Ref. Ares(2025)9382589 - 31/10/2025

CH%#
!#ASE

Combatting online
HAte Speech by
engaging online mEdia

D 3.8
Final Version of CHASE ICT Tool

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Deliverable's Information

Deliverable	**Visualisation Component**
Deliverable No.	D3.8
Task	Final Version of Chase ICT Tool
Task No.	T3.2
Task leader	ITML
Task contributors	Lampros Argyriou (ITML), Andreas Siafakas (ITML)
Deliverable due date	31/10/25
Dissemination level	PU
Version	1

Document History

Version No	Date	Contributor(s)	Organisation	Notes
0.1	26/10/2025	Lampros Argyriou	ITML	Initial Draft
0.2	28/10/2025	Andreas Siafakas	ITML	Internal Review
0.3	29/10/1025	Lampros Argyriou	ITML	First Version of D3.8

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Table of Contents

D3.8 Final Version of CHASE ICT Tool 1
Executive Summary 4
Introduction 5
Purpose of the Document 5
Structure of the Document 5
Deployment and User Management 6
Overview of the deployed CHASE platform environment 6
User Authentication and Management Process 7
Backend and Data Layer Updates 9
N-gram computation workflow 9
Asynchronous Text Analytics Execution 10
Integration of LLM Tracing 10
Conclusion 13

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Executive Summary

This document presents the final version of the CHASE ICT Tool, developed under *Work Package 3 (WP3) – CHASE Technology Components and Platform*.

Building upon the technical specifications and incremental implementations described in earlier deliverables (D3.1, D3.3, D3.5, and D3.6), this report consolidates the complete, deployed version of the CHASE system and summarizes the latest technical refinements leading to its full operational readiness.

With these refinements, CHASE ICT tool reaches its full functional maturity, ready for pilot deployment and real-world validation by partner organizations. It represents the consolidation of all technological achievements within WP3 and establishes a robust foundation for future operational use and sustainability.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Introduction

Purpose of the Document

This deliverable presents the final integrated and operational version of the CHASE ICT tool. It describes the system in its completed form, summarizing the final updates that have been implemented.

The document highlights three main aspects:

- The integration of the platform's components into a unified operational environment supporting multi-partner access and interaction.
- The finalized configuration and management of user access.
- The latest enhancements to data and analytics workflows, developed to improve visualization and analytical responsiveness.

Overall, the document serves as the closing technical report of WP3, demonstrating that the CHASE ICT tool has achieved the intended level of integration and readiness for validation in the pilot phase.

Structure of the Document

The rest of this document is structured as follows:

- Section 2 describes the approach used for managing user access and authentication across partner organizations.
- Section 3 outlines the latest improvements to the system's internal workflows.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Deployment and User Management

Overview of the deployed CHASE platform environment

The final version of the CHASE ICT tool has been fully deployed on a dedicated **Kubernetes (K8s)[1] cluster** hosted on ITML's infrastructure. Each service has been containerized using Docker[2]. This setup provides a secure, controlled, and scalable runtime environment that enables fine-grained operational control. All system components are containerized and managed through Kubernetes.

The cluster is configured to orchestrate service deployment, scaling, and lifecycle management through declarative manifests. Each component is deployed as a separate Kubernetes pod, grouped into namespaces corresponding to the major subsystems of the CHASE ICT tool. Stateful services, such as OpenSearch and Keycloak, make use of **Persistent Volumes (PVs)** to ensure data durability across restarts and upgrades. Ingress controllers expose only the necessary public endpoints (e.g., the CHASE web interface and Keycloak authentication portal) over **HTTPS**, while internal service communication remains restricted to the cluster's private network.

Inter-service communication between CHASE components relies on a combination of **REST APIs** and asynchronous event streaming via **Apache Kafka**. To safeguard this communication layer, **mutual TLS[3] (mTLS)** has been implemented between Kafka producers and consumers, guaranteeing both message confidentiality and bidirectional authentication. This approach ensures that only trusted services within the Kubernetes cluster can publish or consume events, thereby reinforcing the platform's overall security posture.

Continuous monitoring of the deployed environment is achieved using Kubernetes health probes, internal logging mechanisms, and OpenSearch dashboards. This configuration enables real-time observability, fault detection, and simplified maintenance.

In its final deployed form, the CHASE ICT tool operates as a cohesive, production-ready system that integrates all core components in a unified, secure, and maintainable environment, supporting the complete moderation and analytics workflow defined within Work Package 3.

[1] https://kubernetes.io/
[2] https://www.docker.com/
[3] https://en.wikipedia.org/wiki/Transport_Layer_Security

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

User Authentication and Management Process

The implementation of user authentication and access control in the final version of the CHASE ICT tool follows the configuration previously described in *Deliverable D3.1 – Technical Specifications*, where the integration of **Keycloak** as the **Identity and Access Management (IAM)** solution was first introduced. In this final version, Keycloak has been fully embedded in the deployed environment and serves as the central entry point for user authentication and authorization within the CHASE platform.

User management is performed directly through the **Keycloak administration console**, which allows system administrators to create, update, and deactivate user accounts as well as assign them to predefined roles.

Keycloak operates under a dedicated realm named **"chase"**, which defines all active users, clients, and policies associated with the tool. Two main application clients have been configured:

- **chase-client** — representing the CHASE web application, used for user-facing authentication via OpenID Connect.
- **chase-system-client** — representing internal system communication between backend services and the authentication layer.

The process of user registration for the CHASE platform was coordinated centrally by ITML's technical team. Each participating organization was requested to provide the following details for every individual who required access to the tool:

- **Name**
- **Username**
- **Password**
- **Organization name**

Upon collecting this information, the system administrators created a corresponding **group** in Keycloak for each organization represented in the project (e.g., ITML, Proto Thema, MDIG, CESIE, WAN-IFRA, etc.). Each user account was then created manually within the *chase* realm and assigned to the respective organizational group. This group-based structure ensures that user accounts are clearly associated with their organization and simplifies the management of multiple partners within a single environment.

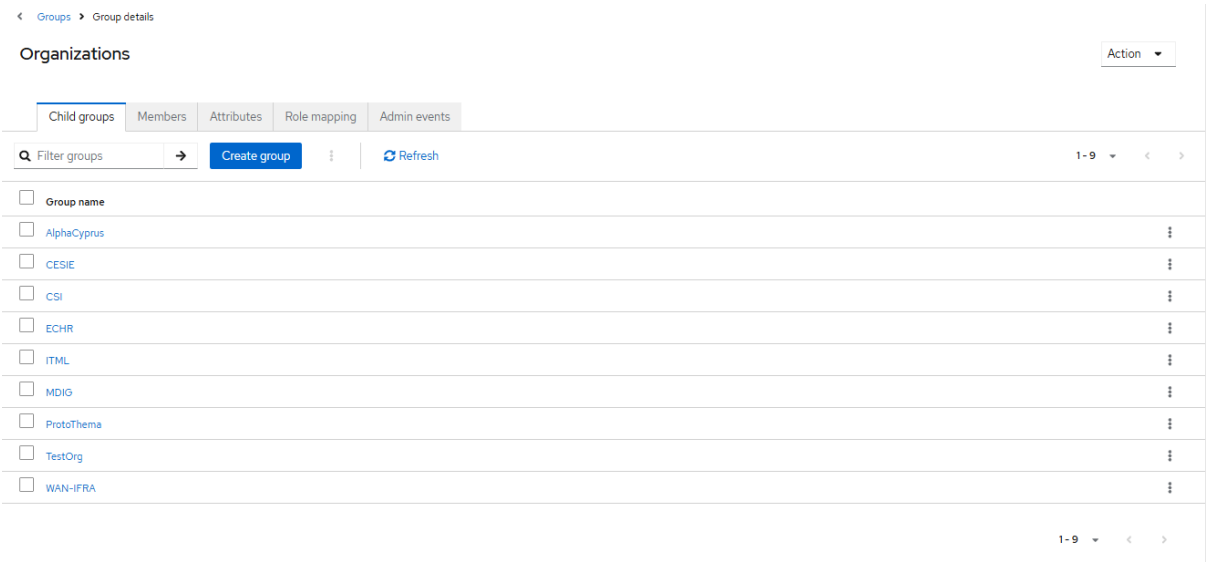

Figure 1:Partner organization groups defined within the CHASE Keycloak realm.

User registration is intentionally **not open to self-signup**. All accounts are created and activated by administrators to preserve control over system access and to maintain a verifiable link between users

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

and their organizations. Password reset and email verification functionalities are disabled, as credential management is handled exclusively through administrative workflows.

This configuration ensures that the entire user management process remains transparent, traceable, and consistent across the consortium. All authentication requests are processed through Keycloak's secure login interface, which integrates seamlessly with the CHASE frontend and backend services (Figure 2). The result is a controlled and auditable access workflow that supports reliable operation of the CHASE ICT tool in a multi-organization context.

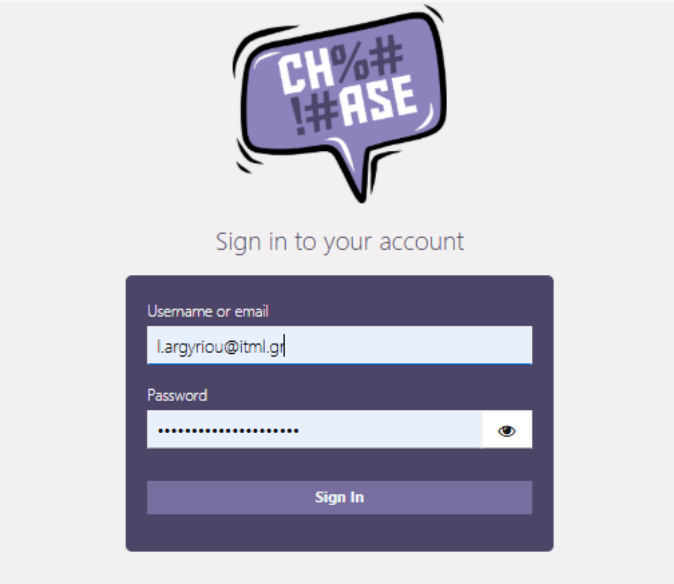

Figure 2:Customized CHASE login page integrated with Keycloak, providing secure authentication for registered users within the CHASE realm.

The CHASE ICT tool is accessible to authorized users through a secure web interface hosted under the project's dedicated domain. Access is provided via the following address:

https://chase-web.itml.space/

This entry point serves as the unified gateway to the CHASE environment, where partner organizations can log in using the credentials assigned to them. The domain supports encrypted communication (HTTPS) and connects directly to the authentication service, ensuring secure access to all functionalities of the CHASE platform.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Backend and Data Layer Updates

Deliverable **D3.3 – CHASE Data Crawler & Data Repository** documented the implementation of the core data infrastructure of the CHASE ICT tool, including the integration of the controller-based backend (CHASE APIs), the adoption of **OpenSearch**[4] for data storage, and the use of **Apache Kafka**[5] for real-time messaging. The deliverable defined the main data flows supporting content ingestion, analysis, and visualization through the indices **chase_posts**, **chase_comments**, and **chase_analytics**, as well as their corresponding Kafka topics. These components established the foundation for end-to-end data processing, from external content retrieval to text analytics and visualization.

Building upon this architecture, the final version of the CHASE ICT tool introduces targeted enhancements to extend analytical capabilities and improve system performance. Two key developments have been implemented:

1. a new **n-gram computation workflow** with dedicated indices to support advanced visualization
2. an **asynchronous execution mechanism** for the text analytics services, enabling faster and more efficient inference.

The following subsections describe these updates in detail.

N-gram computation workflow

To enable the **n-gram visualization functionality** introduced in the visualization component (*Deliverable D3.6 – CHASE Visualization Component*), a new data processing workflow has been developed and integrated into the CHASE ICT tool's backend. This workflow operates as a scheduled task that periodically extracts, processes, and aggregates textual data from the **chase_comments** index to generate n-gram statistics.

Two new OpenSearch indices have been added to support this process:

- **chase_comments_ngram** – storing aggregated n-gram data with frequency counts.
- **chase_comments_ngram_processed** – tracking the processing state of comments to avoid duplication.

The data flow for n-gram calculation follows a structured, automated pipeline. The task executes at predefined intervals (10 minutes). Each run determines the time window to process by checking the most recent entry in **chase_comments_ngram_processed**, or defaults to the earliest available comment if it is the first execution.

Comments within the selected date range are retrieved in batches of ten. For each batch, previously processed comments are filtered out based on their IDs recorded in **chase_comments_ngram_processed.**

Each unprocessed comment undergoes normalization, including the **removal of URLs**, **special characters**, and **line breaks**. The text is tokenized into words, and **stop words are removed**.

Using a sliding-window method, all possible n-grams up to a specified length (configured to five words) are generated. This produces sequences of one to five consecutive words for each comment.

Generated n-grams are aggregated in memory for the current batch. Duplicate entries, sharing the same text, date, category, label, and data source—have their frequency counts summed before persistence.

[4] https://opensearch.org/
[5] https://kafka.apache.org/

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

The aggregated results are upserted into **chase_comments_ngram**, either by creating new records or incrementing existing frequency counts. The processed comment IDs and their dates are then recorded in **chase_comments_ngram_processed**.

The process continues iteratively across all pages within the selected date range and then pauses until the next scheduled run.

This modular design ensures scalability and allows the visualization component to access continuously updated n-gram data for dynamic analytical displays.

Asynchronous Text Analytics Execution

Deliverable **D3.5 – CHASE Real-Time Text Analyser** presented the implementation of the text analytics component responsible for detecting and classifying gender-based hate speech within the CHASE ICT tool. It described two main analytical modules: a **binary classification model** for identifying hateful content and an **agentic workflow** for fine-grained detection of hate-speech patterns. In the final version of the platform, this component has been further optimized to improve responsiveness and scalability.

To achieve this, the execution of the two models has been decoupled and made **asynchronous**, allowing the faster binary classifier to operate independently from the more computationally intensive agentic analysis. This enhancement ensures that classification results are available in real time while the deeper pattern analysis continues in the background.

