and/or investigative journalists. Moreover, Truly

Media is targeting professionals that work in medium size or big media organisations and not small media and fact-checking organisations or individual professionals.

4.3 Technical Infrastructure

Truly Media is a web application that consists of a SinglePageApp (frontend) built in Angular[37] and a loosely coupled backend built in Java Spring Boot[38]. The frontend is interacting with the backend by means of consumable API exposed by the backend. A few more elements implementing various features are also part of the overall architecture. Such elements are fulfilling the requirements for real-time push notifications (Node.js[39], socket.io[40]) and offline AI analysis. Figure 20 below provides an overview of Truly Media's high-level architecture.

The technical implementation of the Truly Media platform is based on the following technical stack:

- Angular 9+ (frontend);
- Java Spring Boot (backend);
- Node.js (socket server, download media module);
- Python[41] (AI + analysis libs).

The hosting environment is a combination of cloud and on-premise topology with the following elements:

- AWS S3[42], Cloudfront[43];
- Heroku[44];
- MongoDB Atlas [45](cloud);
- Solr[46] Server;
- Kubernetes[47] cluster (on premise).

The automated build and deployment process is executed by a CI/CD (Continuous Integration/ Continuous Deployment) pipeline implemented inside BitBucket and Jenkins[48]. The code resides in a versioning repository BitBucket, following the Gitflow principles.

[37] https://angular.dev/
[38] https://spring.io/projects/spring-boot
[39] https://nodejs.org/en
[40] https://socket.io/
[41] https://www.python.org/
[42] https://aws.amazon.com/s3/
[43] https://aws.amazon.com/cloudfront/
[44] https://www.heroku.com/
[45] https://www.mongodb.com/atlas
[46] https://solr.apache.org/
[47] https://kubernetes.io/
[48] https://www.jenkins.io/

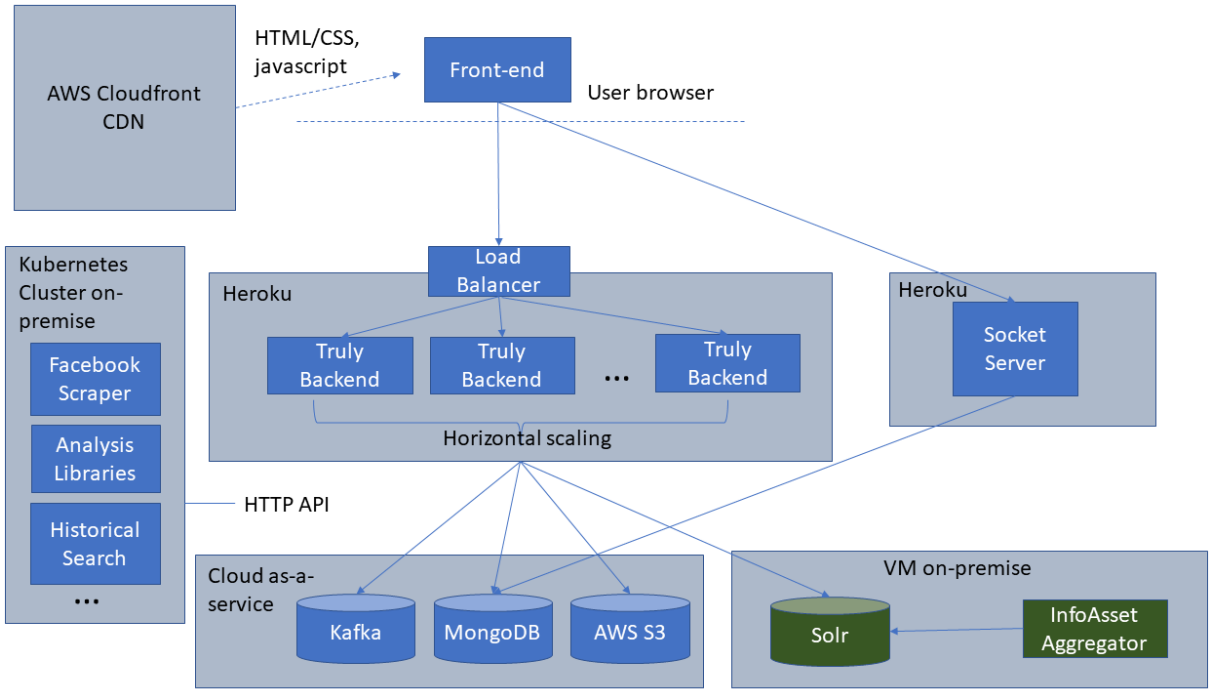

Figure 20 Truly Media architecture diagram

4.4 Integration Activities and other Technical Developments

This section provides details on the integration of AI components into Truly Media. The integration process involved designing and developing user interfaces, as well as establishing communication mechanisms between the demonstrator and the AI components/tools. This encompassed information exchange, interactions, and data structures. In total, three AI components have been integrated in Truly Media during the final reporting period, while several improvements and updates were implemented to already integrated components following feedback from the user evaluation sessions.

4.4.1 Integration with Partner Services/Tools

Integration of Machine Generated Text (MGT) Detection Service by KInIT

The MGT service detects whether an input text is machine generated or written by a human. The service has several applications in the verification domain and provides useful indications and clues as to the authenticity of an account or a piece of content.

The user can choose to analyse either the whole text of an article or post or selected excerpts. Once the analysis is complete, the service generates a probability score ranging from 0% to a 100% of whether the analysed text is machine generated or human written (see Figure 21 below).

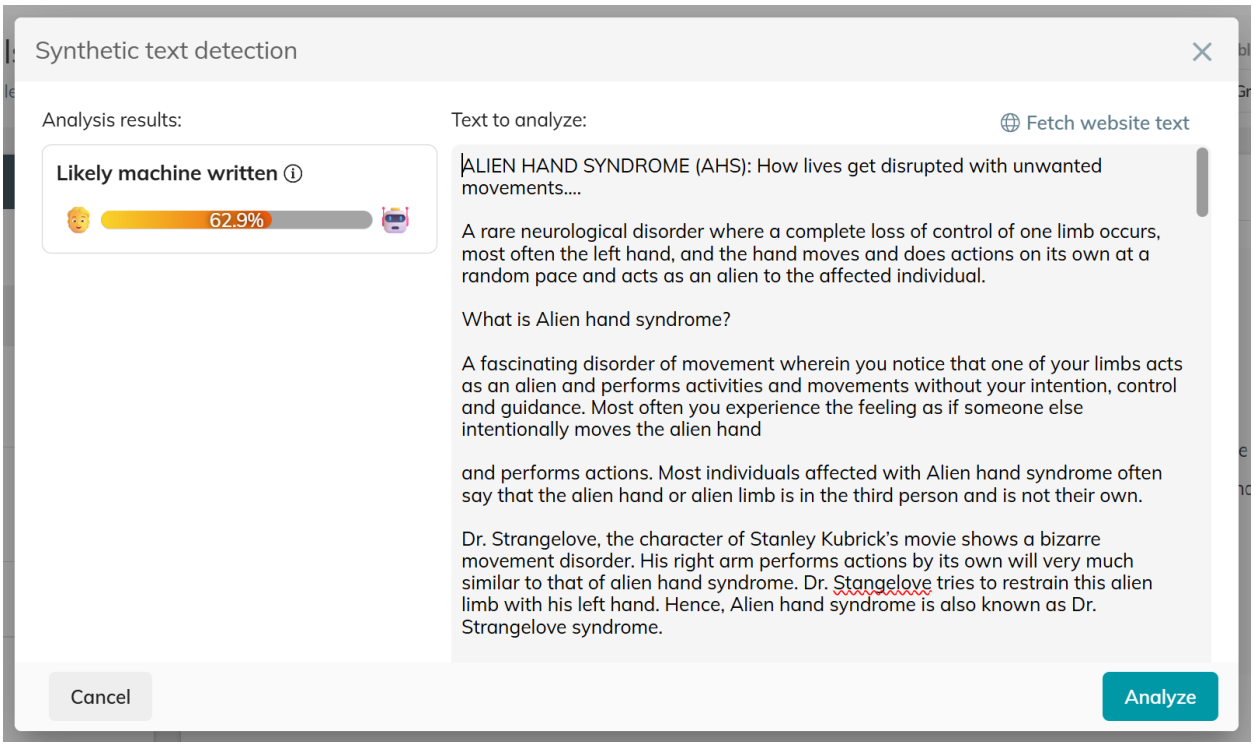

Figure 21 MGT service integrated in Truly Media

To support this functionality, ATC integrated the MGT API endpoint into its Java spring boot backend. The backend handles incoming requests, sends the relevant text to the MGT service for analysis, and returns the results to the frontend of Truly Media. The integration was designed to be stateless, with each request processed in real time to ensure up-to-date analysis.

In parallel, ATC developed a separate python based microservice responsible for extracting and preparing article content. This service scrapes webpages, removes HTML markup, and preserves useful formatting such as paragraphs, titles, and author information. The cleaned and structured text is then passed to the backend for MGT analysis.

Integration of Keyframe Selection and Enhancement Service (KSE) by CERTH

As Figure 22 illustrates, the KSE service allows the user to extract a set of representative keyframes from a video and enhances key elements, i.e. detected faces and text regions. The service is particularly useful when analysing videos to check whether scenes of the video have appeared online before or to investigate in more detail faces or shots in the video that contain text.

The user can further process and edit the extracted keyframes and video shots and perform reverse image searches to investigate whether the video has appeared online in the past or surface other details, such as information about the detected faces.

To integrate the KSE service, ATC connected it to its Java Spring boot backend, which acts as the intermediary between the user interface and the keyframe extraction pipeline. When the user requests the keyframe extraction service, an analysis is triggered with the video input, then Truly Media uses a

polling mechanism to check the processing status at regular intervals. Once the analysis is complete, the backend retrieves the results from the database and returns them to the user.

To optimise performance and reduce redundant processing, result caching was implemented. This allows the system to store and reuse the output of previously analysed videos, minimising repeated calls to the extraction service and speeding up response times for identical inputs.

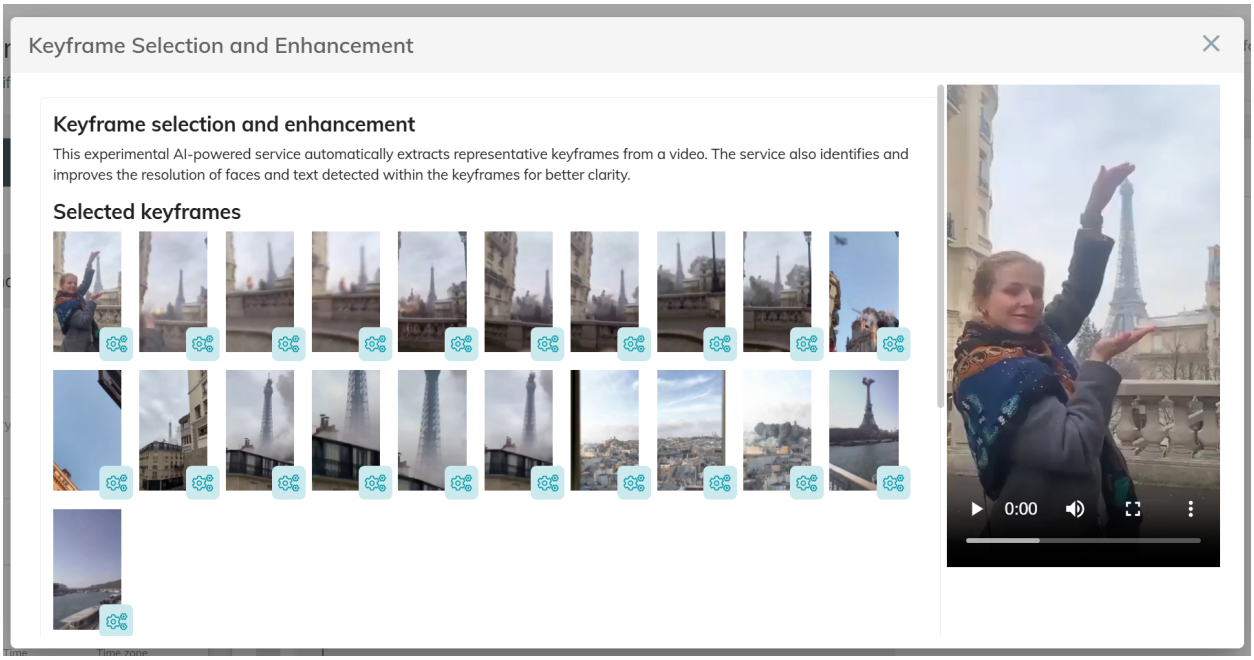

Figure 22 KSE service integrated in Truly Media

Integration of Synthetic Speech Detection Service by IDMT

The synthetic speech detection service supports users 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 be natural speech. The results are given as a waveform that indicates for each segment of the analysed audio the probabilities of the speech being synthetic or not (see Figure 23 below). Colour codes are also displayed below the waveform to help users interpret the results visually.