The updated workflow, described below, preserves the data integrity and message-driven structure introduced in *D3.3* while significantly improving throughput and user experience.

The updated workflow operates as follows:

1. The **Text Analytics component** consumes comments from the **chase_comments** Kafka topic.
2. It first executes the **binary classification model**, identifying whether a comment contains hate speech.
3. The classification results are written to the **chase_analytics** Kafka topic.
4. Through the **Storage Connector**, these results are persisted in the **chase_analytics** OpenSearch index.
5. A secondary service, **hate_speech_patterns_analytics**, independently consumes messages from the **chase_analytics** topic. It performs the **agentic workflow** to detect fine-grained hate-speech patterns and then invokes the updates the corresponding entries in the chase_analytics index.

This asynchronous architecture ensures that the slower agentic workflow does not block the real-time ingestion and classification pipeline. The separation of the two models enhances throughput, reduces latency for alert generation, and provides continuous updates to the analytics index as deeper analysis becomes available.

Integration of LLM Tracing

To enhance transparency, debugging, and performance evaluation of the text analytics component, **LLM tracing has been integrated into the CHASE ICT tool using Langfuse**[6]. This addition provides detailed observability of the Large Language Model (LLM) operations involved in the agentic workflow responsible for hate-speech pattern detection.

Langfuse is an open-source observability and analytics platform specifically designed for applications powered by large language models. It enables detailed tracing of each model interaction, capturing structured metadata such as prompt text, completion output, latency, cost metrics, and hierarchical call relationships between different stages of an agentic workflow. By aggregating these traces in real time, Langfuse provides developers and researchers with a unified interface for analyzing performance, identifying prompt bottlenecks, and validating model behavior across large volumes of executions.

[6] https://langfuse.com/

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

In the CHASE deployment, Langfuse is connected to the **hate-speech pattern detection agentic workflow** within the text analytics service. Each invocation of the workflow (executed by the *chase_agent* module) is automatically logged as a trace. A trace corresponds to one full analytical run of the agentic pipeline for a given post or comment and includes multiple sub-spans representing the LLM calls involved in each reasoning step (e.g., text summarization, pattern classification, validation). For each sub-span, Langfuse records the duration, token count, and status of the operation. This fine-grained logging enables correlation between input characteristics, model performance, and output reliability.

Figure 3 shows the Langfuse *Tracing Dashboard*, which provides an aggregated view of recent executions. Each row in the table represents a single trace recorded for the ***chase_agent*** **workflow**. Alongside **timestamps**, **user identifiers**, and organizational tags, the dashboard displays **quantitative performance indicators** including **latency** (execution time per run), **token usage** (input/output and total), and the **estimated cost** of each operation. This interface allows the technical team to monitor throughput and performance trends over time.

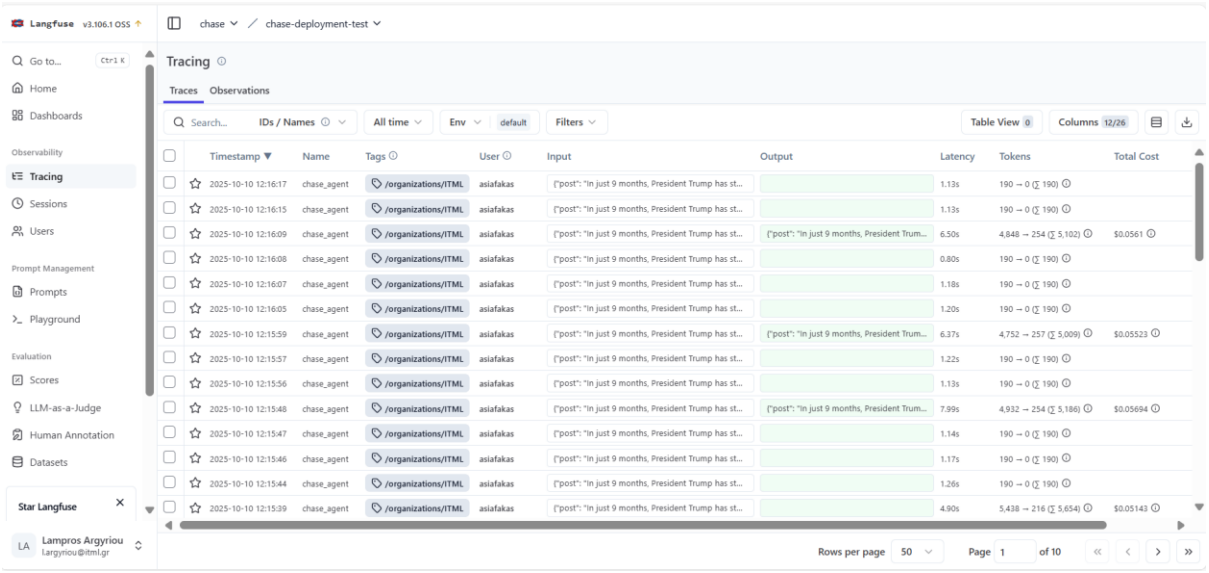

Figure 3:Langfuse Tracing Dashboard displaying recent executions of the CHASE agentic workflow.

Figure 4 presents a detailed view of one specific trace. The left panel visualizes the hierarchical execution graph, where each node represents a distinct LLM call corresponding to steps in the agentic process (e.g., summarize_post, classify_category, evaluate_classification). The central panel lists detailed metrics for each call, including latency and token count, while the right panel displays the structured input and output of the model. The bottom of the view includes a process flow diagram illustrating the logical progression from initialization through intermediate reasoning steps to completion.

Through this integration, the CHASE ICT tool now supports comprehensive monitoring of LLM-based processes. Developers can review historical executions, measure performance trends, and correlate specific prompts with analytical outcomes. This functionality strengthens the reliability and maintainability of the text analytics component by ensuring that all model-driven operations are transparent, measurable, and continuously verifiable during both development and pilot operation phases.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

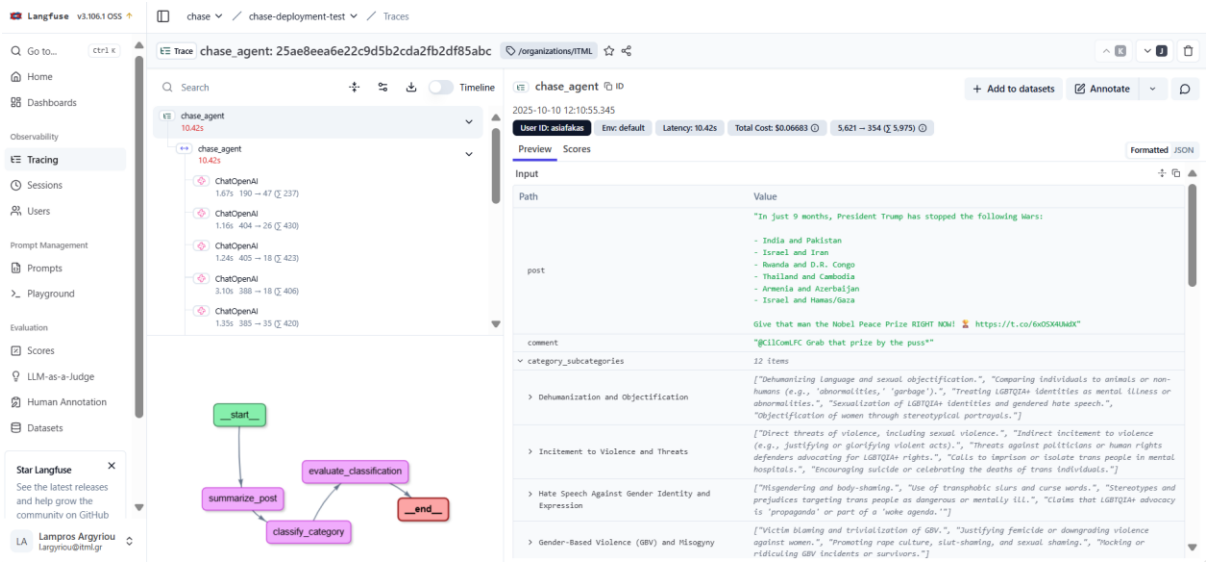

Figure 4: Detailed view of a Langfuse trace representing one execution of the CHASE agentic workflow. The visualization highlights the model's internal reasoning path and corresponding analytical outputs.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Conclusion

The completion of this deliverable marks the finalization of the CHASE ICT tool as a comprehensive and fully integrated platform. Through successive development stages, each component has been refined, connected, and validated to ensure cohesive operation and alignment with the project's objectives.

The result is a stable, functional solution that combines automation, interpretability, and accessibility within a unified framework.

This milestone concludes the technical development activities of WP3 and sets the stage for the evaluation and validation of the CHASE ICT tool during the pilot phase. Feedback from real-world use will guide any final refinements and support the long-term sustainability of the platform beyond the project's lifetime.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

CH%#
!#ASE

Combatting online
HAte Speech by
engaging online mEdia

D 3.8
Version finale de l'outil informatique CHASE

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Table des matières

Résumé exécutif 3
Introduction 3
Objectif du document 3
Structure du document 3
Déploiement et gestion des utilisateurs 4
Vue d'ensemble de l'environnement de la plateforme CHASE déployée 4
Processus d'authentification et de gestion des utilisateurs 4
Mises à jour du backend et de la couche de données 7
Flux de travail de calcul de n-grammes 7
Exécution asynchrone de l'analytique de texte 7
Intégration du traçage LLM 9
Conclusion 11

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Résumé exécutif

Ce document présente la version finale de l'outil informatique CHASE, développé dans le cadre du *Work Package 3 (WP3) -- Composants technologiques et plateforme CHASE*.

S'appuyant sur les spécifications techniques et les implémentations incrémentales décrites dans les livrables antérieurs (D3.1, D3.3, D3.5 et D3.6), ce rapport consolide la version complète et déployée du système CHASE et résume les derniers perfectionnements techniques conduisant à sa pleine disponibilité opérationnelle.

Avec ces perfectionnements, l'outil informatique CHASE atteint sa pleine maturité fonctionnelle, prêt pour un déploiement pilote et une validation dans des conditions réelles par les organisations partenaires. Il représente la consolidation de toutes les réalisations technologiques du WP3 et établit une base robuste pour une utilisation opérationnelle future et la durabilité.

Introduction

Objectif du document

Ce livrable présente la version intégrée et opérationnelle finale de l'outil informatique CHASE. Il décrit le système dans sa forme achevée, en résumant les dernières mises à jour qui ont été implémentées.

Le document met en évidence trois aspects principaux :

- L'intégration des composants de la plateforme dans un environnement opérationnel unifié prenant en charge l'accès et l'interaction multi-partenaires.
- La configuration finalisée et la gestion des accès utilisateurs.
- Les dernières améliorations apportées aux flux de travail de données et d'analytique, développées pour améliorer la visualisation et la réactivité analytique.

Dans l'ensemble, le document constitue le rapport technique final du WP3, démontrant que l'outil informatique CHASE a atteint le niveau d'intégration et de préparation prévu pour la validation dans la phase pilote.

Structure du document

Le reste de ce document est structuré comme suit :

- Section 2 décrit l'approche utilisée pour gérer l'accès utilisateur et l'authentification entre les organisations partenaires.
- Section 3 décrit les dernières améliorations apportées aux flux de travail internes du système.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Déploiement et gestion des utilisateurs

Vue d'ensemble de l'environnement de la plateforme CHASE déployée

La version finale de l'outil informatique CHASE a été entièrement déployée sur un cluster Kubernetes (K8s)[1] dédié hébergé sur l'infrastructure d'ITML. Chaque service a été conteneurisé à l'aide de Docker[2]. Cette configuration fournit un environnement d'exécution sécurisé, contrôlé et évolutif qui permet un contrôle opérationnel précis. Tous les composants du système sont conteneurisés et gérés via Kubernetes.

Le cluster est configuré pour orchestrer le déploiement, la mise à l'échelle et la gestion du cycle de vie des services via des manifestes déclaratifs. Chaque composant est déployé comme un pod Kubernetes distinct, regroupé en espaces de noms correspondant aux principaux sous-systèmes de l'outil informatique CHASE. Les services avec état, tels qu'OpenSearch et Keycloak, utilisent des Volumes Persistants (PV) pour garantir la durabilité des données lors des redémarrages et des mises à niveau. Les contrôleurs Ingress n'exposent que les points de terminaison publics nécessaires (par exemple, l'interface web CHASE et le portail d'authentification Keycloak) via HTTPS, tandis que la communication interne des services reste limitée au réseau privé du cluster.

La communication inter-services entre les composants CHASE repose sur une combinaison d'API REST et de streaming d'événements asynchrone via Apache Kafka. Pour sécuriser cette couche de communication, le TLS mutuel (mTLS)[3] a été implémenté entre les producteurs et consommateurs Kafka, garantissant à la fois la confidentialité des messages et l'authentification bidirectionnelle. Cette approche garantit que seuls les services approuvés au sein du cluster Kubernetes peuvent publier ou consommer des événements, renforçant ainsi la posture de sécurité globale de la plateforme.

La surveillance continue de l'environnement déployé est réalisée à l'aide de sondes de santé Kubernetes, de mécanismes de journalisation internes et de tableaux de bord OpenSearch. Cette configuration permet l'observabilité en temps réel, la détection des pannes et une maintenance simplifiée.

Dans sa forme déployée finale, l'outil informatique CHASE fonctionne comme un système cohérent et prêt pour la production qui intègre tous les composants principaux dans un environnement unifié, sécurisé et maintenable, prenant en charge le flux de travail complet de modération et d'analytique défini dans le Work Package 3.

Processus d'authentification et de gestion des utilisateurs

L'implémentation de l'authentification utilisateur et du contrôle d'accès dans la version finale de l'outil informatique CHASE suit la configuration précédemment décrite dans le *Livrable D3.1 -- Spécifications*

[1] https://kubernetes.io/
[2] https://www.docker.com/
[3] https://en.wikipedia.org/wiki/Transport_Layer_Security

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

techniques, où l'intégration de **Keycloak** comme solution de **gestion des identités et des accès (IAM)** a été introduite pour la première fois. Dans cette version finale, Keycloak a été entièrement intégré dans l'environnement déployé et sert de point d'entrée central pour l'authentification et l'autorisation des utilisateurs au sein de la plateforme CHASE.

La gestion des utilisateurs est effectuée directement via la **console d'administration Keycloak**, qui permet aux administrateurs système de créer, mettre à jour et désactiver des comptes utilisateurs ainsi que de leur attribuer des rôles prédéfinis.

Keycloak fonctionne sous un domaine dédié nommé **"chase"**, qui définit tous les utilisateurs actifs, clients et politiques associés à l'outil. Deux clients d'application principaux ont été configurés :

- **chase-client** --- représentant l'application web CHASE, utilisé pour l'authentification côté utilisateur via OpenID Connect.
- **chase-system-client** --- représentant la communication système interne entre les services backend et la couche d'authentification.

Le processus d'inscription des utilisateurs pour la plateforme CHASE a été coordonné de manière centralisée par l'équipe technique d'ITML. Chaque organisation participante a été invitée à fournir les détails suivants pour chaque personne nécessitant un accès à l'outil :

- Nom
- Nom d'utilisateur
- Mot de passe
- Nom de l'organisation

Après avoir collecté ces informations, les administrateurs système ont créé un **groupe** correspondant dans Keycloak pour chaque organisation représentée dans le projet (par exemple, ITML, Proto Thema, MDIG, CESIE, WAN-IFRA, etc.). Chaque compte utilisateur a ensuite été créé manuellement dans le domaine *chase* et attribué au groupe organisationnel respectif. Cette structure basée sur les groupes garantit que les comptes utilisateurs sont clairement associés à leur organisation et simplifie la gestion de plusieurs partenaires au sein d'un même environnement.

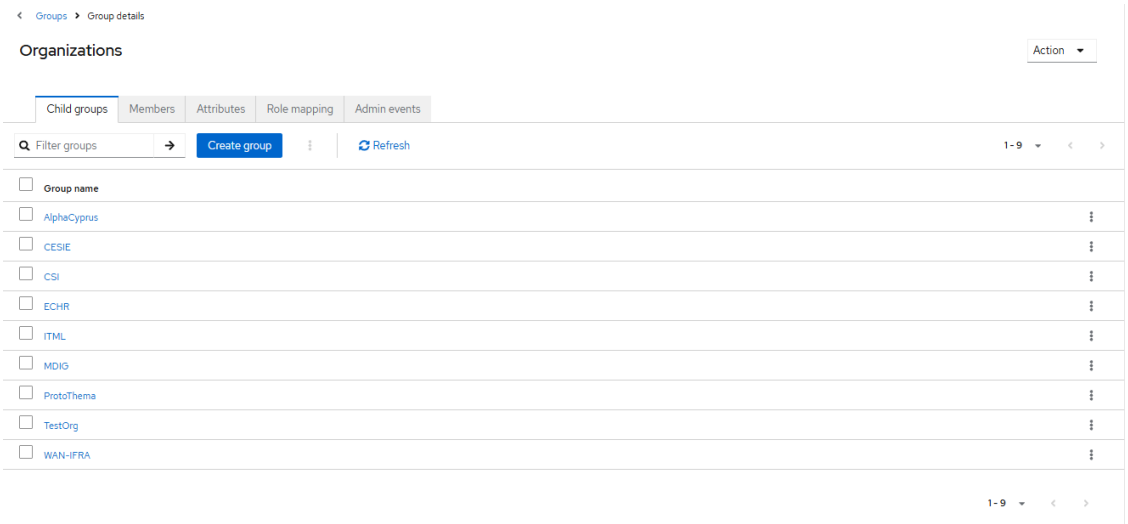

Figure 1 : Groupes d'organisations partenaires définis dans le realm CHASE de Keycloak

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

L'inscription des utilisateurs n'est intentionnellement **pas ouverte à l'auto-inscription**. Tous les comptes sont créés et activés par les administrateurs afin de préserver le contrôle de l'accès au système et de maintenir un lien vérifiable entre les utilisateurs et leurs organisations. Les fonctionnalités de réinitialisation du mot de passe et de vérification par e-mail sont désactivées, car la gestion des identifiants est gérée exclusivement via des flux de travail administratifs.

Cette configuration garantit que l'ensemble du processus de gestion des utilisateurs reste transparent, traçable et cohérent dans tout le consortium. Toutes les demandes d'authentification sont traitées via l'interface de connexion sécurisée de Keycloak, qui s'intègre de manière transparente avec les services frontend et backend de CHASE. Le résultat est un flux de travail d'accès contrôlé et auditable qui prend en charge le fonctionnement fiable de l'outil informatique CHASE dans un contexte multi-organisationnel.

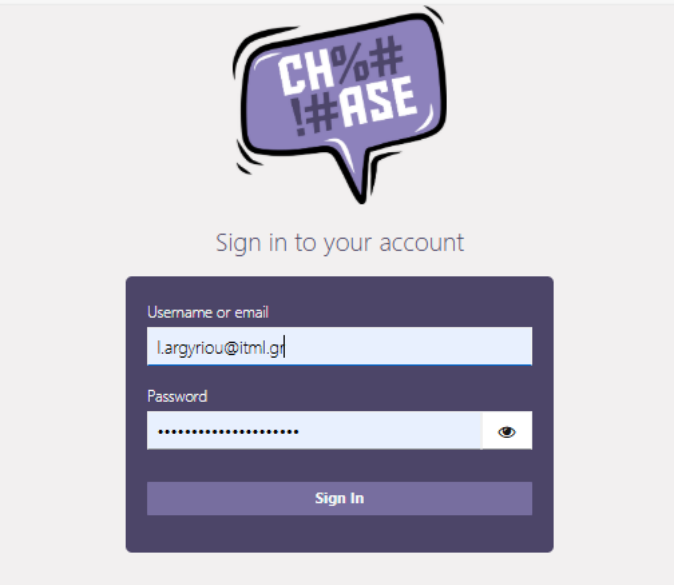

Figure 2 : Page de connexion CHASE personnalisée intégrée avec Keycloak, qui fournit une authentification sécurisée pour les utilisateurs enregistrés dans le realm CHASE

L'outil informatique CHASE est accessible aux utilisateurs autorisés via une interface web sécurisée hébergée sous le domaine dédié du projet. L'accès est fourni via l'adresse suivante :

https://chase-web.itml.space/

Ce point d'entrée sert de passerelle unifiée vers l'environnement CHASE, où les organisations partenaires peuvent se connecter en utilisant les identifiants qui leur sont attribués. Le domaine prend en charge la communication cryptée (HTTPS) et se connecte directement au service d'authentification, garantissant un accès sécurisé à toutes les fonctionnalités de la plateforme CHASE.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Mises à jour du backend et de la couche de données

Le livrable **D3.3 -- CHASE Data Crawler & Data Repository** a documenté l'implémentation de l'infrastructure de données de base de l'outil informatique CHASE, y compris l'intégration du backend basé sur contrôleur (API CHASE), l'adoption d'**OpenSearch**[4] pour le stockage des données et l'utilisation d'**Apache Kafka**[5] pour la messagerie en temps réel. Le livrable a défini les principaux flux de données prenant en charge l'ingestion, l'analyse et la visualisation du contenu via les indices **chase_posts**, **chase_comments** et **chase_analytics**, ainsi que leurs sujets Kafka correspondants. Ces composants ont établi la base du traitement de données de bout en bout, de la récupération de contenu externe à l'analytique de texte et à la visualisation.

S'appuyant sur cette architecture, la version finale de l'outil informatique CHASE introduit des améliorations ciblées pour étendre les capacités analytiques et améliorer les performances du système. Deux développements clés ont été implémentés :

1. un nouveau **flux de travail de calcul de n-grammes** avec des indices dédiés pour prendre en charge la visualisation avancée
2. un **mécanisme d'exécution asynchrone** pour les services d'analytique de texte, permettant une inférence plus rapide et plus efficace.

Les sous-sections suivantes décrivent ces mises à jour en détail.

Flux de travail de calcul de n-grammes

Pour activer la **fonctionnalité de visualisation de n-grammes** introduite dans le composant de visualisation (*Livrable D3.6 -- Composant de visualisation CHASE*), un nouveau flux de travail de traitement des données a été développé et intégré dans le backend de l'outil informatique CHASE. Ce flux de travail fonctionne comme une tâche planifiée qui extrait, traite et agrège périodiquement les données textuelles de l'indice **chase_comments** pour générer des statistiques de n-grammes.

Deux nouveaux indices OpenSearch ont été ajoutés pour prendre en charge ce processus :

- **chase_comments_ngram** -- stockant les données de n-grammes agrégées avec des comptes de fréquence.
- **chase_comments_ngram_processed** -- suivant l'état de traitement des commentaires pour éviter la duplication.

Le flux de données pour le calcul des n-grammes suit un pipeline structuré et automatisé. La tâche s'exécute à intervalles prédéfinis (10 minutes). Chaque exécution détermine la fenêtre temporelle à traiter en vérifiant l'entrée la plus récente dans **chase_comments_ngram_processed**, ou utilise par défaut le commentaire disponible le plus ancien s'il s'agit de la première exécution.

[4] https://opensearch.org/
[5] https://kafka.apache.org/

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Les commentaires dans la plage de dates sélectionnée sont récupérés par lots de dix. Pour chaque lot, les commentaires précédemment traités sont filtrés en fonction de leurs identifiants enregistrés dans chase_comments_ngram_processed.

Chaque commentaire non traité subit une normalisation, y compris la suppression des URL, des caractères spéciaux et des sauts de ligne. Le texte est tokenisé en mots, et les mots vides sont supprimés.

En utilisant une méthode de fenêtre glissante, tous les n-grammes possibles jusqu'à une longueur spécifiée (configurée à cinq mots) sont générés. Cela produit des séquences d'un à cinq mots consécutifs pour chaque commentaire.

Les n-grammes générés sont agrégés en mémoire pour le lot actuel. Les entrées en double, partageant le même texte, la même date, la même catégorie, la même étiquette et la même source de données, voient leurs comptes de fréquence additionnés avant la persistance.

Les résultats agrégés sont insérés ou mis à jour dans chase_comments_ngram, soit en créant de nouveaux enregistrements, soit en incrémentant les comptes de fréquence existants. Les identifiants des commentaires traités et leurs dates sont ensuite enregistrés dans chase_comments_ngram_processed.

Le processus se poursuit de manière itérative sur toutes les pages dans la plage de dates sélectionnée, puis fait une pause jusqu'à la prochaine exécution planifiée.

Cette conception modulaire garantit l'évolutivité et permet au composant de visualisation d'accéder à des données de n-grammes continuellement mises à jour pour des affichages analytiques dynamiques.

Exécution asynchrone de l'analytique de texte

Le livrable D3.5 -- CHASE Real-Time Text Analyser a présenté l'implémentation du composant d'analytique de texte responsable de la détection et de la classification du discours de haine de genre au sein de l'outil informatique CHASE. Il a décrit deux modules analytiques principaux : un modèle de classification binaire pour identifier le contenu haineux et un flux de travail agentique pour la détection précise des modèles de discours de haine. Dans la version finale de la plateforme, ce composant a été davantage optimisé pour améliorer la réactivité et l'évolutivité.

Pour y parvenir, l'exécution des deux modèles a été découplée et rendue asynchrone, permettant au classificateur binaire plus rapide de fonctionner indépendamment de l'analyse agentique plus intensive en calcul. Cette amélioration garantit que les résultats de classification sont disponibles en temps réel tandis que l'analyse de modèles plus approfondie se poursuit en arrière-plan.

Le flux de travail mis à jour, décrit ci-dessous, préserve l'intégrité des données et la structure pilotée par les messages introduite dans *D3.3* tout en améliorant considérablement le débit et l'expérience utilisateur.

Le flux de travail mis à jour fonctionne comme suit :

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

1. Le **composant d'analytique de texte** consomme les commentaires du sujet Kafka **chase_comments**.
2. Il exécute d'abord le **modèle de classification binaire**, identifiant si un commentaire contient un discours de haine.
3. Les résultats de classification sont écrits dans le sujet Kafka **chase_analytics**.
4. Via le **Storage Connector**, ces résultats sont persistés dans l'indice OpenSearch **chase_analytics**.
5. Un service secondaire, **hate_speech_patterns_analytics**, consomme indépendamment les messages du sujet **chase_analytics**. Il effectue le **flux de travail agentique** pour détecter les modèles de discours de haine précis, puis invoque les mises à jour des entrées correspondantes dans l'indice chase_analytics.

Cette architecture asynchrone garantit que le flux de travail agentique plus lent ne bloque pas le pipeline d'ingestion et de classification en temps réel. La séparation des deux modèles améliore le débit, réduit la latence pour la génération d'alertes et fournit des mises à jour continues à l'indice analytique au fur et à mesure que des analyses plus approfondies deviennent disponibles.

Intégration du traçage LLM

Pour améliorer la transparence, le débogage et l'évaluation des performances du composant d'analytique de texte, **le traçage LLM a été intégré dans l'outil informatique CHASE en utilisant Langfuse**[6]. Cet ajout fournit une observabilité détaillée des opérations des grands modèles de langage (LLM) impliquées dans le flux de travail agentique responsable de la détection des modèles de discours de haine.

Langfuse est une plateforme d'observabilité et d'analytique open source spécialement conçue pour les applications alimentées par de grands modèles de langage. Elle permet un traçage détaillé de chaque interaction de modèle, capturant des métadonnées structurées telles que le texte de prompt, la sortie de complétion, la latence, les métriques de coût et les relations d'appel hiérarchiques entre différentes étapes d'un flux de travail agentique. En agrégeant ces traces en temps réel, Langfuse fournit aux développeurs et chercheurs une interface unifiée pour analyser les performances, identifier les goulots d'étranglement de prompt et valider le comportement du modèle sur de gros volumes d'exécutions.