Moreover, to support users with interpreting the detection output, explanations of how the model works and how users should interpret the results are included in a dedicated support page (see Figure 24 below).

From a technical standpoint, the detection service is integrated into a node.js based audio processing microservice. The system initiates analysis tasks asynchronously and uses a polling mechanism to track their progress. Once completed, results are cached to improve performance for repeated queries. On the frontend, the waveform visualisation is powered by waveSurfer.js, enabling interactive playback with colour-coded overlays that correspond to the detection probabilities for each segment.

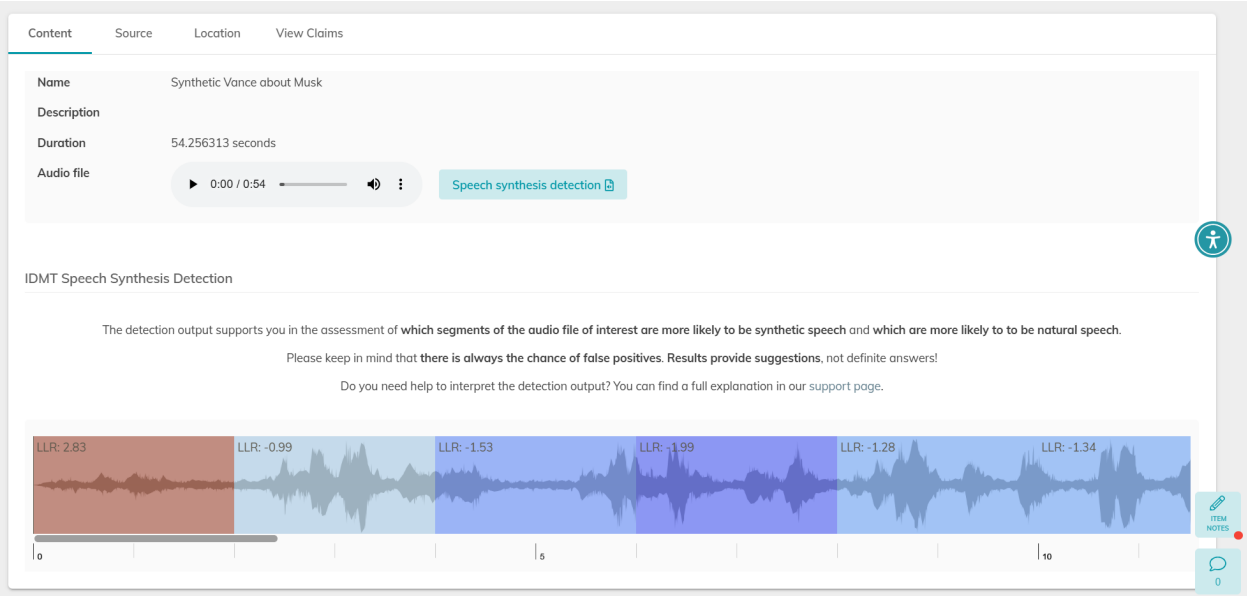

Figure 23 Synthetic Speech Detection service integrated in Truly Media

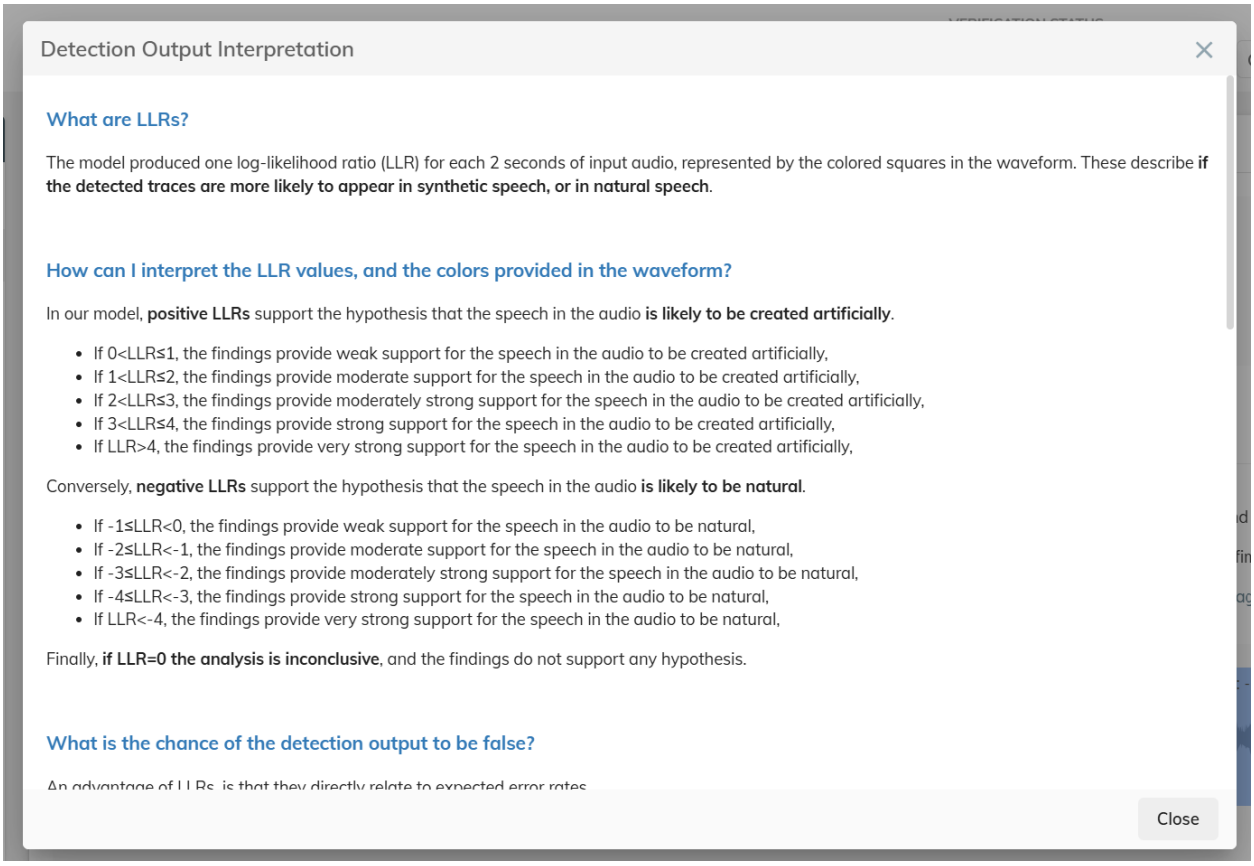

Figure 24 Support page within Truly Media for the Synthetic Speech Detection service

4.4.2 Other Technical Improvements

In addition to integrating and refining external analysis services, several improvements to the core functionality of the Truly Media platform were introduced. These enhancements focus on strengthening collaboration and ensuring better teamwork during verification workflows.

Enhanced Verification Support for Multimedia Items

To address the limitation that many integrated services could not process multimedia hosted on third-party platforms such as YouTube or Facebook, a dedicated service for video retrieval and audio extraction was developed. This service allows users to provide a link to external content, which is then securely downloaded and converted into an audio file within our infrastructure. By making the media file available locally, the service ensures compatibility with the integrated analysis tools in Truly Media and enables users to apply verification and detection workflows on multimedia they do not directly control. This addition significantly broadens the range of content that can be analysed through the platform, improving both flexibility and coverage.

Additionally, more search engines have been included in the reverse image search functionality of the platform, namely Bing, Google Lens, and Google Fact-check, to further facilitate image verification through the platform.

Workflow Improvements

The collaborative verification checklist has been enhanced with a locking mechanism designed to prevent conflicts when multiple users edit the same collection item. Previously, the checklist relied on web sockets for real-time updates, but this approach occasionally caused issues when two or more users attempted to edit simultaneously, leading to overwriting or loss of work. With the new mechanism, a user can explicitly lock a checklist before making edits, ensuring that only they can update it until they choose to unlock it. This enhancement improves the reliability of the platform and the integrity of the data during collaborative verification.

The assignment functionality of collection items has also been extended to support multiple assignees. Until now, an item could only be assigned to a single user, which limited collaboration on complex verification tasks. With this improvement, items can now be assigned to several users simultaneously, enabling teams to distribute responsibilities more flexibly and work together on the same item. This change supports more efficient teamwork and better reflects the collaborative workflows that Truly Media users follow in practice.

4.4.3 Improvements and Updates Following Evaluation Feedback

This subsection summarises the refinements made to Truly Media's integrated AI services in response to the project's participatory evaluations. Across sessions, users consistently emphasised three needs: (i) trust through transparency and explainability, (ii) ease of use and clear, non-ambiguous UI cues, and (iii)

contextualised, actionable outputs (e.g., timestamps, visual markers, concise guidance). These observations guided both algorithmic and interface updates detailed below.

Concretely, each service received targeted changes that align results, wording and visuals with users' expectations. For video deepfake detection, we streamlined the client-service communication, deprecated a legacy face-reenactment path, and aligned the client with CERTH's latest API to ensure compatibility with new models. Synthetic image detection broadened its coverage with additional detectors from CERTH/UNINA and now surfaces the model-provided labels directly in the UI to avoid local re-labelling and reduce misinterpretation.

Text analysis changes focused more on clarity and relevance: the Machine-Generated Text (MGT) service now presents an at-a-glance likelihood bar with a category label and explanatory tooltips, while a dedicated scraping microservice delivers cleaner article text and lets users analyse either selected excerpts or the full page - addressing requests for finer control and more meaningful inputs.

For video verification, the Keyframe Selection and Enhancement (KSE) view now shows dual timelines (keyframe markers and shot boundaries) and hover links from detected faces to their positions, with face/text panels collapsible to reduce clutter, making the analysis easier to follow at a glance.

Finally, for synthetic audio detection, interpretability aids were made more prominent (legend placement, "how it works" entry points, and inline guidance for per-segment scores) and the audio player is synchronised with the waveform, adding quick navigation and slower playback options to support careful inspection.

These changes aimed to keep cognitive load low, while keeping interpretation help and support handily accessible. The updates are elaborated in the following subsections.

Updates and improvements to the Video Deepfake Detection Service by CERTH

In this iteration, the ATC team working on Truly Media enhanced the integration of the video deepfake detection service to improve both stability and maintainability. Parts of the integration codebase were refactored to ensure more robust communication with the external service and better maintainability going forward. In addition, the previously integrated face reenactment detection algorithm was deprecated and removed, and our API client was updated to align with the latest API changes from CERTH, ensuring compatibility with their most recent models.

Updates and Improvements to the Synthetic Image Detection Service by CERTH and UNINA

For the synthetic image detection service, improvements that strengthen its functionality and deliver a more consistent user experience were implemented. On the algorithmic side, additional detection models were integrated, including *itw_rine_mever, itw_augrine_mever, itw_spai_mever, sd21_bfree-dino2reg4_grip, multi_bfree-dino2reg4_grip,* and *sd21_bfree-siglip_grip*[49], broadening the coverage of the analysis. From a usability perspective, the UI/UX was refined by adopting the labels returned directly from the service rather than calculating labels locally on the Truly Media frontend, as Figure 25 and Figure 26 show. This change ensures greater consistency with the underlying detection models, reduces the risk

[49] More information on these models is available in Deliverable D3.2.

of misinterpretation, and provides users with clearer and more reliable results during the verification workflow.