Dans le déploiement CHASE, Langfuse est connecté au **flux de travail agentique de détection des modèles de discours de haine** au sein du service d'analytique de texte. Chaque invocation du flux de travail (exécutée par le module *chase_agent*) est automatiquement enregistrée comme une trace. Une trace correspond à une exécution analytique complète du pipeline agentique pour une publication ou un commentaire donné et inclut plusieurs sous-spans représentant les appels LLM impliqués dans chaque étape de raisonnement (par exemple, résumé de texte, classification de modèle, validation). Pour chaque sous-span, Langfuse enregistre la durée, le nombre de tokens et l'état de l'opération. Cette

[6] https://langfuse.com/

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

journalisation précise permet la corrélation entre les caractéristiques d'entrée, les performances du modèle et la fiabilité de la sortie.

La **Figure 3** montre le *Tableau de bord de traçage* de Langfuse, qui fournit une vue agrégée des exécutions récentes. Chaque ligne du tableau représente une seule trace enregistrée pour le **flux de travail *chase_agent***. Outre les **horodatages**, les **identifiants utilisateur** et les balises organisationnelles, le tableau de bord affiche des **indicateurs de performance quantitatifs** incluant la **latence** (temps d'exécution par exécution), l'**utilisation des tokens** (entrée/sortie et total) et le **coût estimé** de chaque opération. Cette interface permet à l'équipe technique de surveiller le débit et les tendances de performance au fil du temps.

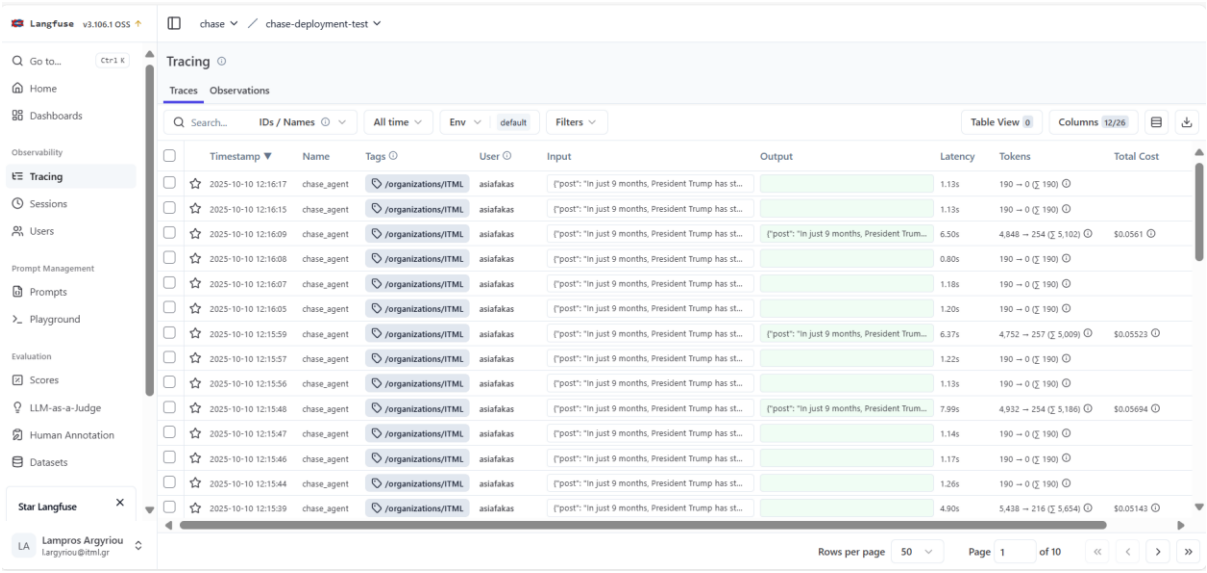

Figure 3 : Tableau de bord de suivi Langfuse montrant les exécutions récentes du workflow agentique CHASE

La **Figure 4** présente une vue détaillée d'une trace spécifique. Le panneau de gauche visualise le graphe d'exécution hiérarchique, où chaque nœud représente un appel LLM distinct correspondant aux étapes du processus agentique (par exemple, summarize_post, classify_category, evaluate_classification). Le panneau central liste les métriques détaillées pour chaque appel, y compris la latence et le nombre de tokens, tandis que le panneau de droite affiche l'entrée et la sortie structurées du modèle. Le bas de la vue comprend un diagramme de flux de processus illustrant la progression logique de l'initialisation à travers les étapes de raisonnement intermédiaires jusqu'à l'achèvement.

Grâce à cette intégration, l'outil informatique CHASE prend désormais en charge une surveillance complète des processus basés sur LLM. Les développeurs peuvent examiner les exécutions historiques, mesurer les tendances de performance et corréler des prompts spécifiques avec des résultats analytiques. Cette fonctionnalité renforce la fiabilité et la maintenabilité du composant d'analytique de texte en garantissant que toutes les opérations pilotées par modèle sont transparentes, mesurables et continuellement vérifiables pendant les phases de développement et d'opération pilote.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

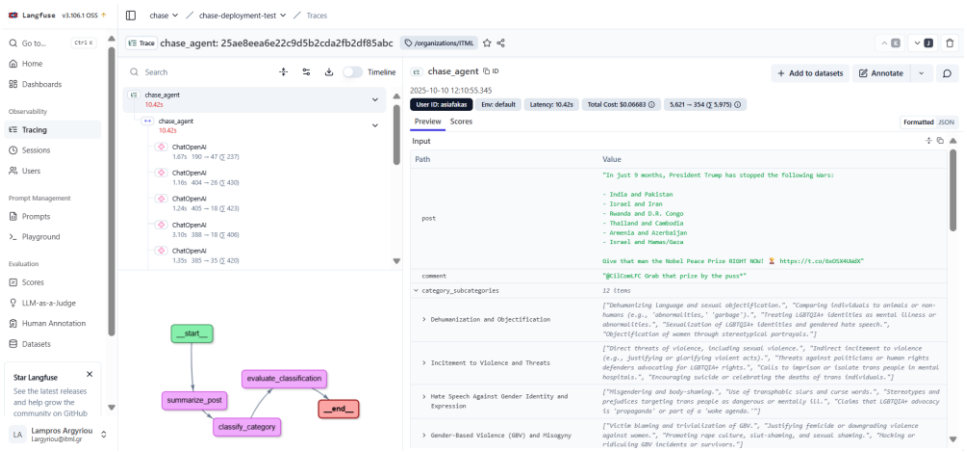

Figure 4 : Visualisation détaillée d'une trace Langfuse représentant une exécution du workflow agentique CHASE. La visualisation met en évidence le cheminement de raisonnement interne du modèle et les sorties analytiques correspondantes

Conclusion

L'achèvement de ce livrable marque la **finalisation de l'outil informatique CHASE** en tant que plateforme complète et entièrement intégrée. À travers des étapes de développement successives, chaque composant a été affiné, connecté et validé pour assurer un fonctionnement cohérent et un alignement avec les objectifs du projet.

Le résultat est une solution stable et fonctionnelle qui combine automatisation, interprétabilité et accessibilité dans un cadre unifié.

Cette étape conclut les activités de développement technique du WP3 et prépare le terrain pour l'évaluation et la validation de l'outil informatique CHASE pendant la phase pilote. Les retours d'expérience de l'utilisation en conditions réelles guideront les derniers perfectionnements et soutiendront la durabilité à long terme de la plateforme au-delà de la durée de vie du projet.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Combatting online HAte Speech by engaging online mEdia

D3.8
Τελική Έκδοση του Πληροφοριακού Εργαλείου CHASE

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Περιεχόμενα

D3.8 Final Version of CHASE ICT Tool 1
Περίληψη 4
Εισαγωγή 5
Στόχος του Παραδοτέου 5
Δομή του παρόντος Εγγράφου 5
Επισκόπηση του αναπτυγμένου περιβάλλοντος της πλατφόρμας CHASE 6
Διαδικασία Αυθεντικοποίησης και Διαχείρισης Χρηστών 7
Ενημερώσεις στο Backend και στο Επίπεδο Δεδομένων 10
Ροή Υπολογισμού N-gram 10
Ασύγχρονη Εκτέλεση Ανάλυσης Κειμένου 11
Ενσωμάτωση Ανίχνευσης LLM (LLM Tracing) 12
Κατακλείδα 15

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Περίληψη

Το παρόν έγγραφο παρουσιάζει την τελική έκδοση του Εργαλείου Πληροφορικής CHASE (CHASE ICT Tool), το οποίο αναπτύχθηκε στο πλαίσιο του Πακέτου Εργασίας 3 (WP3) – Τεχνολογικά Στοιχεία και Πλατφόρμα CHASE.

Βασισμένο στις τεχνικές προδιαγραφές και τις σταδιακές υλοποιήσεις που περιγράφηκαν σε προηγούμενα παραδοτέα (D3.1, D3.3, D3.5 και D3.6), η παρούσα αναφορά συγκεντρώνει την πλήρη και αναπτυγμένη εκδοχή του συστήματος CHASE και συνοψίζει τις πλέον πρόσφατες τεχνικές βελτιώσεις που οδήγησαν στην πλήρη επιχειρησιακή του ετοιμότητα.

Με τις συγκεκριμένες βελτιστοποιήσεις, το εργαλείο CHASE ICT φτάνει στο πλήρες λειτουργικό του δυναμικό, έτοιμο για πιλοτική εφαρμογή και επιβεβαίωση σε πραγματικές συνθήκες από τους εταίρους του έργου. Αντιπροσωπεύει τη συγκέντρωση όλων των τεχνολογικών επιτευγμάτων του WP3 και θέτει ένα ισχυρό θεμέλιο για τη μελλοντική επιχειρησιακή αξιοποίηση και βιωσιμότητά του.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Εισαγωγή

Στόχος του Παραδοτέου

Το παρόν παραδοτέο παρουσιάζει την τελική, ολοκληρωμένη και λειτουργική έκδοση του εργαλείου πληροφορικής CHASE (CHASE ICT tool). Περιγράφει το σύστημα στην τελική του μορφή και συνοψίζει τις τελευταίες ενημερώσεις και βελτιώσεις που έχουν υλοποιηθεί.

Το έγγραφο επικεντρώνεται σε τρεις βασικές πτυχές:

- την ολοκλήρωση και ενοποίηση των επιμέρους στοιχείων της πλατφόρμας σε ένα ενιαίο λειτουργικό περιβάλλον που υποστηρίζει πρόσβαση και αλληλεπίδραση πολλαπλών εταίρων,
- την τελική διαμόρφωση και διαχείριση της πρόσβασης των χρηστών,
- και τις πιο πρόσφατες βελτιώσεις στις ροές δεδομένων και ανάλυσης, που έχουν αναπτυχθεί με στόχο τη βελτίωση της οπτικοποίησης και της αναλυτικής ανταπόκρισης.
-

Συνολικά, το παρόν έγγραφο λειτουργεί ως η τελική τεχνική αναφορά του Πακέτου Εργασίας 3 (WP3), αποδεικνύοντας ότι το εργαλείο CHASE ICT έχει επιτύχει το επιθυμητό επίπεδο ολοκλήρωσης και ετοιμότητας για επικύρωση κατά τη φάση του πιλοτικού προγράμματος.

Δομή του παρόντος Εγγράφου

Η υπόλοιπη δομή του εγγράφου έχει ως εξής:

- Η Ενότητα 2 περιγράφει την προσέγγιση που χρησιμοποιήθηκε για τη διαχείριση της πρόσβασης και της ταυτοποίησης των χρηστών μεταξύ των οργανισμών-εταίρων.

- Η Ενότητα 3 παρουσιάζει τις πιο πρόσφατες βελτιώσεις στις εσωτερικές ροές λειτουργίας του συστήματος.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Επισκόπηση του αναπτυγμένου περιβάλλοντος της πλατφόρμας CHASE

Η τελική έκδοση του εργαλείου CHASE ICT έχει αναπτυχθεί πλήρως σε ένα αποκλειστικό cluster ***Kubernetes (K8s)***[1], το οποίο φιλοξενείται στις υποδομές της ITML (Information Technology for Market Leadership). Κάθε υπηρεσία έχει αναπτυχθεί ως *container (containerized)* χρησιμοποιώντας *Docker*[2].

Η συγκεκριμένη αρχιτεκτονική παρέχει ένα ασφαλές, ελεγχόμενο και επεκτάσιμο περιβάλλον εκτέλεσης, το οποίο επιτρέπει λεπτομερή λειτουργικό έλεγχο. Όλα τα συστατικά του συστήματος έχουν αναπτυχθεί ως *containers* και διαχειρίζονται μέσω ***Kubernetes***.

Το cluster έχει ρυθμιστεί ώστε να ενορχηστρώνει την ανάπτυξη, την κλιμάκωση και τη διαχείριση του κύκλου ζωής των υπηρεσιών μέσω δηλωτικών (*manifests)*. Κάθε συνιστώσα αναπτύσσεται ως ξεχωριστό ***Kubernetes pod***, το οποίο ομαδοποιείται σε ***namespaces*** που αντιστοιχούν στα κύρια υποσυστήματα του εργαλείου CHASE.

Οι υπηρεσίες που διατηρούν κατάσταση (*stateful services*), όπως το ***OpenSearch*** και το ***Keycloak***, χρησιμοποιούν ***Persistent Volumes (PVs)*** για να διασφαλίζουν τη διατήρηση των δεδομένων κατά τη διάρκεια επανεκκινήσεων ή αναβαθμίσεων.

Οι ***Ingress controllers*** εκθέτουν μόνο τα απαραίτητα δημόσια σημεία πρόσβασης (όπως το *CHASE web interface* και το ***Keycloak authentication portal***) μέσω *HTTPS*, ενώ η εσωτερική επικοινωνία μεταξύ υπηρεσιών παραμένει περιορισμένη στο ιδιωτικό δίκτυο του cluster.

Η επικοινωνία μεταξύ των επιμέρους στοιχείων του CHASE βασίζεται σε συνδυασμό ***REST APIs*** και ασύγχρονης ροής γεγονότων (*event streaming*) μέσω ***Apache Kafka***. Για την ασφάλεια αυτού του επιπέδου επικοινωνίας έχει υλοποιηθεί ***mutual TLS (mTLS)***[3] μεταξύ των *Kafka producers* και *consumers*, εξασφαλίζοντας τόσο την εμπιστευτικότητα των μηνυμάτων όσο και τον αμοιβαίο έλεγχο ταυτότητας. Με αυτόν τον τρόπο, μόνο αξιόπιστες υπηρεσίες εντός του *Kubernetes cluster* μπορούν να δημοσιεύουν ή να καταναλώνουν γεγονότα, ενισχύοντας τη συνολική ασφάλεια της πλατφόρμας.

Η συνεχής παρακολούθηση του αναπτυγμένου περιβάλλοντος επιτυγχάνεται μέσω *Kubernetes health probes*, εσωτερικών μηχανισμών καταγραφής (*logging mechanisms*) και *OpenSearch dashboards*. Αυτή η διαμόρφωση παρέχει σε πραγματικό χρόνο ορατότητα, έγκαιρη ανίχνευση σφαλμάτων και απλοποιημένη διαδικασία συντήρησης.

Στην τελική, παραγωγικά έτοιμη μορφή του, το εργαλείο CHASE ICT λειτουργεί ως ένα συνεκτικό και ολοκληρωμένο σύστημα, το οποίο ενσωματώνει όλα τα βασικά του στοιχεία σε ένα ενιαίο, ασφαλές και συντηρήσιμο περιβάλλον, υποστηρίζοντας πλήρως τη ροή εποπτείας και ανάλυσης που ορίστηκε στο Πακέτο Εργασίας 3.

[1] https://kubernetes.io/
[2] https://www.docker.com/
[3] https://en.wikipedia.org/wiki/Transport_Layer_Security

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Διαδικασία Αυθεντικοποίησης και Διαχείρισης Χρηστών

Η υλοποίηση της διαδικασίας αυθεντικοποίησης χρηστών και ελέγχου πρόσβασης στην τελική έκδοση του εργαλείου CHASE ICT ακολουθεί τη ρύθμιση που είχε αρχικά περιγραφεί στο Παραδοτέο *D3.1 – Technical Specifications*, όπου παρουσιάστηκε για πρώτη φορά η ενσωμάτωση του *Keycloak* ως λύση *Identity and Access Management (IAM)*.

Στην τελική αυτή εκδοχή, το *Keycloak* έχει ενσωματωθεί πλήρως στο αναπτυγμένο περιβάλλον και λειτουργεί ως το κεντρικό σημείο εισόδου για την αυθεντικοποίηση και την εξουσιοδότηση των χρηστών εντός της πλατφόρμας CHASE.

Η διαχείριση των χρηστών πραγματοποιείται απευθείας μέσω της κονσόλας διαχείρισης του *Keycloak*, η οποία επιτρέπει στους διαχειριστές του συστήματος να δημιουργούν, ενημερώνουν και απενεργοποιούν λογαριασμούς χρηστών, καθώς και να τους αναθέτουν προκαθορισμένους ρόλους.

Το *Keycloak* λειτουργεί σε ένα ειδικά διαμορφωμένο *realm* με το όνομα **“chase”**, το οποίο περιλαμβάνει όλους τους ενεργούς χρήστες, τους πελάτες (*clients*) και τις πολιτικές που σχετίζονται με το εργαλείο. Έχουν ρυθμιστεί δύο βασικές εφαρμογές-πελάτες (*application clients*):

- chase-client — αντιπροσωπεύει την διαδικτυακή εφαρμογή του CHASE και χρησιμοποιείται για την αυθεντικοποίηση χρηστών μέσω *OpenID Connect*.
- chase-system-client — αντιπροσωπεύει την εσωτερική επικοινωνία του συστήματος μεταξύ των υπηρεσιών *backend* και του επιπέδου αυθεντικοποίησης.

Η διαδικασία εγγραφής χρηστών για την πλατφόρμα CHASE συντονίστηκε κεντρικά από την τεχνική ομάδα της ITML (Information Technology for Market Leadership). Κάθε οργανισμός-εταίρος κλήθηκε να παρέχει τα ακόλουθα στοιχεία για κάθε άτομο που χρειαζόταν πρόσβαση στο εργαλείο:

- Όνομα
- Όνομα χρήστη (*Username*)
- Κωδικό πρόσβασης (*Password*)
- Ονομασία οργανισμού

Αφού συλλέχθηκαν τα δεδομένα αυτά, οι διαχειριστές του συστήματος δημιούργησαν στο *Keycloak* μία ξεχωριστή ομάδα (*group*) για κάθε οργανισμό του έργου (π.χ. ITML, Proto Thema, MDIG, CESIE, WAN-IFRA κ.λπ.). Κάθε λογαριασμός χρήστη δημιουργήθηκε χειροκίνητα εντός του *chase realm* και ανατέθηκε στη σχετική οργανωτική ομάδα.

Η ομαδοποιημένη αυτή δομή εξασφαλίζει ότι οι λογαριασμοί χρηστών συνδέονται με σαφή τρόπο με τον οργανισμό τους και διευκολύνει τη διαχείριση πολλαπλών εταίρων σε ένα ενιαίο περιβάλλον.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

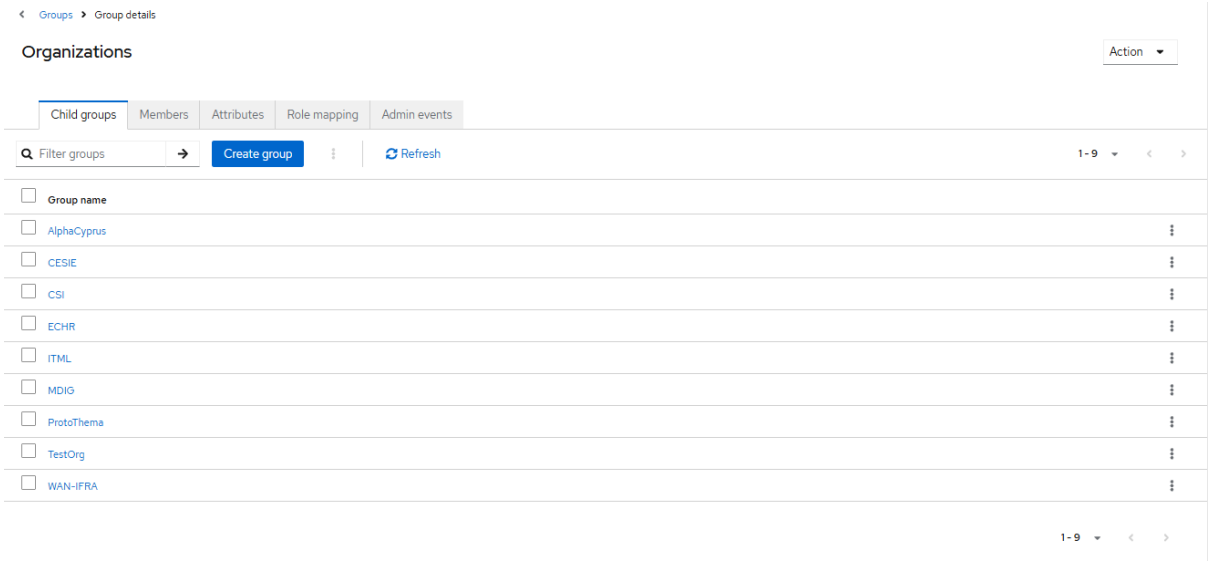

Figure 1: Ομάδες οργανισμών-εταίρων που ορίζονται στο CHASE Keycloak realm.

Η εγγραφή χρηστών δεν είναι ανοικτή σε αυτόματη αυτο-εγγραφή (*self-signup*). Όλοι οι λογαριασμοί δημιουργούνται και ενεργοποιούνται αποκλειστικά από τους διαχειριστές, προκειμένου να διατηρείται ο πλήρης έλεγχος της πρόσβασης στο σύστημα και να εξασφαλίζεται επαληθεύσιμη σύνδεση κάθε χρήστη με τον οργανισμό του.

Οι λειτουργίες επαναφοράς κωδικού πρόσβασης (*password reset*) και επιβεβαίωσης ηλεκτρονικής διεύθυνσης (*email verification*) είναι απενεργοποιημένες, καθώς η διαχείριση των διαπιστευτηρίων πραγματοποιείται αποκλειστικά μέσω διοικητικών διαδικασιών.

Η διαμόρφωση αυτή εξασφαλίζει ότι η συνολική διαδικασία διαχείρισης χρηστών παραμένει διαφανής, ιχνηλάσιμη και συνεπής σε όλο το εταιρικό σχήμα. Όλα τα αιτήματα αυθεντικοποίησης διεκπεραιώνονται μέσω της ασφαλούς διεπαφής σύνδεσης του *Keycloak*, η οποία έχει ενσωματωθεί πλήρως τόσο στο *frontend* όσο και στο *backend* του CHASE (Figure 2). Το αποτέλεσμα είναι μία ελεγχόμενη και καταγεγραμμένη ροή πρόσβασης που εξασφαλίζει αξιόπιστη λειτουργία του εργαλείου CHASE ICT σε περιβάλλον πολλαπλών οργανισμών.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

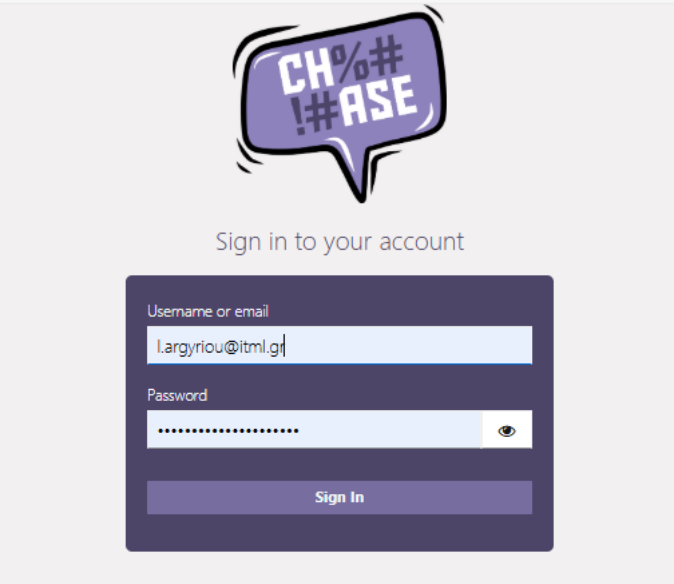

Figure 2: Προσαρμοσμένη σελίδα σύνδεσης του CHASE ενσωματωμένη με το Keycloak, η οποία παρέχει ασφαλή πιστοποίηση για εγγεγραμμένους χρήστες εντός του CHASE realm..

Το εργαλείο (CHASE ICT Tool) είναι προσβάσιμο σε εξουσιοδοτημένους χρήστες μέσω μιας ασφαλούς διαδικτυακής διεπαφής, η οποία φιλοξενείται στο αποκλειστικό domain για τις ανάγκες του έργου:

https://chase-web.itml.space/

Το συγκεκριμένο σημείο εισόδου λειτουργεί ως ενιαία πύλη προς το περιβάλλον CHASE, όπου οι εταίροι του έργου μπορούν να συνδεθούν χρησιμοποιώντας τα διαπιστευτήρια που τους έχουν ανατεθεί. Ο τομέας υποστηρίζει κρυπτογραφημένη επικοινωνία (*HTTPS*) και συνδέεται απευθείας με την υπηρεσία αυθεντικοποίησης, εξασφαλίζοντας ασφαλή πρόσβαση σε όλες τις λειτουργίες της πλατφόρμας CHASE.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Ενημερώσεις στο Backend και στο Επίπεδο Δεδομένων

Το **Παραδοτέο *D3.3 – CHASE Data Crawler & Data Repository*** τεκμηρίωσε την υλοποίηση της βασικής υποδομής δεδομένων του εργαλείου CHASE ICT, συμπεριλαμβανομένης της ενσωμάτωσης του backend ελεγκτή **(*controller-based backend – CHASE APIs*),** της χρήσης του ***OpenSearch***[4] για αποθήκευση δεδομένων, καθώς και της αξιοποίησης του ***Apache Kafka***[5] για ροή μηνυμάτων σε πραγματικό χρόνο (*real-time messaging*).

Το εν λόγω παραδοτέο καθόρισε τις κύριες ροές δεδομένων που υποστηρίζουν τη συλλογή περιεχομένου, την ανάλυση και την οπτικοποίηση, μέσω των *indices* **chase_posts**, **chase_comments** και **chase_analytics**, καθώς και των αντίστοιχων *Kafka topics*. Τα στοιχεία αυτά αποτέλεσαν το θεμέλιο για την ολοκληρωμένη ροή επεξεργασίας δεδομένων — από την εξωτερική ανάκτηση περιεχομένου έως την ανάλυση κειμένου και την απεικόνιση των αποτελεσμάτων.

Με βάση αυτή την αρχιτεκτονική, η τελική έκδοση του εργαλείου CHASE ICT εισάγει στοχευμένες βελτιώσεις που επεκτείνουν τις αναλυτικές δυνατότητες και βελτιώνουν την απόδοση του συστήματος. Υλοποιήθηκαν δύο βασικές αναπτύξεις:

1. Μια νέα ροή υπολογισμού *n-gram* με ειδικά *indices* που υποστηρίζουν προηγμένες δυνατότητες οπτικοποίησης.
2. Ένας μηχανισμός ασύγχρονης εκτέλεσης των υπηρεσιών ανάλυσης κειμένου (*text analytics services*), που επιτρέπει ταχύτερη και πιο αποδοτική εξαγωγή συμπερασμάτων (*inference*).

Οι ακόλουθες υποενότητες περιγράφουν αναλυτικά τις συγκεκριμένες ενημερώσεις.

Ροή Υπολογισμού N-gram

Για την υποστήριξη της λειτουργικότητας οπτικοποίησης *n-gram*, η οποία εισήχθη στο στοιχείο οπτικοποίησης του έργου **(*Deliverable D3.6 – CHASE Visualization Component*)**, αναπτύχθηκε και ενσωματώθηκε μια νέα ροή επεξεργασίας δεδομένων στο *backend* του εργαλείου CHASE.