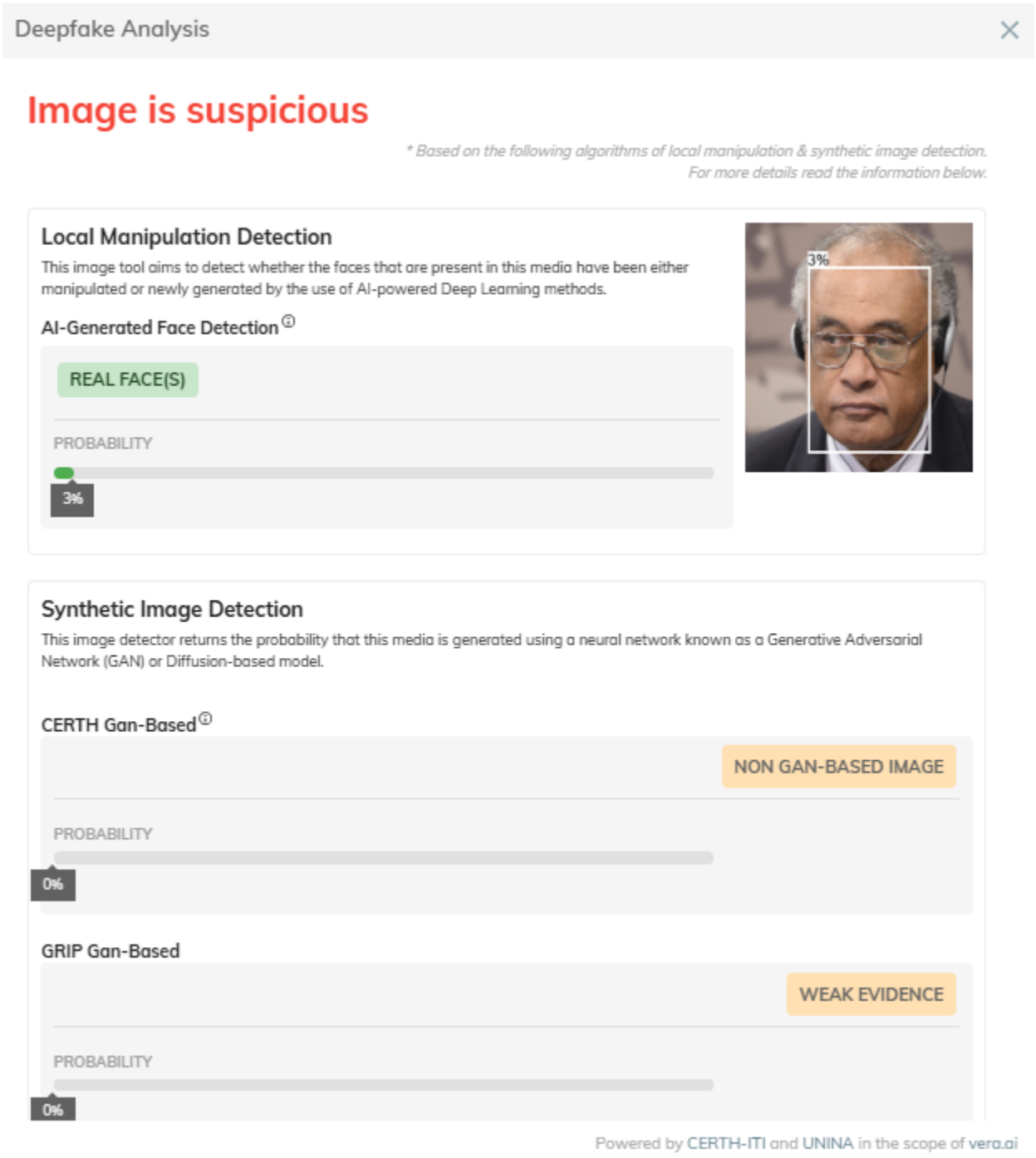

Figure 25 Result of a synthetic image detection analysis

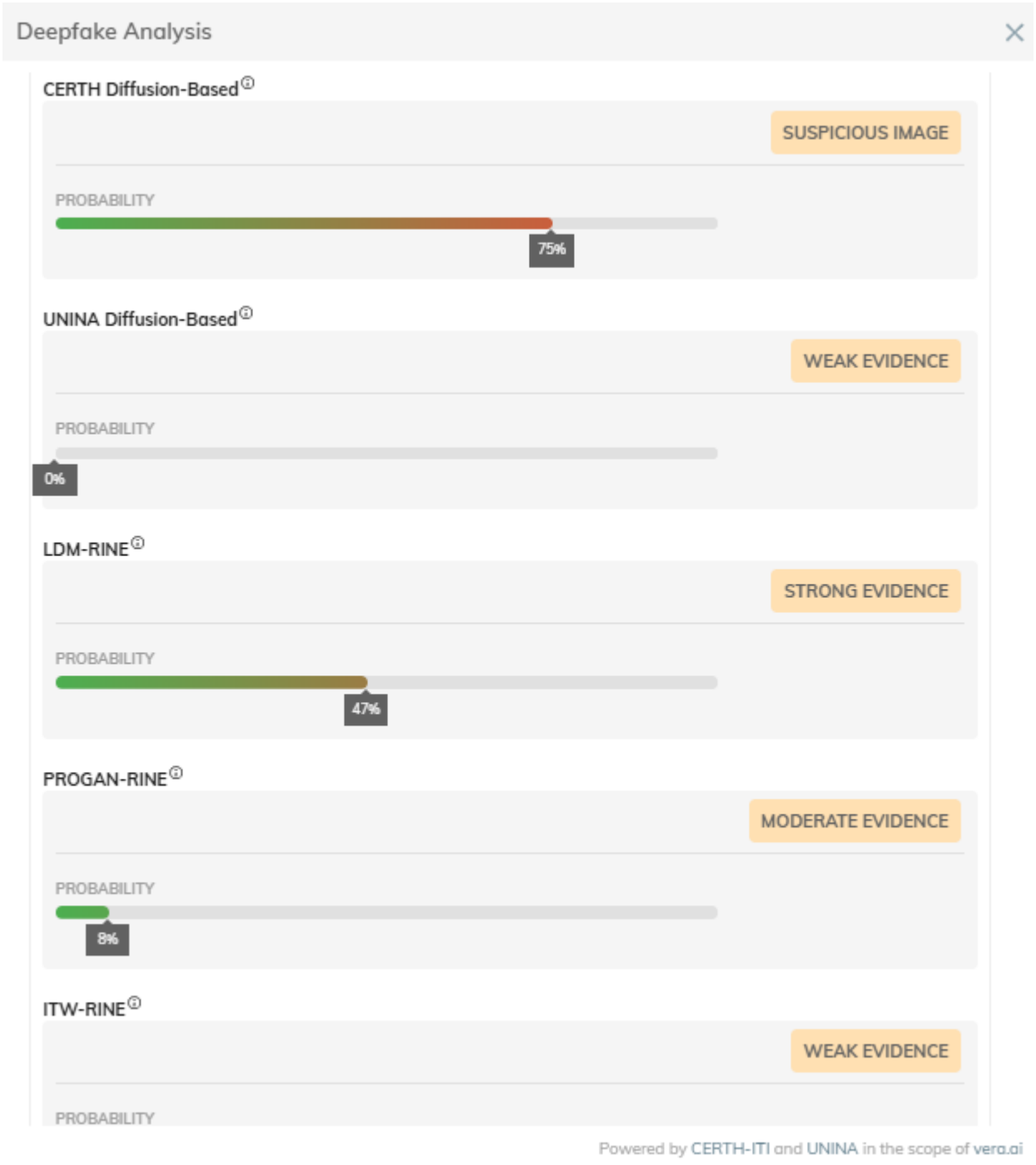

Figure 26 Result of a synthetic image detection analysis (continuation of Figure 25)

Updates and Improvements to the Machine Generated Text Service by KInIT

Based on the evaluation recommendations, the Truly Media UI was updated to improve user understanding of the Machine Generated Text Service's purpose and results. A bar indicating the likelihood of machine generation was introduced, a categorisation label added, and tooltips to explain the result's meaning were included.

One of the key findings from the evaluation was that the text scraped for analysis, especially from articles and websites, was often noisy and unstructured. To address this, a dedicated microservice for web scraping that removes irrelevant markup and content and outputs clean text with paragraph formatting was developed.

Additionally, a feature that allows users to analyse either a selected portion of the text or the entire article, as illustrated in the Figure 27, was implemented.

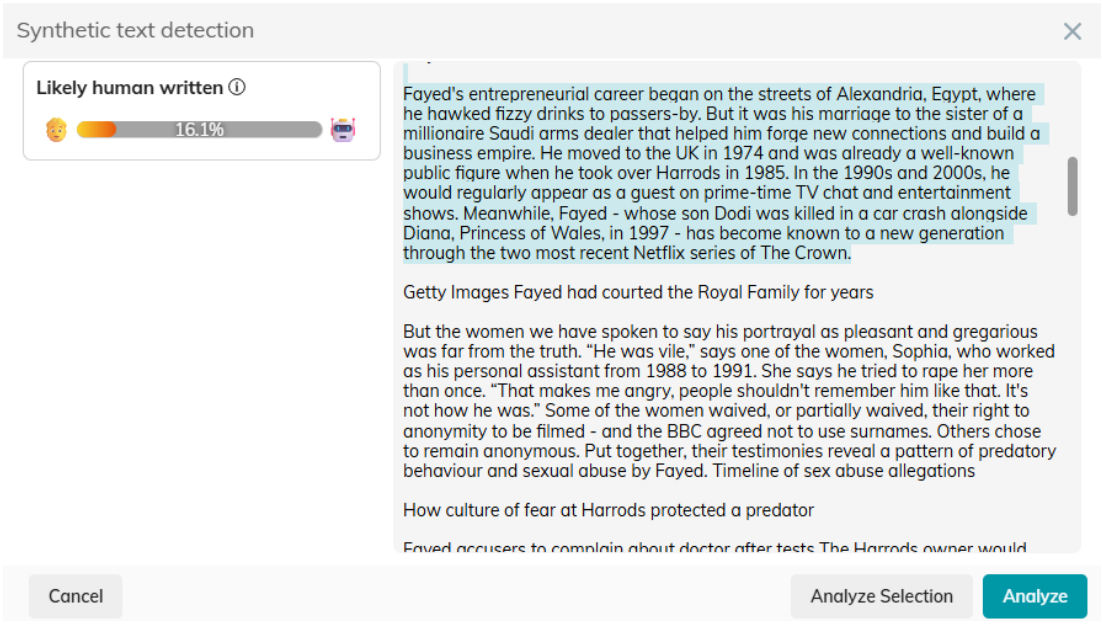

Figure 27 MGT improved analysis interface

Updates and Improvements to the Keyframe Selection and Enhancement Service by CERTH

Following user feedback during the evaluation of the Keyframe Selection and Enhancement Service, several usability improvements were implemented. To help users better understand the position of each keyframe within the context of the video, two timelines were added.

One timeline displays vertical lines representing individual keyframes placed along the video's duration, while the second shows shot changes using alternating background segments to distinguish different shots. Additionally, when a user hovers over a detected face image, a highlighted line appears on the keyframe timeline to indicate where that face occurs. This enhancement provides a clearer representation of the video's structure.

Additionally, to minimise visual clutter, especially in cases where a large number of faces or text regions are detected, the Detected faces and Detected text sections were redesigned as collapsible accordion panels. This allows users to expand and explore these details only when necessary, resulting in a more focused and user-friendly interface. Figure 28 shows the updated UI of the KSE service results in Truly Media.

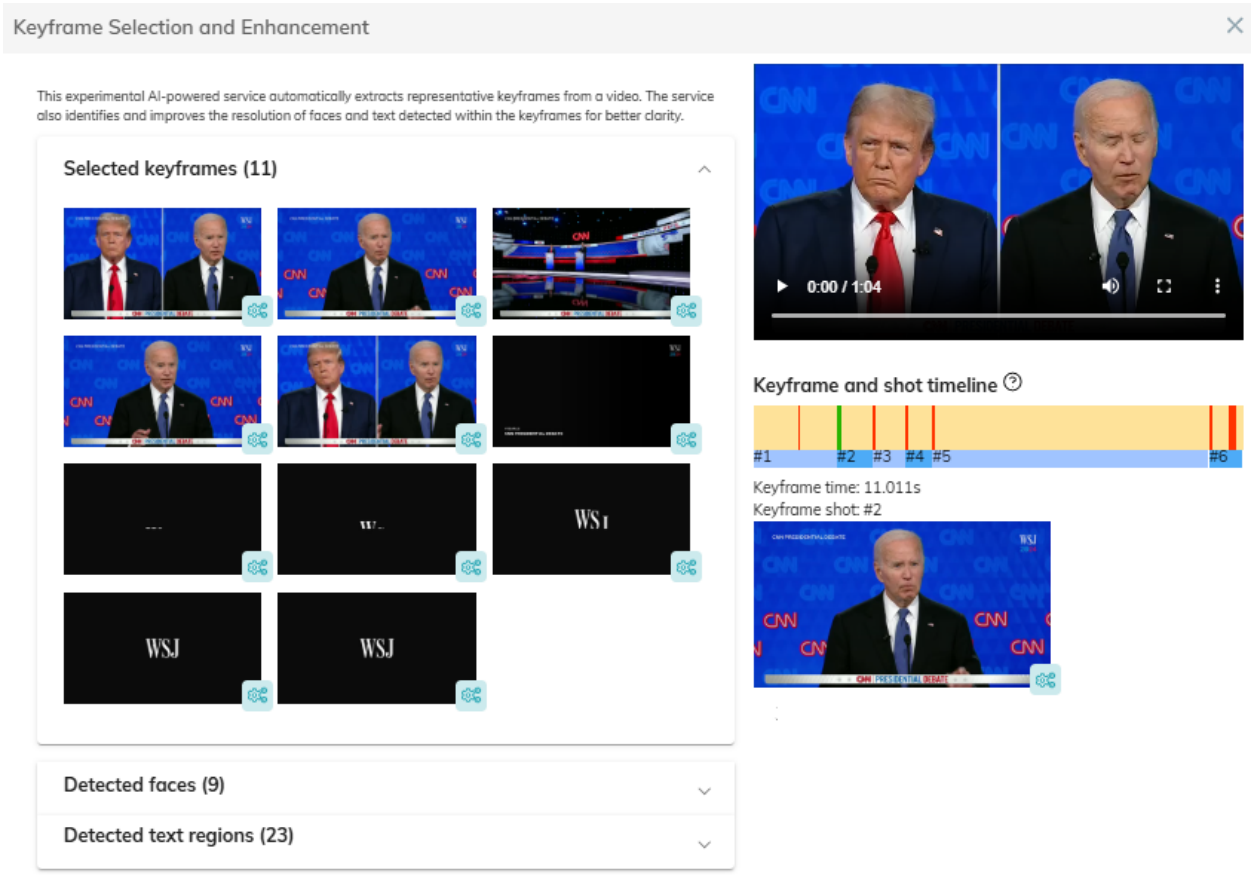

Figure 28 KSE collapsible accordion panels

Updates and Improvements to the Synthetic Audio Detection Service by IDMT

In response to evaluation feedback, the Synthetic Audio Detection service was refined to make interpretation of per-segment log-likelihood ratio (LLR) scores clearer and more accessible. A context tooltip was added directly on the waveform, explaining how to read the LLR values for each section and what they imply for authenticity assessment. In addition, the "How it works" entry point was made more prominent so users can access methodological guidance without leaving the results view. The colour-coding legend was moved above the waveform, ensuring users familiarise themselves with the meaning of colours before assessing the timeline, thereby reducing ambiguity and misreads.