Η ροή αυτή εκτελείται ως προγραμματισμένη διεργασία (*scheduled task*) που περιοδικά εξάγει, επεξεργάζεται και συγκεντρώνει δεδομένα κειμένου από το *index* **chase_comments**, με σκοπό τη δημιουργία στατιστικών *n-gram*.

Για την υποστήριξη της διαδικασίας, προστέθηκαν δύο νέα *OpenSearch indices*:
• **chase_comments_ngram** – αποθηκεύει συγκεντρωτικά δεδομένα *n-gram* με μετρήσεις συχνότητας.
• **chase_comments_ngram_processed** – παρακολουθεί την κατάσταση επεξεργασίας των σχολίων, ώστε να αποφεύγεται η διπλή επεξεργασία.

Η ροή δεδομένων για τον υπολογισμό των ***n-gram*** ακολουθεί μια δομημένη και αυτοματοποιημένη διαδικασία. Η διεργασία εκτελείται σε προκαθορισμένα χρονικά διαστήματα (ανά 10 λεπτά). Σε κάθε εκτέλεση, προσδιορίζει το χρονικό εύρος προς επεξεργασία, ελέγχοντας την πιο πρόσφατη καταχώριση στο *chase_comments_ngram_processed* ή, εάν είναι η πρώτη εκτέλεση, ξεκινά από το παλαιότερο διαθέσιμο σχόλιο.

Τα σχόλια στο επιλεγμένο χρονικό διάστημα ανακτώνται σε ομάδες των δέκα. Για κάθε ομάδα, τα σχόλια που έχουν ήδη επεξεργαστεί φιλτράρονται βάσει των *IDs* τους που έχουν καταγραφεί στο ***chase_comments_ngram_processed***.

[4] https://opensearch.org/
[5] https://kafka.apache.org/

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Κάθε νέο σχόλιο υποβάλλεται σε διαδικασία *normalization*, συμπεριλαμβανομένης της αφαίρεσης *URLs*, ειδικών χαρακτήρων και αλλαγών γραμμής. Το κείμενο χωρίζεται σε λέξεις (*tokenization*), και αφαιρούνται οι *stop words*.

Με τη μέθοδο ***sliding window***, παράγονται όλα τα πιθανά *n-grams* έως ένα συγκεκριμένο μήκος (μέχρι πέντε λέξεις, όπως έχει καθοριστεί στις ρυθμίσεις). Έτσι δημιουργούνται ακολουθίες από μία έως πέντε διαδοχικές λέξεις για κάθε σχόλιο.
Τα παραγόμενα *n-grams* συγκεντρώνονται προσωρινά στη μνήμη για την τρέχουσα ομάδα. Διπλότυπες εγγραφές, με κοινό κείμενο, ημερομηνία, κατηγορία, ετικέτα και πηγή δεδομένων, συγχωνεύονται με άθροιση των μετρήσεων συχνότητας πριν την αποθήκευση.

Τα συγκεντρωτικά αποτελέσματα ενημερώνουν (μέσω *upsert operations*) το ***index*** **chase_comments_ngram**, είτε δημιουργώντας νέες εγγραφές είτε αυξάνοντας τις υπάρχουσες μετρήσεις συχνότητας. Τα *IDs* των επεξεργασμένων σχολίων και οι αντίστοιχες ημερομηνίες τους καταγράφονται στο ***chase_comments_ngram_processed***.

Η διαδικασία συνεχίζεται επαναληπτικά για όλα τα διαθέσιμα χρονικά παράθυρα και στη συνέχεια τίθεται σε παύση έως την επόμενη προγραμματισμένη εκτέλεση.

Η αρθρωτή (modular) αυτή σχεδίαση εξασφαλίζει επεκτασιμότητα και επιτρέπει στο στοιχείο οπτικοποίησης να έχει συνεχή πρόσβαση σε **ενημερωμένα δεδομένα *n-gram*** για δυναμικές αναλυτικές απεικονίσεις.

Ασύγχρονη Εκτέλεση Ανάλυσης Κειμένου

Το **Παραδοτέο *D3.5 – CHASE Real-Time Text Analyser*** παρουσίασε την υλοποίηση του στοιχείου ανάλυσης κειμένου (*text analytics component*) που είναι υπεύθυνο για τον εντοπισμό και την ταξινόμηση διαδικτυακού λόγου μίσους με βάση το φύλο (*gender-based hate speech*) στο εργαλείο CHASE.

Στο εν λόγω παραδοτέο περιγράφηκαν δύο κύριες αναλυτικές ενότητες:

1. Ένα **δυαδικό μοντέλο ταξινόμησης** (*binary classification model*) για την αναγνώριση περιεχομένου μίσους.
2. Ένα **agentic workflow** για την πιο λεπτομερή ανάλυση και ανίχνευση προτύπων ρητορικής μίσους.

Στην τελική έκδοση της πλατφόρμας, το στοιχείο αυτό έχει βελτιστοποιηθεί περαιτέρω ώστε να βελτιωθεί η απόκριση και η επεκτασιμότητα του συστήματος.

Για την επίτευξη του στόχου αυτού, η εκτέλεση των δύο μοντέλων έχει **αποσυνδεθεί και καταστεί ασύγχρονη**, επιτρέποντας στο ταχύτερο δυαδικό μοντέλο να λειτουργεί ανεξάρτητα από την πιο απαιτητική, σε υπολογιστικούς πόρους, *agentic analysis*.

Αυτή η βελτίωση εξασφαλίζει ότι τα αποτελέσματα της ταξινόμησης είναι διαθέσιμα σε πραγματικό χρόνο, ενώ η βαθύτερη ανάλυση των προτύπων ρητορικής μίσους εκτελείται στο παρασκήνιο.
Η αναβαθμισμένη ροή, που περιγράφεται παρακάτω, διατηρεί την ακεραιότητα των δεδομένων και τη δομή που βασίζεται σε μηνύματα (*message-driven structure*) όπως καθορίστηκε στο *D3.3*, ενώ παράλληλα βελτιώνει σημαντικά τον ρυθμό επεξεργασίας και την εμπειρία του χρήστη.

Η ενημερωμένη ροή λειτουργεί ως εξής:

1. Το στοιχείο *Text Analytics* καταναλώνει σχόλια από το *Kafka topic* **chase_comments**.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

2. Εκτελεί πρώτα το **δυαδικό μοντέλο ταξινόμησης**, το οποίο εντοπίζει εάν ένα σχόλιο περιέχει λόγο μίσους.
3. Τα αποτελέσματα της ταξινόμησης αποστέλλονται στο *Kafka topic* **chase_analytics**.
4. Μέσω του *Storage Connector*, τα αποτελέσματα αποθηκεύονται στο *OpenSearch index* **chase_analytics**.
5. Μία δευτερεύουσα υπηρεσία, με όνομα **hate_speech_patterns_analytics**, καταναλώνει ανεξάρτητα τα μηνύματα από το *chase_analytics topic*. Εκτελεί το *agentic workflow* για τον εντοπισμό λεπτομερών προτύπων ρητορικής μίσους και ενημερώνει τις αντίστοιχες εγγραφές στο *index chase_analytics*.

Η ασύγχρονη αυτή αρχιτεκτονική εξασφαλίζει ότι το πιο αργό *agentic workflow* δεν επιβραδύνει την κύρια ροή εισαγωγής και ταξινόμησης δεδομένων σε πραγματικό χρόνο.

Ο διαχωρισμός των δύο μοντέλων βελτιώνει τη **διεκπεραιωτική ικανότητα (throughput)**, μειώνει τη **χρονική καθυστέρηση (latency)** κατά τη δημιουργία ειδοποιήσεων και επιτρέπει **συνεχή ενημέρωση** του *analytics index* καθώς προκύπτουν νέα, πιο λεπτομερή αποτελέσματα ανάλυσης.

Ενσωμάτωση Ανίχνευσης LLM (LLM Tracing)

Για την ενίσχυση της **διαφάνειας**, της **παρακολούθησης απόδοσης** και της **διευκόλυνσης αποσφαλμάτωσης** του στοιχείου *text analytics*, ενσωματώθηκε στο εργαλείο CHASE ICT η δυνατότητα **LLM tracing**, μέσω της πλατφόρμας **Langfuse**[6].

Η λειτουργικότητα αυτή επιτρέπει λεπτομερή παρακολούθηση των λειτουργιών του *Large Language Model (LLM)* που συμμετέχει στο *agentic workflow* για την ανίχνευση προτύπων ρητορικής μίσους με βάση το φύλο (*gendered hate-speech pattern detection*).

Το **Langfuse** είναι μια ανοιχτού κώδικα πλατφόρμα παρατηρησιμότητας και ανάλυσης (*open-source observability and analytics platform*), σχεδιασμένη ειδικά για εφαρμογές που αξιοποιούν μεγάλα γλωσσικά μοντέλα. Παρέχει δυνατότητα αναλυτικής καταγραφής κάθε αλληλεπίδρασης με το μοντέλο, συλλέγοντας δομημένα μεταδεδομένα, όπως:

- το *prompt text*,
- το *completion output*,
- τη *latency*,
- τα *cost metrics*, και
- τις ιεραρχικές σχέσεις κλήσεων (*hierarchical call relationships*) μεταξύ των διαφορετικών σταδίων ενός *agentic workflow*.

Με την πραγματικού χρόνου συγκέντρωση αυτών των *traces*, το *Langfuse* παρέχει στους τεχνικούς και ερευνητές ένα ενοποιημένο περιβάλλον παρακολούθησης και ανάλυσης της απόδοσης, διευκολύνοντας τον εντοπισμό καθυστερήσεων στα *prompts* και την επαλήθευση της συμπεριφοράς του μοντέλου σε μεγάλο όγκο εκτελέσεων.

Στο περιβάλλον ανάπτυξης του CHASE, το *Langfuse* έχει συνδεθεί με το *agentic workflow* ανάλυσης προτύπων ρητορικής μίσους, που λειτουργεί εντός της υπηρεσίας *text analytics*. Κάθε εκτέλεση αυτής της ροής, που πραγματοποιείται μέσω του *chase_agent module*, καταγράφεται αυτόματα ως **trace**.
Κάθε *trace* αντιστοιχεί σε μία πλήρη εκτέλεση της αναλυτικής διαδικασίας για ένα συγκεκριμένο δημοσίευμα ή σχόλιο και περιλαμβάνει πολλαπλά **sub-spans**, δηλαδή επιμέρους βήματα συλλογισμού του LLM (π.χ. *text summarization*, *pattern classification*, *validation*).

[6] https://langfuse.com/

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Για κάθε *sub-span*, το *Langfuse* καταγράφει τη **διάρκεια εκτέλεσης (duration)**, τον **αριθμό tokens**, και την **κατάσταση ολοκλήρωσης (status)**. Αυτή η λεπτομερής καταγραφή επιτρέπει τη συσχέτιση μεταξύ χαρακτηριστικών εισόδου, απόδοσης μοντέλου και αξιοπιστίας εξόδου.

Το Σχήμα 3 παρουσιάζει τον **Πίνακα Ανίχνευσης *Langfuse* (*Langfuse Tracing Dashboard*),** ο οποίος προσφέρει συγκεντρωτική προβολή των πιο πρόσφατων εκτελέσεων της ροής *chase_agent*.
Κάθε γραμμή στον πίνακα αντιπροσωπεύει ένα ***trace*** που έχει καταγραφεί για το ***agentic workflow*** **του CHASE.**

Πέρα από τα ***timestamps***, τα ***user identifiers*** και τα ***organizational tags***, ο πίνακας εμφανίζει ποσοτικά στοιχεία απόδοσης, όπως:

- **latency** (χρόνος εκτέλεσης ανά ανάλυση),
- **token usage** (εισροή, εκροή και συνολικά tokens), και
- **estimated cost** (εκτιμώμενο κόστος λειτουργίας).

Η διεπαφή επιτρέπει στην τεχνική ομάδα να παρακολουθεί την απόδοση, να συγκρίνει εκτελέσεις και να εντοπίζει τυχόν αποκλίσεις ή προβλήματα σε βάθος χρόνου.

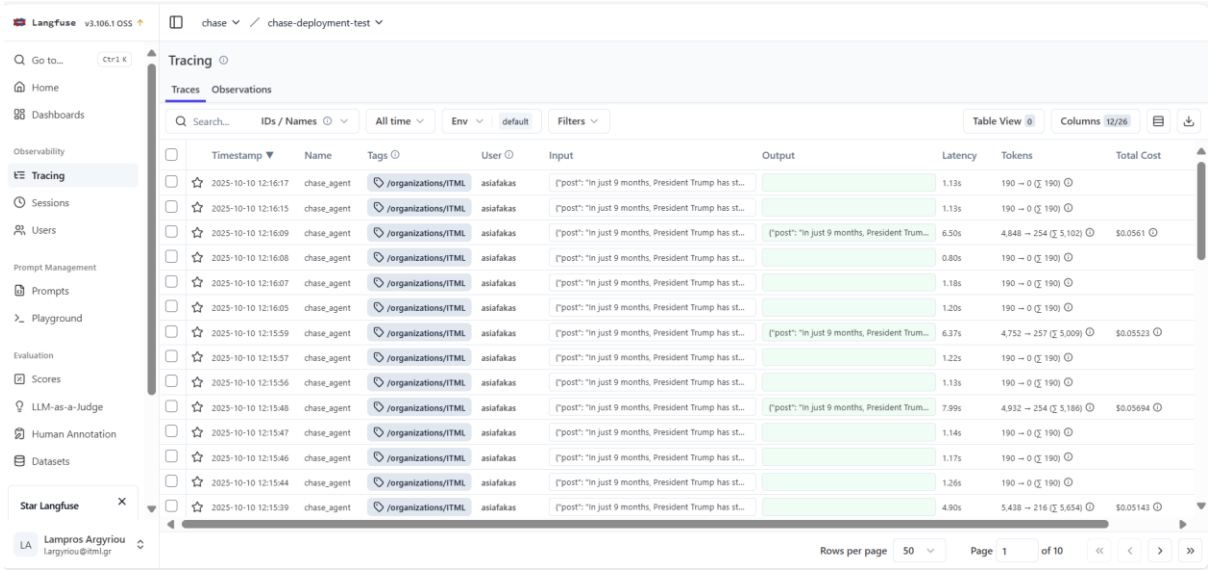

Figure 3: Πίνακας Tracing του Langfuse που εμφανίζει πρόσφατες εκτελέσεις της πρακτορικής ροής εργασιών (agentic workflow) του CHASE..

Το **Σχήμα 4** παρουσιάζει μια αναλυτική προβολή ενός συγκεκριμένου *trace*.

- Στο **αριστερό τμήμα**, απεικονίζεται το **ιεραρχικό γράφημα εκτέλεσης (hierarchical execution graph)**, όπου κάθε κόμβος αντιπροσωπεύει μία ξεχωριστή κλήση LLM που αντιστοιχεί σε στάδιο της ανάλυσης (π.χ. *summarize_post*, *classify_category*, *evaluate_classification*).
- Στο **κεντρικό τμήμα**, εμφανίζονται αναλυτικά στατιστικά για κάθε κλήση, όπως η *latency* και ο αριθμός *tokens*.
- Στο **δεξιό τμήμα**, παρουσιάζονται τα δομημένα δεδομένα εισόδου και εξόδου του μοντέλου.

Στο κάτω μέρος του πίνακα εμφανίζεται ένα **διάγραμμα ροής διαδικασίας (process flow diagram)** που απεικονίζει τη λογική ακολουθία από την αρχικοποίηση έως τα ενδιάμεσα βήματα συλλογισμού και την τελική έξοδο.

Μέσω αυτής της ενσωμάτωσης, το εργαλείο CHASE ICT υποστηρίζει πλέον **πλήρη και διαφανή παρακολούθηση των διαδικασιών που βασίζονται σε LLM**. Οι προγραμματιστές μπορούν να

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

ανατρέχουν σε ιστορικά *traces*, να μετρούν την απόδοση και να συσχετίζουν συγκεκριμένα *prompts* με τα αντίστοιχα αναλυτικά αποτελέσματα.

Αυτή η λειτουργικότητα ενισχύει την **αξιοπιστία**, τη **συντηρησιμότητα** και τη **διαφάνεια** του στοιχείου *text analytics*, διασφαλίζοντας ότι όλες οι λειτουργίες που βασίζονται σε γλωσσικά μοντέλα παραμένουν **μετρήσιμες, ελέγξιμες και επαληθεύσιμες** καθ' όλη τη διάρκεια της ανάπτυξης και της πιλοτικής εφαρμογής της πλατφόρμας CHASE.

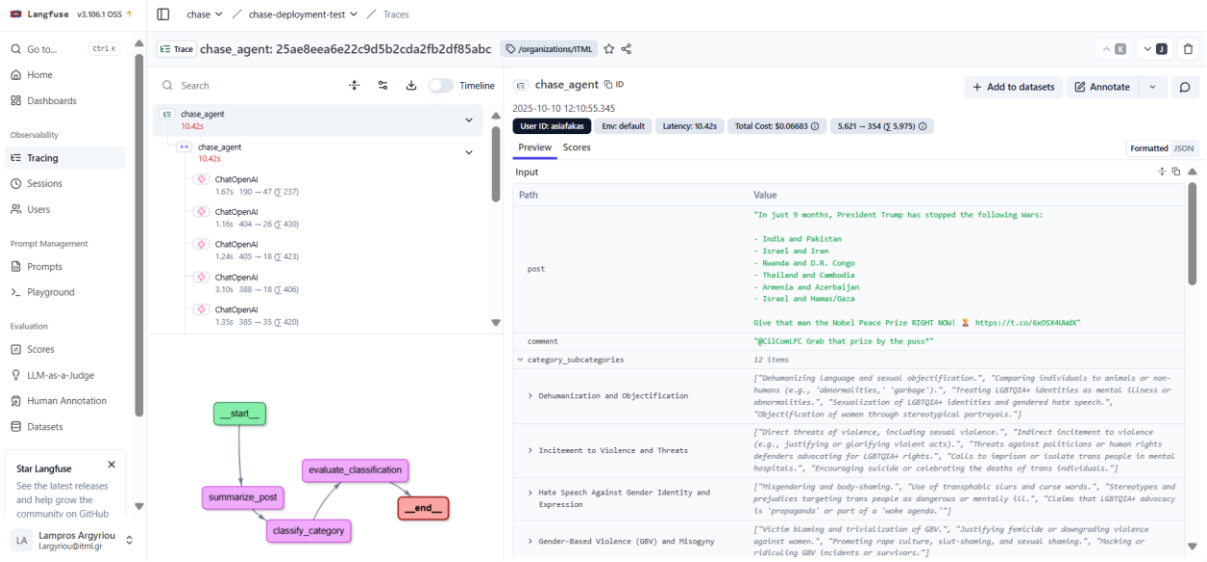

Figure 4: Αναλυτική προβολή ενός trace στο Langfuse που αναπαριστά μία εκτέλεση της πρακτορικής ροής εργασιών του CHASE. Η απεικόνιση αναδεικνύει την εσωτερική πορεία συλλογισμού του μοντέλου και τα αντίστοιχα αναλυτικά αποτελέσματα..

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Κατακλείδα

Η ολοκλήρωση αυτού του παραδοτέου σηματοδοτεί την τελική διαμόρφωση του εργαλείου CHASE ως μία ολοκληρωμένη και πλήρως ενσωματωμένη πλατφόρμα. Μέσα από διαδοχικά στάδια ανάπτυξης, κάθε στοιχείο έχει βελτιωθεί, διασυνδεθεί και επικυρωθεί, ώστε να διασφαλιστεί η συνεκτική λειτουργία και η ευθυγράμμιση με τους στόχους του έργου.

Το αποτέλεσμα είναι μια σταθερή, λειτουργική λύση που συνδυάζει αυτοματοποίηση, ερμηνευσιμότητα και προσβασιμότητα σε ένα ενιαίο πλαίσιο.

Το ορόσημο αυτό ολοκληρώνει τις τεχνικές δραστηριότητες ανάπτυξης του WP3 και δημιουργεί τις προϋποθέσεις για την αξιολόγηση και επικύρωση του εργαλείου CHASE ICT κατά τη φάση των πιλοτικών δοκιμών.

Η ανατροφοδότηση από την πραγματική χρήση θα καθοδηγήσει τις τελικές βελτιώσεις και θα στηρίξει τη μακροπρόθεσμη βιωσιμότητα της πλατφόρμας πέρα από τη διάρκεια ζωής του έργου.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

CH%#
!#ASE

Combatting online
HAte Speech by
engaging online mEdia

D 3.8
Versione finale dello strumento informatico CHASE

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Indice

Sommario 3
Introduzione 3
Scopo del documento 3
Struttura del documento 3
Deployment e gestione utenti 4
Panoramica dell'ambiente della piattaforma CHASE distribuita 4
Processo di autenticazione e gestione degli utenti 4
Aggiornamenti del backend e del layer dati 7
Flusso di lavoro di calcolo degli n-grammi 7
Esecuzione asincrona della text analytics 7
Integrazione del tracing LLM 9
Conclusione 11

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Sommario esecutivo

Questo documento presenta la versione finale dello **strumento informatico CHASE**, sviluppato nell'ambito del *Work Package 3 (WP3) -- Componenti tecnologici e piattaforma CHASE*.

Basandosi sulle specifiche tecniche e le implementazioni incrementali descritte nei deliverable precedenti (D3.1, D3.3, D3.5 e D3.6), questo rapporto consolida la versione completa e distribuita del sistema CHASE e sintetizza gli ultimi perfezionamenti tecnici che hanno condotto alla sua piena operatività.

Con questi perfezionamenti, lo strumento informatico CHASE raggiunge la sua piena maturità funzionale, pronto per il deployment pilota e la validazione nel mondo reale da parte delle organizzazioni partner. Rappresenta il consolidamento di tutti i risultati tecnologici all'interno del WP3 e stabilisce una base solida per il futuro uso operativo e la sostenibilità.

Introduzione

Scopo del documento

Questo deliverable presenta la **versione finale integrata e operativa** dello strumento informatico CHASE. Descrive il sistema nella sua forma completata, riassumendo gli ultimi aggiornamenti che sono stati implementati.

Il documento evidenzia tre aspetti principali:

- L'integrazione dei componenti della piattaforma in un ambiente operativo unificato che supporta l'accesso e l'interazione multi-partner.
- La configurazione finalizzata e la gestione degli accessi utente.
- Gli ultimi miglioramenti ai flussi di lavoro di dati e analytics, sviluppati per migliorare la visualizzazione e la reattività analitica.

Nel complesso, il documento funge da rapporto tecnico conclusivo del WP3, dimostrando che lo strumento informatico CHASE ha raggiunto il livello previsto di integrazione e prontezza per la validazione nella fase pilota.

Struttura del documento

Il resto di questo documento è strutturato come segue:

- **Sezione 2** descrive l'approccio utilizzato per gestire l'accesso utente e l'autenticazione tra le organizzazioni partner.
- **Sezione 3** delinea gli ultimi miglioramenti ai flussi di lavoro interni del sistema.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Deployment e gestione utenti

Panoramica dell'ambiente della piattaforma CHASE distribuita

La versione finale dello strumento informatico CHASE è stata completamente distribuita su un **cluster Kubernetes (K8s)**[1] dedicato ospitato sull'infrastruttura di ITML. Ogni servizio è stato containerizzato utilizzando Docker[2]. Questa configurazione fornisce un ambiente di runtime sicuro, controllato e scalabile che consente un controllo operativo granulare. Tutti i componenti del sistema sono containerizzati e gestiti tramite Kubernetes.

Il cluster è configurato per orchestrare il deployment dei servizi, lo scaling e la gestione del ciclo di vita attraverso manifest dichiarativi. Ogni componente è distribuito come un pod Kubernetes separato, raggruppato in namespace corrispondenti ai principali sottosistemi dello strumento informatico CHASE. I servizi stateful, come OpenSearch e Keycloak, utilizzano **Volumi Persistenti (PV)** per garantire la durabilità dei dati attraverso riavvii e aggiornamenti. Gli ingress controller espongono solo gli endpoint pubblici necessari (ad esempio, l'interfaccia web CHASE e il portale di autenticazione Keycloak) tramite **HTTPS**, mentre la comunicazione interna dei servizi rimane limitata alla rete privata del cluster.

La comunicazione inter-servizi tra i componenti CHASE si basa su una combinazione di **API REST** e streaming di eventi asincrono tramite **Apache Kafka**. Per salvaguardare questo livello di comunicazione, è stato implementato il **TLS mutuo (mTLS)**[3] tra i producer e consumer Kafka, garantendo sia la riservatezza dei messaggi che l'autenticazione bidirezionale. Questo approccio assicura che solo i servizi fidati all'interno del cluster Kubernetes possano pubblicare o consumare eventi, rafforzando così la postura di sicurezza complessiva della piattaforma.

Il monitoraggio continuo dell'ambiente distribuito è ottenuto utilizzando probe di health Kubernetes, meccanismi di logging interni e dashboard OpenSearch. Questa configurazione abilita l'osservabilità in tempo reale, il rilevamento dei guasti e una manutenzione semplificata.

Nella sua forma distribuita finale, lo strumento informatico CHASE opera come un sistema coeso e production-ready che integra tutti i componenti core in un ambiente unificato, sicuro e manutenibile, supportando il flusso di lavoro completo di moderazione e analytics definito all'interno del Work Package 3.

Processo di autenticazione e gestione degli utenti

L'implementazione dell'autenticazione utente e del controllo degli accessi nella versione finale dello strumento informatico CHASE segue la configurazione precedentemente descritta nel *Deliverable D3.1 -- Specifiche tecniche*, dove l'integrazione di **Keycloak** come soluzione di **gestione delle identità e degli accessi (IAM)** è stata introdotta per la prima volta. In questa versione finale, Keycloak è stato completamente integrato nell'ambiente distribuito e serve come punto di ingresso centrale per l'autenticazione e l'autorizzazione degli utenti all'interno della piattaforma CHASE.

[1] https://kubernetes.io/
[2] https://www.docker.com/
[3] https://en.wikipedia.org/wiki/Transport_Layer_Security

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

La gestione degli utenti viene eseguita direttamente attraverso la **console di amministrazione Keycloak**, che consente agli amministratori di sistema di creare, aggiornare e disattivare account utente nonché di assegnare loro ruoli predefiniti.

Keycloak opera sotto un realm dedicato denominato **"chase"**, che definisce tutti gli utenti attivi, i client e le policy associate allo strumento. Sono stati configurati due client applicativi principali:

- **chase-client** --- che rappresenta l'applicazione web CHASE, utilizzato per l'autenticazione lato utente tramite OpenID Connect.
- **chase-system-client** --- che rappresenta la comunicazione di sistema interna tra i servizi backend e il livello di autenticazione.

Il processo di registrazione degli utenti per la piattaforma CHASE è stato coordinato centralmente dal team tecnico di ITML. A ciascuna organizzazione partecipante è stato richiesto di fornire i seguenti dettagli per ogni individuo che necessitava di accesso allo strumento:

- Nome
- Username
- Password
- Nome dell'organizzazione

Dopo aver raccolto queste informazioni, gli amministratori di sistema hanno creato un **gruppo** corrispondente in Keycloak per ciascuna organizzazione rappresentata nel progetto (ad esempio, ITML, Proto Thema, MDIG, CESIE, WAN-IFRA, ecc.). Ogni account utente è stato quindi creato manualmente all'interno del realm *chase* e assegnato al rispettivo gruppo organizzativo. Questa struttura basata sui gruppi garantisce che gli account utente siano chiaramente associati alla loro organizzazione e semplifica la gestione di più partner all'interno di un singolo ambiente.