Usability of the audio player was also improved to support careful inspection. The player is now synchronised with the waveform so the playhead and highlighted segment remain in lockstep, and quick navigation controls allow users to jump to the beginning or end of the track. For detailed listening, new playback speed options (0,5 and 0,25) make it easier to scrutinise suspicious passages. Together, these changes lower cognitive load, align the UI with users' perceptions and expectations, and make the service's outputs more actionable within existing verification workflows.

4.5 Challenges and Lessons Learned

Integrating research intensive AI technologies into a media production system that supports users in content verification tasks presents various technical challenges and associated risks. These challenges encompass a broad spectrum, ranging from technical complexities like interoperability and performance optimisation to regulatory compliance and user transparency and trust. By understanding and addressing these challenges proactively, stakeholders can navigate the integration process more effectively and mitigate potential risks to ensure the seamless functionality and reliability of the final solution. Below we briefly discuss various challenges and risks related to the technical work in T5.4 that have been faced during the integration process, but also - more generally - challenges and risks that can come up in similar contexts when dealing with actual media production systems.

One of the primary challenges lies in **delivering user-friendly results**, especially in making complex analysis outputs understandable to a diverse user base. While advanced libraries and visualization techniques can enhance interpretability, ongoing efforts are required to simplify results interpretation further. This involves not only improving the accessibility of analysis outputs but also ensuring that users can easily interact with the tools to derive actionable insights. Advanced user interfaces and intuitive design principles play a crucial role in facilitating user engagement and adoption.

Scalability concerns also arise, particularly regarding external modules integrated into media production systems like Truly Media. As the demand for analysis grows and users require processing of larger and more complex media assets, scalability becomes a critical factor. Without adequate measures in place, scalability challenges can lead to delays in response times, potentially undermining user experience and satisfaction. Proactive strategies, such as pre-processing assets offline and optimising resource allocation, are essential to ensure smooth scalability and uninterrupted service delivery.

Furthermore, the integration of externally hosted AI modules introduces other challenges, including the lack of visibility into internal workings and potential disruptions caused by changes or downtime. Effective communication with module owners and proactive monitoring of external libraries are needed to mitigate these integration challenges and maintain consistency in feature delivery. **Increased deployment costs** associated with containerised applications and on-premises Kubernetes clusters underscore the importance of cost management and resource optimisation in the integration process. Balancing the need for scalability and performance with cost considerations is essential to ensure the long-term sustainability of AI-powered media production systems. Integrating diverse AI components from different partners may also present **interoperability challenges**. Ensuring seamless communication and compatibility between various components, each with its own technical specifications and requirements, is complex. Addressing interoperability issues during integration is crucial to avoid disruptions in the functionality of the overall system.

Moreover, integrating advanced AI components into media production systems not only demands **specialised technical expertise and user training** but also necessitates overcoming challenges in change management and stakeholder alignment. These challenges include acquiring or developing new skills within the organization, as well as managing resistance to change, differing priorities, and communication gaps among teams and stakeholders. Investing in training programs, change management strategies, stakeholder engagement, and clear communication is vital for building the capabilities required for

successfully deploying and operating AI-powered systems while ensuring smooth integration within the organization.

4.6 Next Steps beyond vera.ai

Looking beyond the conclusion of the project, it is essential to mention that Truly Media is already a commercial product with a user base that includes well known news outlets like Reuters' fact-checking and User-Generated Content (UGC) teams, as well as members of the European Digital Media Observatory (EDMO) community. As such, the basis of the demonstrator is continuously maintained and enhanced with new features in response to user needs.

ATC and DW are both very well placed within the disinformation detection and respective research domain, participating in several initiatives and networks like EDMO, relevant industry and research clusters, media networks, as well as European and national R&D projects dealing with the development of technologies and tools to tackle disinformation. Knowledge, know-how, and experience generated through vera.ai, coming from different project activities like user requirements identification, integration activities, insights from the evaluation etc., will be leveraged by both DW and ATC and will be used as valuable input for various purposes, including product development, consulting services, and new research and development projects and activities.

Moreover, the synergies and relationships with other research and industry partners that have been fostered within the vera.ai network will be sustained and further strengthened through continued collaboration in subsequent projects. For example, ATC, DW, and CERTH are all partners in the Horizon Europe AI-CODE project dealing with the development of advanced AI solutions against advanced disinformation techniques for media professionals. Outcomes from vera.ai are used as background knowledge, and respective technology will be further improved and enhanced in projects like the one mentioned above.

In addition, ATC and DW will explore the feasibility of transferring mature enough components from vera.ai partners into Truly Media's production environment, extending accessibility to real-world users. This will depend on the maturity of the integrated technologies and the ability to establish suitable service or licensing agreements among stakeholders. Critical aspects to be addressed in such agreements encompass licensing terms, fees, service level requirements, deployment, and hosting arrangements.

Finally, Truly Media is currently undergoing a major re-design to meet more modern technology stack standards and UI/UX design principles. Know-how and knowledge from vera.ai around user needs and expectations, as well as valuable feedback gathered through the user evaluation sessions, are already feeding into the design and development of the new platform. Moreover, the integrated vera.ai components will be migrated to the new version of the platform with upgraded interfaces and a more tight integration with user workflows.

For more details we refer the reader to D6.3 Final Exploitation and Sustainability Report and to ATC's exploitation plan under Annex I of D6.3.

5 Database of Known Fakes – Final Release

The following sections describe the final release of the Database of Known Fakes (DBKF) within the scope of vera.ai. The additions and improvements to the database in this past year and a half since deliverable D5.2 were carried out in two main directions: extending search capabilities through integration of WP4-led research outputs and aligning existing functionalities with user feedback.

5.1 Tool Purpose and Overview

Building on the earlier description in D5.2, it should be noted that along with IFCN members, the DBKF now includes sources from the EDMO fact-checking community[50] too. In addition, the database already enables search through other types of documents such as social media posts, news articles and on the more granular level of claims and claim reviews (in contrast to previously having these combined, in wholeness, as debunks). The integration of a conversational interface, leveraging the capabilities of large language models (LLMs), turns the DBKF into an even more powerful, publicly available[51] tool for fact-checking content exploration and analysis.

5.2 End Users

The primary stakeholders who can benefit most from DBKF use remain fact-checkers and media professionals. With the addition of the cluster search feature, the database becomes even more relevant for researchers, especially the ones interested in studying disinformation narrative patterns.

5.3 Technical Infrastructure

Short descriptions of the main DBKF components are available in deliverable D5.2. Figure 29 below presents an updated version of the system architecture with new services included (marked with red rectangles). These additions are as follows:

- The DBKF Chatbot is a Python service that allows users to explore the database via asking questions to an AI agent.
- The Manual Ingester Service supports the manual upload of zipped social media exports (such as Telegram chats, Facebook/X/TikTok posts among others), which are then processed and converted into a proper data format to be stored in GraphDB[52], the underlying graph storage database.

The Cluster Search and the former Multilingual Search, now Translated Search, are endpoints in the Search Service API, hence are not present as standalone components. The system also makes calls to two

[50] https://edmo.eu/areas-of-activities/fact-checking/fact-checking-community-how-to-join/#1704890419295-17674ceb-9076

[51] https://dbkf.ontotext.com/

[52] https://www.ontotext.com/products/graphdb/

additional external APIs, namely the Google FactCheck Explorer API[53] for complementary data collection and OpenAI Assistants API[54] for the conversational functionality.

Due to the addition of new services, the requirements for processing power increased significantly, straining the existing infrastructure. Therefore, the three Hetzner-hosted environments (DEV, STAGE, PROD) have been upgraded to a 24-core Intel® Xeon® Gold 5412U CPU with 256 GB RAM (double the size of the previous machines) and 3.84 TB NVMe SSD to meet the higher demand.

[53] https://developers.google.com/fact-check/tools/api/
[54] https://platform.openai.com/docs/assistants/overview

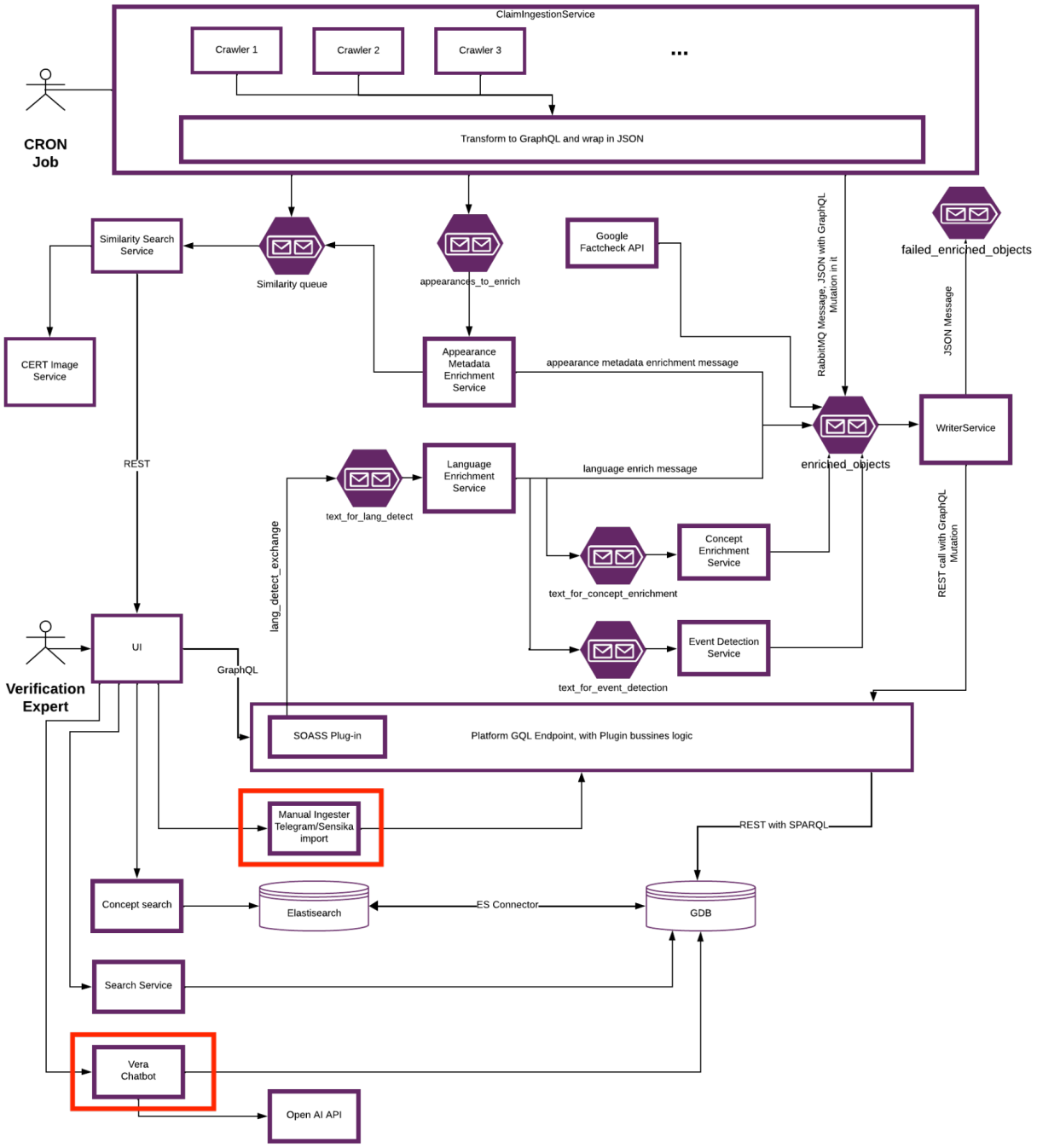

Figure 29 Updated DBKF architecture diagram

5.4 Integration Activities and other Technical Developments

Technical work in the reported period largely focused on integrating the results of WP4 research as well as addressing user evaluation feedback in a satisfactory manner, complemented by some other search/upload functionalities and UI/UX improvements. The following subsections elaborate on the details of these activities.

5.4.1 Integration with Partner Services/Tools

The DBKF team at ONTO has collaborated with partner USFD to integrate relevant functionalities in the Verification Assistant. One of them is the so-called DBKF text search, triggered alongside the other tools in the webpage link analysis. When the user provides a URL, the Verification Assistant scrapes the text, media and links from the webpage and calls a multitude of services on those items, one of which is the DBKF Search Service endpoint. It checks for documents in the database, similar to the text on the webpage, and flags them to the user as warnings (see Figure 30 below).

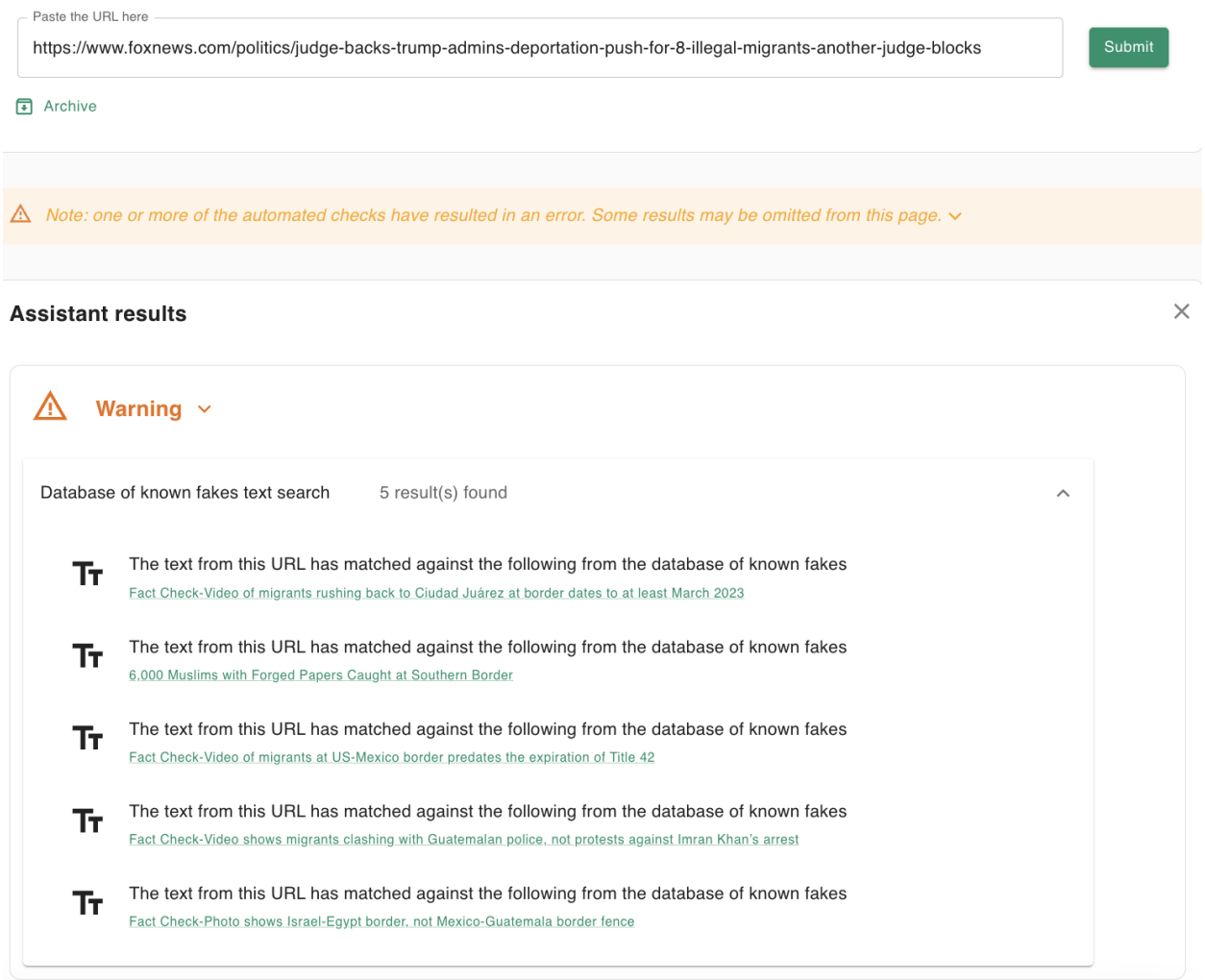

Figure 30 Example of DBKF text search in the Verification Assistant

5.4.2 Other Technical Developments

This section outlines the multiple technical developments to the DBKF over the past year and a half, excluding the functionalities that have been improved as a result of evaluations, which are described in a subsequent section. Here the focus is on:

- extended content collection;
- novel capabilities (Cluster Search and Chatbot), integrating WP4-led research;
- new event type search facet and support for different document types;
- other improvements to the overall system, services and UI/UX.

Content Collection

In order to expand fact-checking content coverage of the database, the DBKF team took advantage of the Google FactCheck Explorer API. The integration of this API was partially carried out in the scope of the Digital Europe project BROD (grant agreement No 101083730), and in the part where it was a requirement of the project to extend coverage of Bulgarian and Romanian sources, and the inclusion of a wider pool of fact-checking publications from other countries was done within vera.ai. The Google FactCheck Explorer API allows to query ClaimReview[55] structured data and provides information about the following objects: claim, author, publishing organisation. Claim reviews, appearance and evidence links cannot be retrieved through this API, which is the reason for the zero values in the respective columns in Table 4 further below.

Combining the claims from the Google FactCheck Explorer API and DBKF custom ingestion components, claims count has increased with more than 50,000 claims, amounting to around 190,500 claims in total available for search in the database (as of 28 August 2025). The DBKF now covers well over 70% of IFCN-approved organisations and almost 60% of EDMO fact-checking community members as well as almost 80% out of the 24 official EU languages. In terms of geographic diversity, the database has extended coverage of sources from France, Germany and Italy, and already ingests content from fact-checkers in countries that were previously not present such as Austria, Belgium, Denmark, Estonia, Greece and the Netherlands. Overall statistics for claims per source and distribution by language are presented below (Table 5 and Figure 31).

Table 5 Statistics about fact-checking sources in DBKF

Source	Claims	Appearances	Evidences
AFP	41,753	234,946	600,058
Snopes	23,165	34,656	135,751
Leadstories	12,425	24,606	113,480
CheckYourFact	8,903	17,731	86,225
Maldita	8,649	10,235	70,774
Poligrafo	7,975	2,229	41,246

[55] https://developers.google.com/search/docs/appearance/structured-data/factcheck

DemagogPL	7,506	0	0
FullFact	7,210	8,210	86,150
AfricaCheck	6,603	11,853	67,945
PolitifactFB	6,380	5,572	42,345
Factly	5,996	11,137	69,270
Boomlive	4,625	39,355	48,629
Reuters	3,634	12,776	32,036
Newtral	3,588	6,700	34,779
Correctiv	3,490	4,988	50,014
TaiwanFC	3,297	0	20,127
Ellinika Hoaxes	3,227	0	0
Observador	3,153	0	0
VoxUkraine	2,727	4,977	23,870
Проверено.Медиа	2,378	21,058	21,074
Dubawa	2,334	0	0
StopFake	2,072	6,055	8,639
VerificaRTVE	1,829	10,791	10,608
Science Feedback	1,454	4,373	27,756
EFEVerifica	1,435	0	0
Delfi	1,414	1,918	14,553
Facta	1,352	2,192	11,938
Verificat	1,260	899	11,958
Pagella Politica	1,170	0	0
FactCheck	1,168	295	13,280
ReBaltica	1,156	2,772	22,847
Tjekdet	835	0	0
IrishJournal	679	168	8,529
20 Minutes	662	0	0
Fakenews.pl	593	0	0
FactCheckBG	514	758	4,046
Knack	463	0	0

Factual.ro	457	315	0
Bayerischer Rundfunk	426	0	0
Franceinfo	394	0	0
investigatebel.org	346	0	0
DemagogCZ	317	564	6,316
Nieuwscheckers	310	0	0
Austria Presse Agentur	281	0	0
Greece Fact Check	276	0	0
FactReview	162	0	0
Deutsche Welle	149	0	0
Factcheck.Vlaanderen	74	0	0
BNT	70	0	0
Open	30	0	0
Faktabaari	19	0	0
France24	17	14	166
Health Feedback	15	0	0
Climate Feedback	7	0	0
VRT	3	0	0
AGI	3	0	0

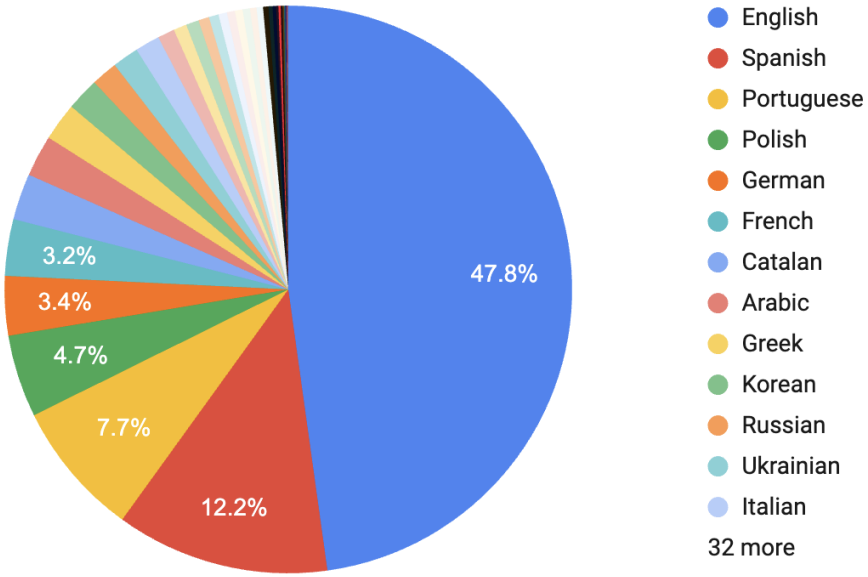

Figure 31 Language distribution of claims in DBKF

Narrative Cluster Search

One of the user requirements, elicited at the start of the project, indicated the ability to detect disinformation narratives in multiple languages. With this need in mind, the ONTO research team developed an innovative AI-driven approach for multilingual clustering of disinformation claims within the scope of WP4. The method is well-described in deliverable D4.2 Coordinated sharing behaviour detection and disinformation campaign modelling methods. The technical implementation of the clustering process in WP5 made the resulting clusters accessible through the DBKF interface to facilitate the discovery and analysis of cross-national and multilingual disinformation narratives.

The Cluster Search loads as a separate screen[56] in the DBKF UI, where all precomputed clusters are listed and the one with the most recent end date is on top. In addition to the usual sorting options by date and relevance, these groups of similar claims can be ordered by the size of the cluster, in other words the number of the claims therein (see Figure 32 below). For more efficient navigation users can combine it with the faceted filters and concept search to narrow down the clustered data based on various criteria.

[56] https://dbkf.ontotext.com/#!/clusterSearchViewResults

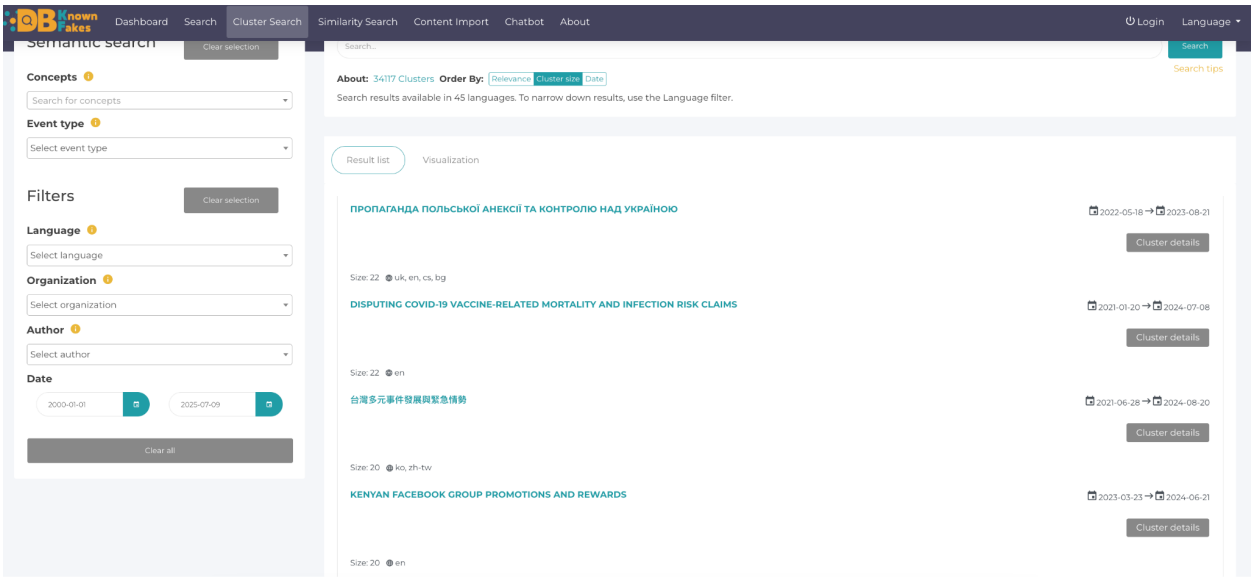

Figure 32 Cluster Search Result list, sorted by cluster size

Each cluster is represented by its cluster card (see Figure 33 below), which contains the most important metadata such as the size, the languages of claims and the start-end dates (these indicate when the earliest claim, respectively the latest claim were published).

Figure 33 Example of a cluster card

A closer exploration of each cluster is enabled through the Cluster details screen, which in addition to the basic metadata includes a list of the key concepts. A vertical timeline visualisation of the claims provides a more intuitive way to follow their distribution in time (see Figure 34 below). This was inspired by the DBKF participatory evaluation session, where media professionals indicated the need to see how narratives evolve over time.

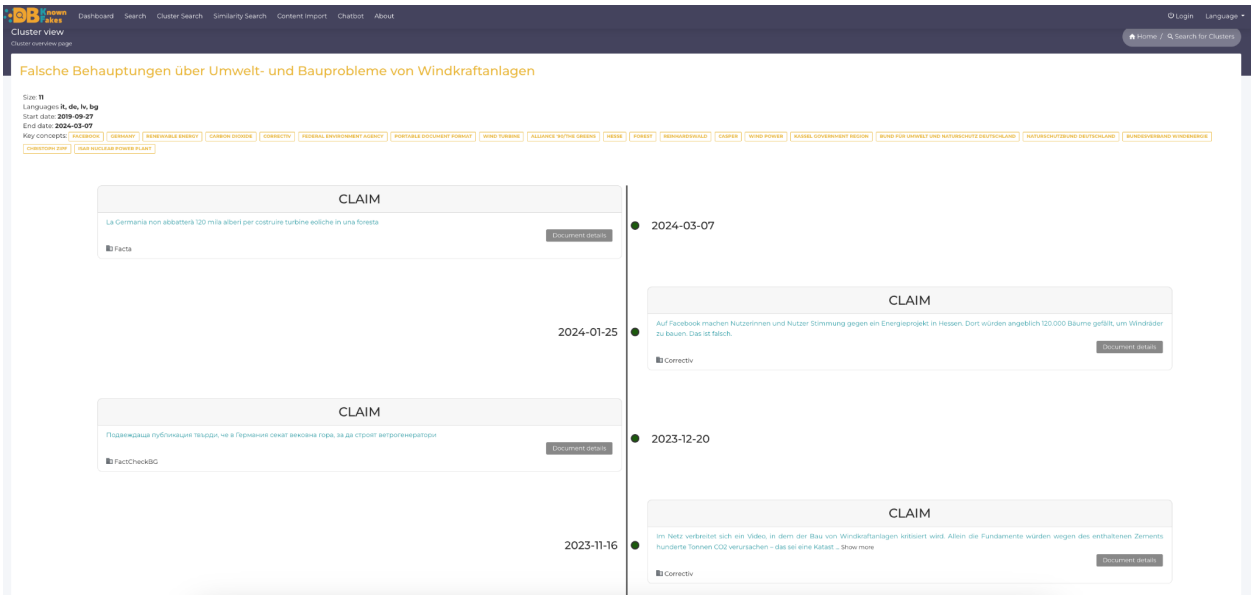

Figure 34 Example of a cluster view screen

Besides the Result list tab, the Cluster Search screen also has a Visualisation tab that plots the start and end dates of clusters from the current page of results (see Figure 35 below). This provides another perspective to analysing content reuse.

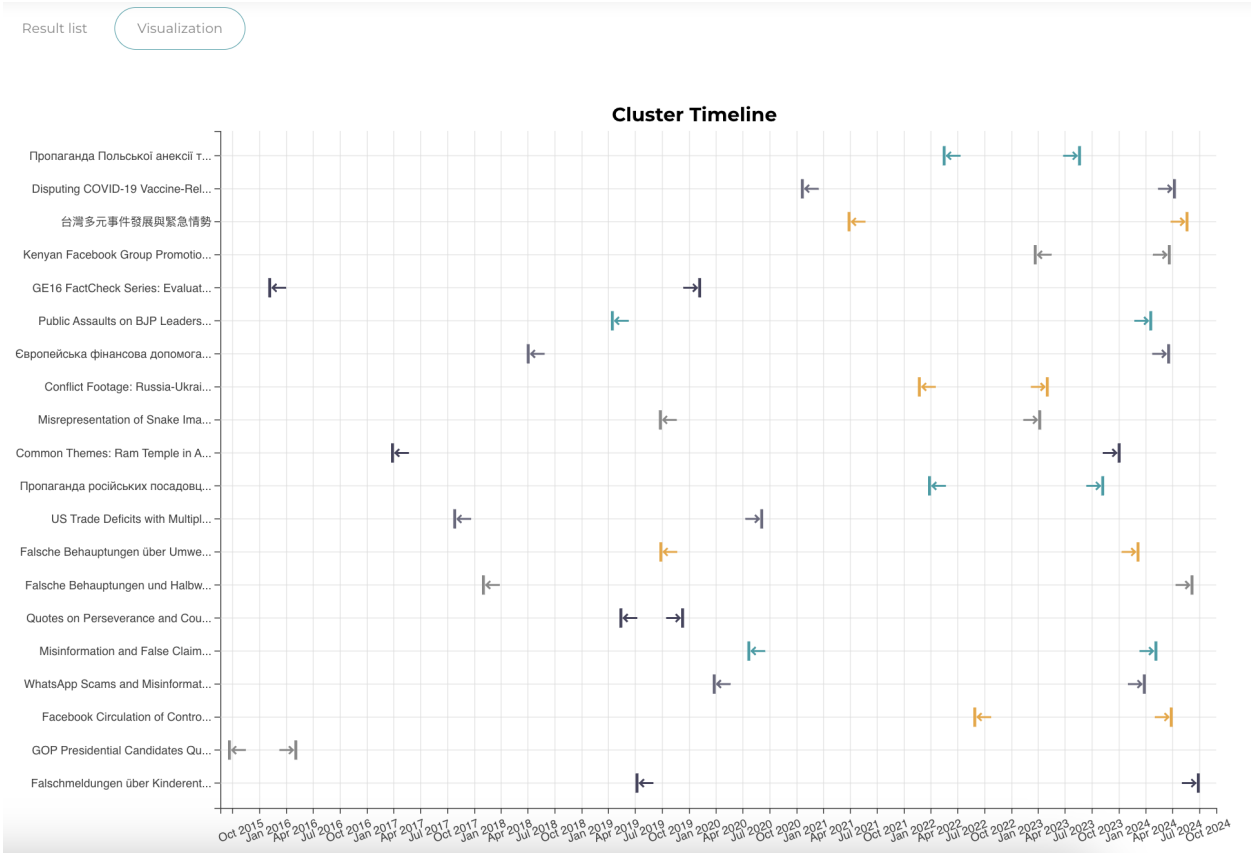

Figure 35 Example of a cluster visualization screen

DBKF Chatbot

In order to get to know and benefit the most from the diverse DBKF search capabilities, it is helpful especially for new users to have assisted interactions with the database. That is why the DBKF team integrated a chatbot functionality, developed by ONTO researchers within WP4 (more details are provided in deliverable D4.2). This AI agent, powered by OpenAI Assistants API, is potentially capable of leveraging the different tools created by partners in WP4, but in its specific DBKF implementation it has access to all the search functionalities of the database.

The DBKF Chatbot is developed as a Python service with REST endpoints to create a conversation with an AI agent, exchange messages with it and get explanations for its actions. There is a separate screen in the DBKF UI[57], where users can ask their questions in natural language, in a language of their choice. The chatbot can respond to queries about all the available data (claims and claim reviews) and metadata (authors, publishing organisations, publish dates) as well as the enrichments to it (concepts, events, clusters). At the bottom of the conversation there is an Info icon, which allows users to understand how the chatbot came to the response such as what query methods it used. Figure 36, Figure 37 and Figure 38 shows screenshots of sample conversations.

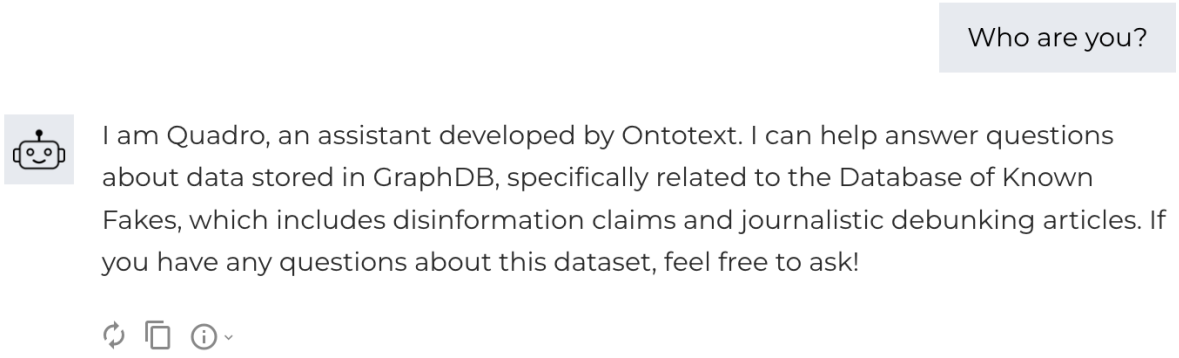

Figure 36 DBKF Chatbot conversation: what the chatbot knows about itself

[57] https://dbkf.ontotext.com/#!/chat

Give me the most recent claim clusters.

Here are the most recent claim clusters:

1. **Vitamin A and measles prevention/treatment**
   - Date First: November 3, 2025
   - Link to Cluster
2. **Common Side Effects of Pfizer-BioNTech COVID-19 Vaccine**
   - Date First: October 1, 2025
   - Link to Cluster
3. **Unauthorized Access to Bank Accounts**
   - Date First: May 15, 2025
   - Link to Cluster
4. **Misinformation about COVID-19 vaccine effects**
   - Date First: May 2, 2025
   - Link to Cluster
5. **Misleading Fox News Chyron on Trump El Salvador Sentencing**
   - Date First: April 23, 2025
   - Link to Cluster

If you need more information on any of these clusters, feel free to ask!

Figure 37 DBKF Chatbot conversation: request to provide the most recent clusters

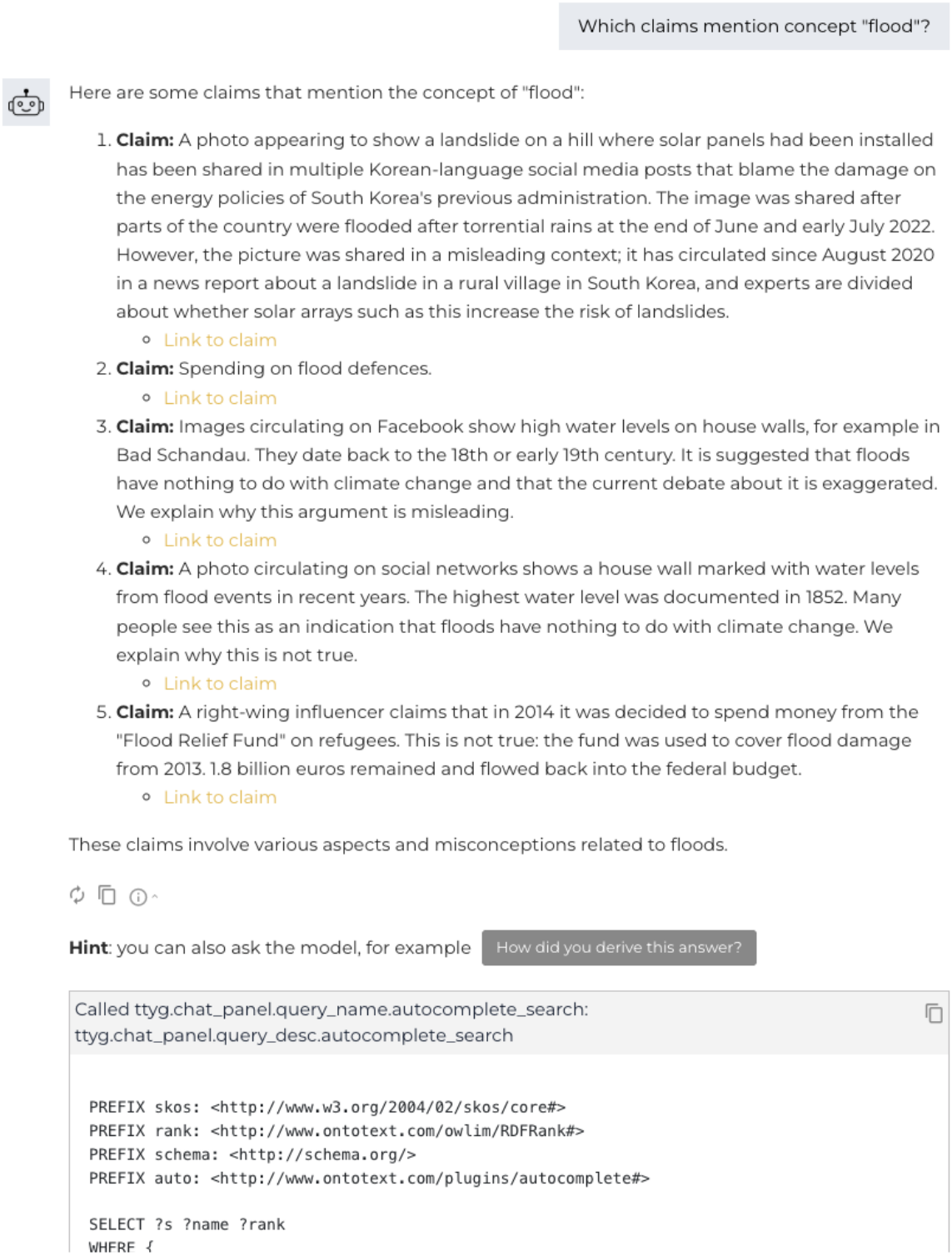

Figure 38 DBKF Chatbot conversation: request to provide claims, containing a certain concept, plus an excerpt of the explanatory message

Event Type Search

As a continuation of the work described in D5.2 (section 5.4.2, subsection *Content Enrichment with Events*), the DBKF team integrated an Event Detection Service, developed within the scope of the VIGILANT[58] project. The event extraction model identifies 36 event types in English, hence only English language content in the DBKF was re-annotated to enrich it with this information. A dedicated search functionality was then enabled through a facet in the UI. The various event types are listed in a multiselect dropdown box, which works as a logical "OR". For example, if the user chooses "Attack" and "Demonstrate", the search will return results that contain either "Attack", or "Demonstrate" (see Figure 39 below).

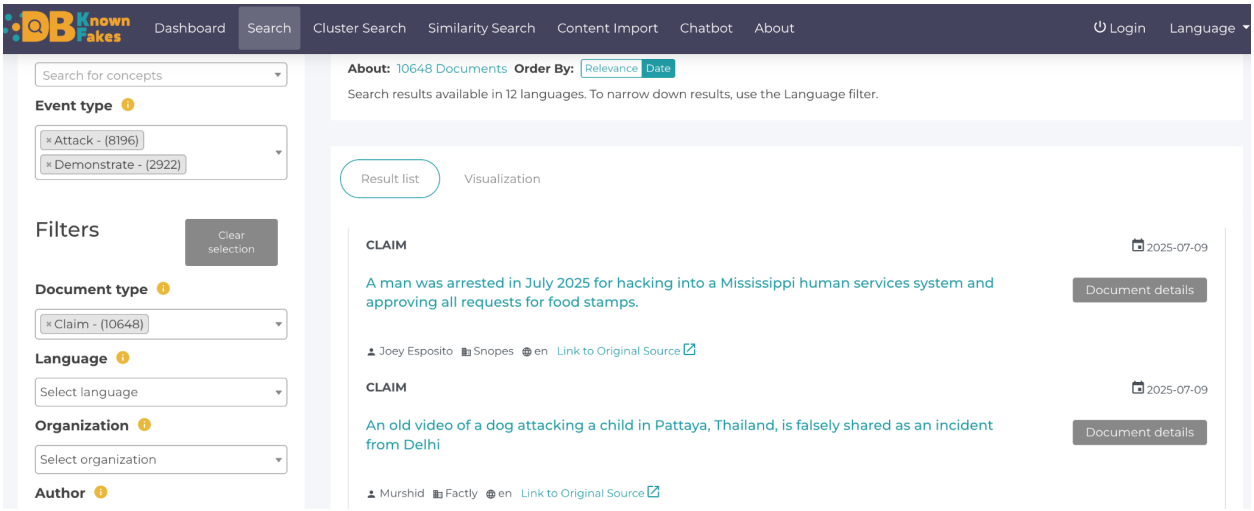

Figure 39 Event Type search results for "Attack" and "Demonstrate" events

Similar to the concept quotes, the Document details view now also features a dedicated section with event quotes. The so-called trigger word, which describes best the particular event, is highlighted in bold for better visibility. The excerpts for each event instance are contained in separate boxes, and the type of the event is written in the bottom left part of the box (see Figure 40 below).

[58] VIGILANT is a Horizon Europe project, funded under grant agreement No 101073921. Project website: https://www.vigilantproject.eu/

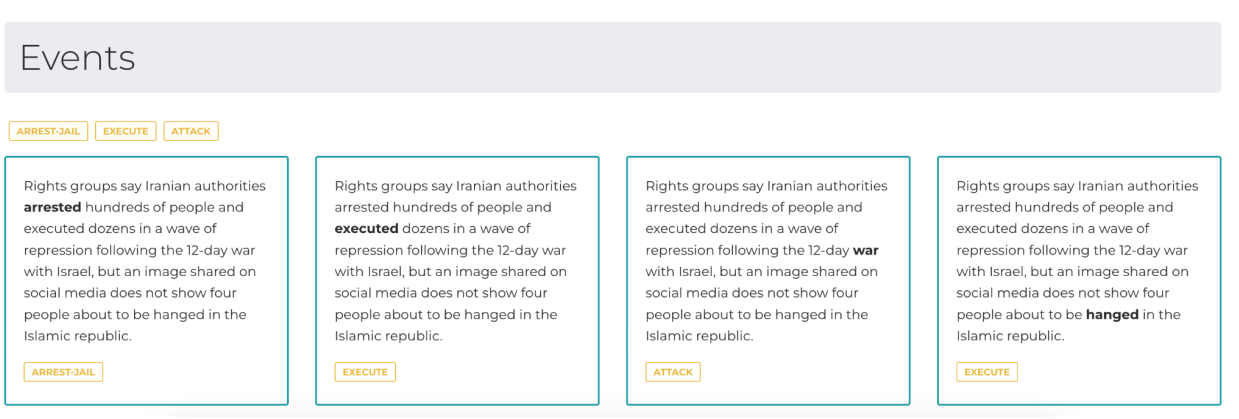

Figure 40 Events section with quotes

Document Type Search

The DBKF Search Service was completely redesigned to allow for exploration of debunks on a more granular level (of claims and claim reviews) as well as to enable search for other types of documents such as social media posts for the needs of WP4. As a result, there are now five types of objects that the database supports (see Figure 41):

- Claims – There are more claims than claim reviews because the Google FactCheck Explorer API retrieves only claims.
- Claim reviews – In the Document details view, there is a section for Linked documents, which allows switching between the claim review and the associated claim.
- News articles – These are text documents, extracted from appearances. They are interesting from the perspective of analysing what media sources disinformation claims refer to.
- Social media posts – X/Twitter, TikTok, Facebook
- Comments – These are comments under social media posts. No such data is currently available in the database.

It should be noted that social media posts and comments are not sourced in an automated way, rather through a manual content ingestion service. This service was originally developed for the needs of stakeholders in the VIGILANT project but has proven useful for adding content to the DBKF, in particular for the purpose of building a proof of concept for the integration of WP4 partner services that work on different types of documents other than debunks.

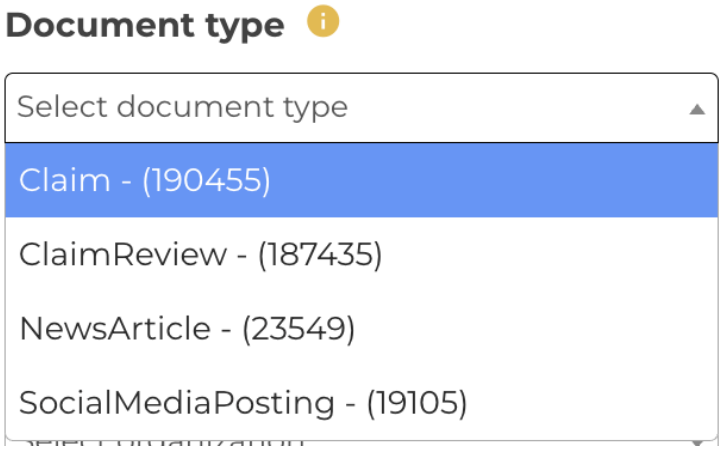

Figure 41 Document type filter

Multiple tiny but necessary UI changes were done to reflect the support of different document types such as dynamically changing the document type, written on the header of a document card; replacing the "Link to Debunking Article" button with a hyperlinked text saying "Link to Original Source". In addition, the DBKF Dashboard page now displays statistics for claims only, as it would be too confusing and overwhelming to visualise these for all types of documents at once.

Miscellaneous Improvements to System, Services and UI/UX

An important improvement to the overall system setup and management is the introduction of ArgoCD[59] – a tool for deployment automation. This has significantly streamlined the process of service deployment across all three environments: DEV, STAGE and PROD.

In terms of enhancements to services and components, the main lines of work included:

- Maintenance of the content ingestion components and creation of a last ingest claim date healthcheck to continuously monitor their status;
- Updating the claim extraction logic for a couple of sources (Snopes, Africa Check, FullFact) so as to pick the more appropriate field for claim from the html of their fact-checks;
- Improvement to the Concept Search API behaviour – Previously, when performing the matching by Wikidata concept label, the pipeline put a higher weight on the preferred label (the one in English) and lower weight on the alternative labels in languages other than English. This sometimes resulted in concepts not surfacing as suggestions or appearing at the very bottom of the list when written in a language other than English. An improvement in the ranking of results has been noticed after tweaking the algorithm to treat preferred and alternative labels with an equal weight.
- Modifying the Language Enrichment Service – Previously, oftentimes more than one language was associated with a debunk because the service picked all the different languages, contained in the text. For example, here we have an excerpt from a claim review from *AFP* that is predominantly in Bulgarian but also contains phrases/sentences in English: “Лекарят във видеото е Джефри

[59] https://argo-cd.readthedocs.io/en/stable/

Барк, основател на организацията America's Frontline Doctors, чиито подвеждащи и неточни твърдения са били многократно опровергавани от AFP, например тук и тук. В клипа Барк чете листовката на Vaxelis – ваксина, която едновременно предпазва от дифтерия, тетанус, коклюш, хепатит B, полиомиелит и Haemophilus influenzae тип b." The service has been tweaked to associate the claim/claim review with the language with the highest probability (i.e. the predominant language of the text). Documents with more than one language tag were re-annotated to apply this fix.

- Adjusting the appearance and evidence extraction logic for *AFP*, as its fact-checking sites around the world introduced specific fields for these items.

On the UI/UX side, some minor modifications – not insignificant from the UX perspective – have been done (see Figure 42), such as:

- Replacing the yellow colour of all links (for appearances, evidences, source text, etc.) with green, which is more readable and pleasing to the eye.
- Keeping the headline text displayed on a document card with the same length for all documents, also adding "Show More" and "Show Less" for longer headlines.
- Displaying the count of languages for a search query and hinting users to use the language filter to narrow down their choice, if needed.

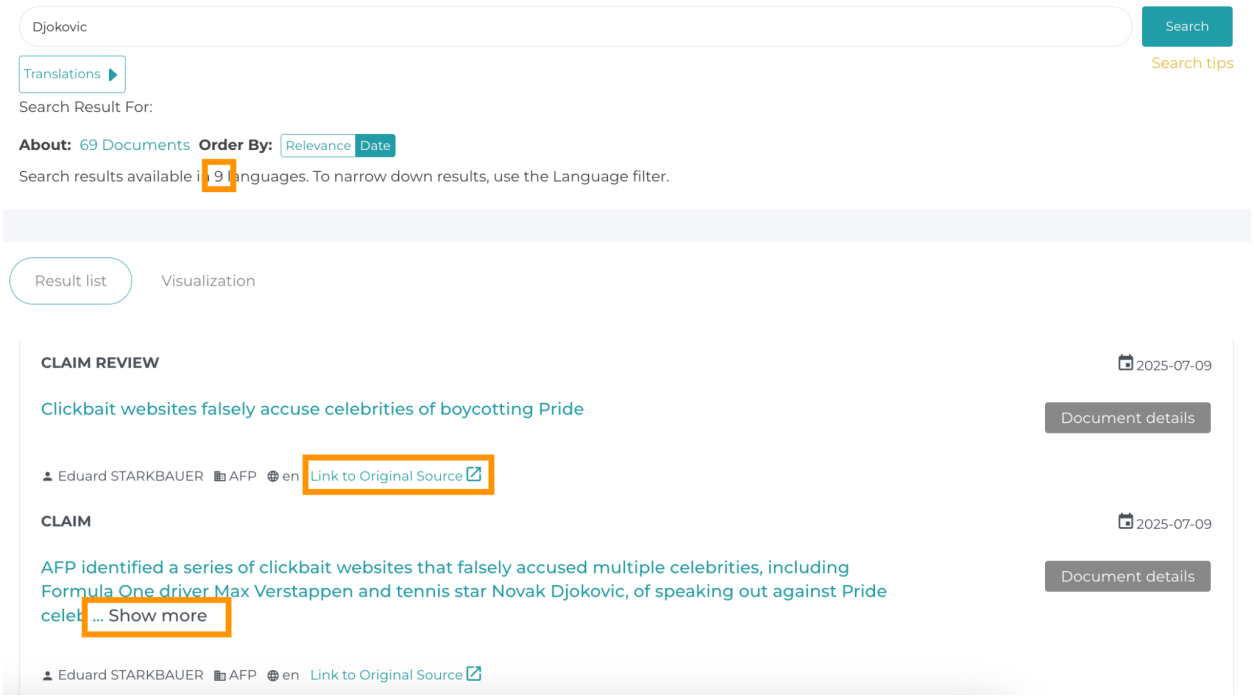

Figure 42 UI changes example for a sample search query ("Djokovic")

5.4.3 Improvements and Updates Following Evaluation Feedback

With the close support of and collaboration with WP2 partners EBU and DW, the database went through a thorough participatory evaluation session that focused on the Multilingual Search in particular. Alongside their feedback on this specific functionality, participants provided the DBKF team with a lot more valuable comments on how to improve the system.

As a result of the evaluation, the Multilingual Search has transformed into a Translated Search. It is no longer a separate screen, which prevented users from easily switching between the basic monolingual to the multilingual option. The workflow is now the following: Once the user runs a full-text search, they are offered to explore the translations of their search term/phrase. These translations are now visible unlike before and are also selectable. By default, the language on top is the original language of the typed-in word(s). The user has to click "Apply" to confirm the selection and trigger the search (see Figure 43 below).

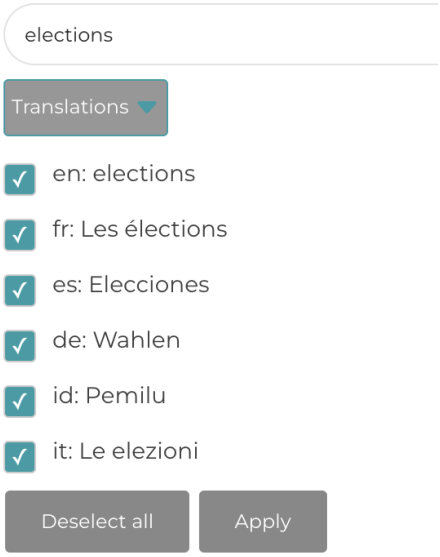

Figure 43 Example of a Translated Search

Other notable changes introduced to the UI, based on the evaluation feedback, included:

- Improving readability of headlines by lowering the size of the font and un-bolding them.
- Adding organisation and language to the key metadata on the document card, as media professionals indicated it is of utmost importance for them to know the name of the organisation rather than the author.
- Replacing "published x days ago" with the exact publishing date.

5.5 Challenges and Lessons Learned

The development and upkeep of a complex system such as the Database of Known Fakes is accompanied by challenges on different levels such as:

- Data collection – The database would be of little value without data to explore and analyse. At the same time, access to data and metadata (rich, well-structured) continues to be problematic

(because of the specifics of each fact-checking organisation) or utterly impossible (in the case of live data retrieval from X/Twitter or Facebook).

- Maintenance – The complexity of the system, the growing number of services with higher processing demand - all of this requires a constant effort to keep running, both in terms of human involvement and expertise as well as proper infrastructure.
- Reliance on external components – Providing users with an intuitive conversational interface, driven by the latest advances in LLMs, comes with a caveat: the DBKF Chatbot is not always consistent in its behaviour. This is on one hand due to the nature of LLMs but is also complicated by the upcoming deprecation of the OpenAI Assistants API. The team is working to mitigate this situation by developing an updated Chatbot version that would combine OpenAI Completions API[60] and LangGraph[61].

5.6 Next Steps beyond vera.ai

As outlined in the vera.ai Final Exploitation and Sustainability Report (deliverable D6.3), the Database of Known Fakes, along with the advanced search and exploration tools it provides, serve as a valuable showcase of ONTO's strong capabilities in semantic technologies and in leveraging the combination of AI and knowledge graphs. The services and models employed can be reused in solving other (including commercially viable) use cases, where analytical tasks could benefit from searchable data enrichments. Therefore, the DBKF team will aim to sustain a locally deployed instance of the database with selected content to use as a demonstrator for potential customers. This instance will be publicly available so it could still be useful in the work of fact-checkers and disinformation researchers. The team at ONTO could also seek opportunities for further development of DBKF functionalities through involvement in other relevant EC-funded projects or through commercial avenues.

[60] https://platform.openai.com/docs/api-reference/completions
[61] https://github.com/langchain-ai/langgraph

6 Integration with Relevant AI Platforms

In the scope of T5.6, consortium partners made available the WP3 and WP4 AI-based verification technologies through relevant AI platforms, with special emphasis on AI4EU and ELG (the European Language Grid). The aim was to make vera.ai outputs readily available to the AI research community and companies wishing to use or integrate such technology in their products and generally to increase the reach and impact of vera.ai's work. Integration with such platforms and repositories is additionally a key action for strengthening the exploitation and sustainability potentials of the project's technologies, as it increases discoverability, widens access to practitioners and researchers and facilitates re-use by other projects and industry.

In total, **19** AI assets were contributed to the AIoD and ELG platforms (exceeding the KPI of 10 assets contributed), whereas a dedicated project page was created in both platforms[62]. Most of the contributed assets are ML models and datasets. Project announcements and events were also shared through the AIoDP to increase their reach[63]. Other synergies with the AI4EU community include the delivery of an AI-on-Demand webinar by vera.ai partners, titled "AI Meets Media: Tackling Disinformation with Cutting-Edge Innovation"[64], presenting vera.ai's tools on synthetic media detection.

To facilitate the work, ATC contacted the administrators of the platforms to set up the project pages and created tutorials to guide the partners through the process.

[62] For reference see https://live.european-language-grid.eu/catalogue/project/23623 and https://www.ai4europe.eu/ai-community/projects/veraai.

[63] See https://www.ai4europe.eu/ai-community/projects/veraai?category=event and

[64] https://www.ai4europe.eu/news-and-events/events/ai-meets-media-tackling-disinformation-join-our-free-webinar-april-3rd

7 Conclusion and Outlook beyond vera.ai

Over the final reporting period, vera.ai's end-user tools - the Verification Plugin, the Verification Assistant, Truly Media, and the Database of Known Fakes (DBKF) - matured along three axes: (i) systematic integration of WP3/WP4 research outcomes into production workflows, (ii) explainability-first UI/UX patterns informed by participatory evaluations, and (iii) tighter interoperability across independently deployed services. The result is broader modality coverage (text, image, video, audio) and smoother hand-offs between tools, while keeping cognitive load low and interpretation help one click away. The emergence and proliferation of ChatGPT and similar LLMs underscored the dynamic nature of AI development and its impact on media production processes. The results of this process also highlighted the importance of real-world data in refining AI models, underscoring in parallel the ongoing challenges and complexities inherent in deploying AI technologies in real-world scenarios.

Overall, the active guidance provided by industry partners to research partners throughout the integration process played a crucial role in achieving successful results. This emphasises the significance of industry and research collaboration right from the project's inception, highlighting that a lab-to-market transfer process requires joint efforts. Moreover, the direct involvement of end-users in iterative and agile development processes further amplifies the potential market adoption of AI-related innovations, fostering a user-centric approach and ensuring the practical relevance of the developed solutions.

The work that was implemented throughout the duration of WP5 highlights the challenges and the problems that are entailed in the process of transferring research-intensive AI technologies into end-user products, platforms, and tools, as well as in the process of industry-research collaboration. Through these challenges (as briefly discussed below), both industry and research partners have accumulated valuable experiences and knowledge that allow them to design effective mitigation strategies and responses and pave the way for a more successful and effective collaboration beyond the scope of the project.

In total, throughout the duration of WP5, **23** AI components were integrated in the end user tools, along with several other technical developments in the tools themselves, following feedback from the user evaluation sessions, but also in response to market and technological developments.

Integrating research-intensive AI technologies into media systems and tools presented several technical challenges and associated risks. These encompass technical complexities, regulatory compliance, user trust, scalability, interoperability, data privacy, performance optimisation, dependency management, testing complexity, and change management. Below is a concise summary of the challenges, risks, and mitigation measures identified in the course of work carried out in WP5:

Explainability and trust as product features: Across tools, making complex outputs actionable for professionals consistently was prioritised over adding raw model capacity. Teams converged on progressive-disclosure interfaces (headline takeaways with drill-downs), traffic-light gauges, per-segment/timeline visual cues, and contextual "how it works" support to prevent misreads and encourage informed use. This design stance emerged directly from user evaluations.

Managing false positives and model drift demands process, not just models: The Verification Plugin introduced a staged release flow (evaluation/beta) to contain the impact of immature detectors and

explicitly called out the risk of propagating errors through Near-Duplicate Detection (NDD) - a reminder that infrastructure can amplify model mistakes if governance is weak.

Platform volatility is a constant challenge: The Verification Assistant's extractors break when social platforms change; automated scraper tests and user-facing fallbacks (e.g., upload local files) are now part of the operating playbook. Likewise, the Verification Plugin's social-network analysis (SNA) re-architecture after Twitter/X API deprecations underscores the need for resilient data-collection patterns.

Scale and external dependencies shape user experience: Truly Media's integrations highlighted latency, black-box modules, and service downtime as practical risks; mitigations include polling, caching, proactive communication with component owners, and careful cost control for containerised/on-premise deployments. Interoperability among heterogeneous partner components also requires change management and shared practices, not only APIs.

Human-AI collaboration: Ensuring effective integration of AI technologies within existing workflows, particularly for tasks requiring human supervision and decision-making, is a significant challenge for both research and industry partners. Collaborative approaches that emphasise human-AI interaction and facilitate smooth integration are necessary.

Data availability and maintenance remain foundational: the DBKF's value is bound by the volume/quality of ingestible fact-checks and the stability of external APIs. The team reports persistent challenges with source heterogeneity, system upkeep, and the fragility of LLM-based components, prompting architectural adjustments.

Concluding this work, significant advancements have been made throughout the vera.ai project in the integration of AI technologies into various media-related use cases, paving the way for continued innovation and collaboration in the field. The project's findings and initiatives offer valuable insights and opportunities for future endeavours. Partnerships forged during the project have facilitated knowledge transfer and collaboration between the media industry and the AI research community. Leveraging the knowledge and experience generated through vera.ai, partners can drive product development, consulting services, and new research initiatives.

Looking beyond the vera.ai horizon, all partners are committed to sustaining and maintaining their tools through their own resources and new funding. We briefly present below the plans outlined for the continuation of the operation of the end-user tools.

Verification Plugin: The Verification Plugin will continue as a maintained, free and open-source front end, with a post-project roadmap shaped by user needs. Priorities include watermarking detection for provenance cues, richer SNA (including coordinated-behaviour views), narrative-analysis and video-highlighting tools, optional local history with shareable links, renewed Firefox support, and Telegram-to-RSS conversion—implemented under the same staged-rollout discipline to protect trust.

Verification Assistant: Sustainability hinges on securing follow-on funding. Near-term technical work focuses on code refactoring, continued Python upgrades, and robust extraction from JavaScript-heavy sites—paired with ongoing maintenance to keep pace with fast-moving platform changes.

Truly Media: As an established commercial product used by major newsrooms and EDMO members, Truly Media will keep integrating research outputs where maturity and licensing allow. ATC/DW will leverage consortium know-how in new R&D and industry collaborations (e.g., AI-CODE), explore service/SLA frameworks for partner components, and complete a major platform redesign that migrates vera.ai integrations with tighter workflow fit and a modern technological stack.

Database of Known Fakes (DBKF): ONTO will sustain a publicly available, locally hosted demonstrator with selected content as an exploitation showcase for knowledge graph/AI capabilities, while pursuing further development via EC projects or commercial avenues. In parallel, the chatbot stack will be reworked (e.g., Completions API + LangGraph) to mitigate upstream LLM/API changes.

Looking forward, cross-cutting recommendations for future work include **keeping explainability central** by preserving progressive-disclosure UIs, embedding provenance cues (e.g., C2PA where available), and standardising interpretation aids across modalities; **institutionalising staged evaluation** by treating model onboarding (evaluation → beta → public) and telemetry/feedback loops as part of the governance structure and not as ad-hoc practices; **designing with dependency risk in mind** by monitoring partner services, defining SLAs, and building modular, gracefully degrading integrations; **investing in data pipelines** and expanding coverage, normalising metadata, and maintaining multilingual reach to underpin DBKF and Assistant features; and finally **planning for constant platform churn** with automated extractor tests, rapid hot-fix capacity, and user fallbacks (e.g., local uploads) to keep workflows resilient when the web changes.

In sum, vera.ai leaves behind a robust, interoperable toolchain that demonstrably answers user requirements while codifying practices (explainability, staged releases, resilience to external change) that will keep these tools reliable in production. The partners have laid out credible, tool-specific sustainability paths and a shared posture toward exploitation and further research, ensuring that the project's results continue to reach - and benefit - professionals beyond the project's lifetime.

vera.ai

vera.ai is a Horizon Europe Research and Innovation Project co-financed by the European Union under Grant Agreement ID: 101070093, an Innovate UK grant 10039055 and the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number 22.00245.