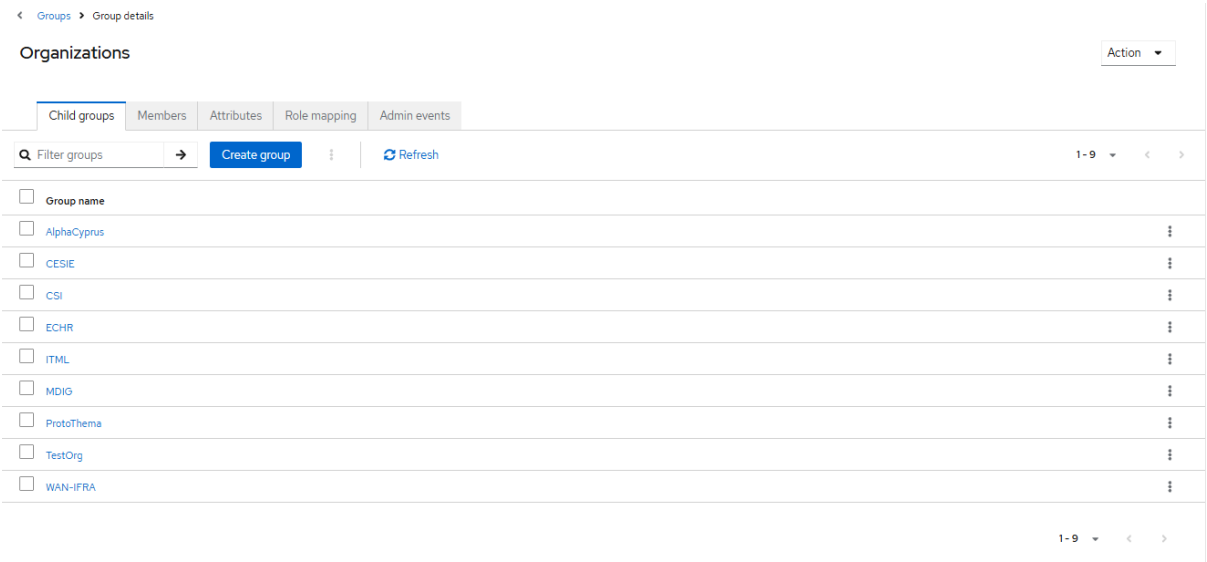

Figura 1: Gruppi di organizzazioni partner definiti all'interno del realm CHASE di Keycloak.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

La registrazione degli utenti non è intenzionalmente **aperta all'auto-registrazione**. Tutti gli account sono creati e attivati dagli amministratori per preservare il controllo sull'accesso al sistema e per mantenere un collegamento verificabile tra gli utenti e le loro organizzazioni. Le funzionalità di reset della password e verifica email sono disabilitate, poiché la gestione delle credenziali è gestita esclusivamente attraverso flussi di lavoro amministrativi.

Questa configurazione garantisce che l'intero processo di gestione degli utenti rimanga trasparente, tracciabile e coerente in tutto il consorzio. Tutte le richieste di autenticazione sono elaborate attraverso l'interfaccia di login sicura di Keycloak, che si integra perfettamente con i servizi frontend e backend di CHASE. Il risultato è un flusso di lavoro di accesso controllato e verificabile che supporta il funzionamento affidabile dello strumento informatico CHASE in un contesto multi-organizzativo.

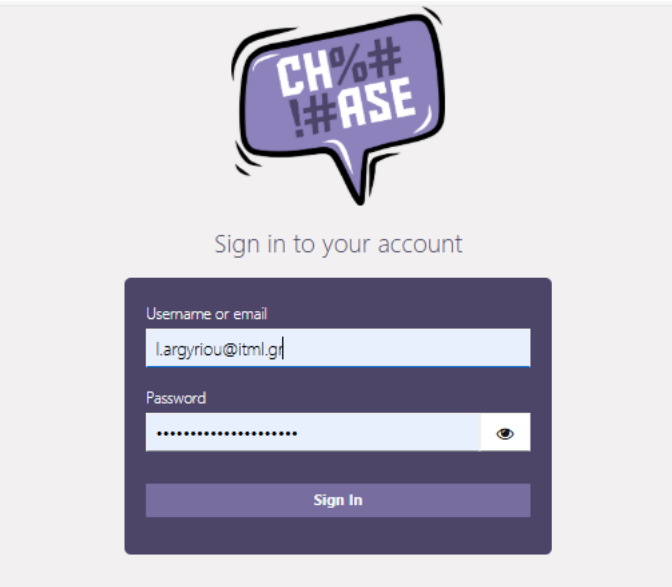

Figura 2: Pagina di login CHASE personalizzata integrata con Keycloak, che fornisce autenticazione sicura per gli utenti registrati all'interno del realm CHASE

Lo strumento informatico CHASE è accessibile agli utenti autorizzati tramite un'interfaccia web sicura ospitata sotto il dominio dedicato del progetto. L'accesso è fornito tramite il seguente indirizzo:

https://chase-web.itml.space/

Questo punto di ingresso funge da gateway unificato all'ambiente CHASE, dove le organizzazioni partner possono effettuare il login utilizzando le credenziali loro assegnate. Il dominio supporta la comunicazione crittografata (HTTPS) e si connette direttamente al servizio di autenticazione, garantendo l'accesso sicuro a tutte le funzionalità della piattaforma CHASE.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Aggiornamenti del backend e del layer dati

Il deliverable **D3.3 CHASE Data Crawler & Data Repository** ha documentato l'implementazione dell'infrastruttura dati core dello strumento informatico CHASE, inclusa l'integrazione del backend basato su controller (API CHASE), l'adozione di **OpenSearch**[4] per lo storage dei dati e l'uso di **Apache Kafka**[5] per la messaggistica in tempo reale. Il deliverable ha definito i principali flussi di dati che supportano l'ingestione, l'analisi e la visualizzazione dei contenuti attraverso gli indici **chase_posts**, **chase_comments** e **chase_analytics**, nonché i corrispondenti topic Kafka. Questi componenti hanno stabilito le fondamenta per l'elaborazione end-to-end dei dati, dal recupero di contenuti esterni all'analytics testuale e alla visualizzazione.

Basandosi su questa architettura, la versione finale dello strumento informatico CHASE introduce miglioramenti mirati per estendere le capacità analitiche e migliorare le prestazioni del sistema. Sono stati implementati due sviluppi chiave:

1. un nuovo **flusso di lavoro di calcolo degli n-grammi** con indici dedicati per supportare la visualizzazione avanzata
2. un **meccanismo di esecuzione asincrona** per i servizi di text analytics, che abilita un'inferenza più veloce ed efficiente.

Le sottosezioni seguenti descrivono questi aggiornamenti in dettaglio.

Flusso di lavoro di calcolo degli n-grammi

Per abilitare la **funzionalità di visualizzazione degli n-grammi** introdotta nel componente di visualizzazione (*Deliverable D3.6 -- Componente di visualizzazione CHASE*), è stato sviluppato e integrato nel backend dello strumento informatico CHASE un nuovo flusso di lavoro di elaborazione dei dati. Questo flusso di lavoro opera come un task pianificato che estrae, elabora e aggrega periodicamente i dati testuali dall'indice **chase_comments** per generare statistiche di n-grammi.

Due nuovi indici OpenSearch sono stati aggiunti per supportare questo processo:

- **chase_comments_ngram** -- che memorizza i dati di n-grammi aggregati con conteggi di frequenza.
- **chase_comments_ngram_processed** -- che traccia lo stato di elaborazione dei commenti per evitare duplicazioni.

Il flusso di dati per il calcolo degli n-grammi segue una pipeline strutturata e automatizzata. Il task viene eseguito a intervalli predefiniti (10 minuti). Ogni esecuzione determina la finestra temporale da elaborare controllando l'entrata più recente in **chase_comments_ngram_processed**, o usa come default il commento disponibile più vecchio se si tratta della prima esecuzione.

[4] https://opensearch.org/
[5] https://kafka.apache.org/

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

I commenti all'interno dell'intervallo di date selezionato vengono recuperati in batch di dieci. Per ogni batch, i commenti precedentemente elaborati vengono filtrati in base ai loro ID registrati in chase_comments_ngram_processed.

Ogni commento non elaborato subisce una normalizzazione, inclusa la rimozione degli URL, dei caratteri speciali e degli a capo. Il testo viene tokenizzato in parole, e le stop word vengono rimosse.

Utilizzando un metodo a finestra scorrevole, vengono generati tutti i possibili n-grammi fino a una lunghezza specificata (configurata a cinque parole). Questo produce sequenze da una a cinque parole consecutive per ogni commento.

Gli n-grammi generati vengono aggregati in memoria per il batch corrente. Le voci duplicate, che condividono lo stesso testo, data, categoria, etichetta e fonte dati, vedono i loro conteggi di frequenza sommati prima della persistenza.

I risultati aggregati vengono inseriti o aggiornati in chase_comments_ngram, creando nuovi record o incrementando i conteggi di frequenza esistenti. Gli ID dei commenti elaborati e le loro date vengono quindi registrati in chase_comments_ngram_processed.

Il processo continua iterativamente su tutte le pagine all'interno dell'intervallo di date selezionato e poi si mette in pausa fino alla successiva esecuzione pianificata.

Questo design modulare garantisce la scalabilità e consente al componente di visualizzazione di accedere a dati di n-grammi continuamente aggiornati per display analitici dinamici.

Esecuzione asincrona della text analytics

Il deliverable D3.5 -- CHASE Real-Time Text Analyser ha presentato l'implementazione del componente di text analytics responsabile del rilevamento e della classificazione dell'hate speech di genere all'interno dello strumento informatico CHASE. Ha descritto due moduli analitici principali: un modello di classificazione binaria per identificare contenuti di odio e un flusso di lavoro agentico per il rilevamento granulare dei pattern di hate speech. Nella versione finale della piattaforma, questo componente è stato ulteriormente ottimizzato per migliorare la reattività e la scalabilità.

Per raggiungere questo obiettivo, l'esecuzione dei due modelli è stata disaccoppiata e resa asincrona, consentendo al classificatore binario più veloce di operare indipendentemente dall'analisi agentica più intensiva dal punto di vista computazionale. Questo miglioramento garantisce che i risultati di classificazione siano disponibili in tempo reale mentre l'analisi più approfondita dei pattern continua in background.

Il flusso di lavoro aggiornato, descritto di seguito, preserva l'integrità dei dati e la struttura message-driven introdotta in D3.3 migliorando significativamente il throughput e l'esperienza utente.

Il flusso di lavoro aggiornato opera come segue:

1. Il componente di Text Analytics consuma i commenti dal topic Kafka chase_comments.
2. Esegue prima il modello di classificazione binaria, identificando se un commento contiene hate speech.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them. [Project Number: 101143159]

3. I risultati di classificazione vengono scritti nel topic Kafka **chase_analytics**.
4. Attraverso lo **Storage Connector**, questi risultati vengono persistiti nell'indice OpenSearch **chase_analytics**.
5. Un servizio secondario, **hate_speech_patterns_analytics**, consuma indipendentemente i messaggi dal topic **chase_analytics**. Esegue il **flusso di lavoro agentico** per rilevare pattern granulari di hate speech e quindi invoca gli aggiornamenti delle corrispondenti entry nell'indice chase_analytics.

Questa architettura asincrona garantisce che il flusso di lavoro agentico più lento non blocchi la pipeline di ingestione e classificazione in tempo reale. La separazione dei due modelli migliora il throughput, riduce la latenza per la generazione di alert e fornisce aggiornamenti continui all'indice analytics man mano che le analisi più approfondite diventano disponibili.

Integrazione del tracing Large Language Model (LLM)

Per migliorare la trasparenza, il debugging e la valutazione delle prestazioni del componente di text analytics, **il tracing LLM è stato integrato nello strumento informatico CHASE utilizzando Langfuse**[6]. Questa aggiunta fornisce un'osservabilità dettagliata delle operazioni dei LLM coinvolte nel flusso di lavoro agentico responsabile del rilevamento dei pattern di hate speech.

Langfuse è una piattaforma open-source di osservabilità e analytics specificamente progettata per applicazioni alimentate da large language model. Abilita il tracing dettagliato di ogni interazione del modello, catturando metadati strutturati come il testo del prompt, l'output di completion, la latenza, le metriche di costo e le relazioni gerarchiche tra diverse fasi di un flusso di lavoro agentico. Aggregando questi trace in tempo reale, Langfuse fornisce a sviluppatori e ricercatori un'interfaccia unificata per analizzare le prestazioni, identificare colli di bottiglia nei prompt e validare il comportamento del modello su grandi volumi di esecuzioni.

Nel deployment CHASE, Langfuse è connesso al **flusso di lavoro agentico di rilevamento dei pattern di hate speech** all'interno del servizio di text analytics. Ogni invocazione del flusso di lavoro (eseguita dal modulo *chase_agent*) viene automaticamente registrata come trace. Un trace corrisponde a un'esecuzione analitica completa della pipeline agentica per un determinato post o commento e include più sub-span che rappresentano le chiamate LLM coinvolte in ogni step di reasoning (ad esempio, riassunto del testo, classificazione dei pattern, validazione). Per ogni sub-span, Langfuse registra la durata, il conteggio dei token e lo stato dell'operazione. Questo logging granulare consente la correlazione tra caratteristiche di input, prestazioni del modello e affidabilità dell'output.

La **Figura 3** mostra la *Dashboard di tracing* di Langfuse, che fornisce una vista aggregata delle esecuzioni recenti. Ogni riga nella tabella rappresenta un singolo trace registrato per il **flusso di lavoro *chase_agent***. Accanto a **timestamp**, **identificatori utente** e tag organizzativi, la dashboard visualizza **indicatori di performance quantitativi** inclusi **latency** (tempo di esecuzione per run), **uso dei token** (input/output e totale) e il **costo stimato** di ogni operazione. Questa interfaccia consente al team tecnico di monitorare il throughput e le tendenze delle prestazioni nel tempo.

[6] https://langfuse.com/

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them. [Project Number: 101143159]

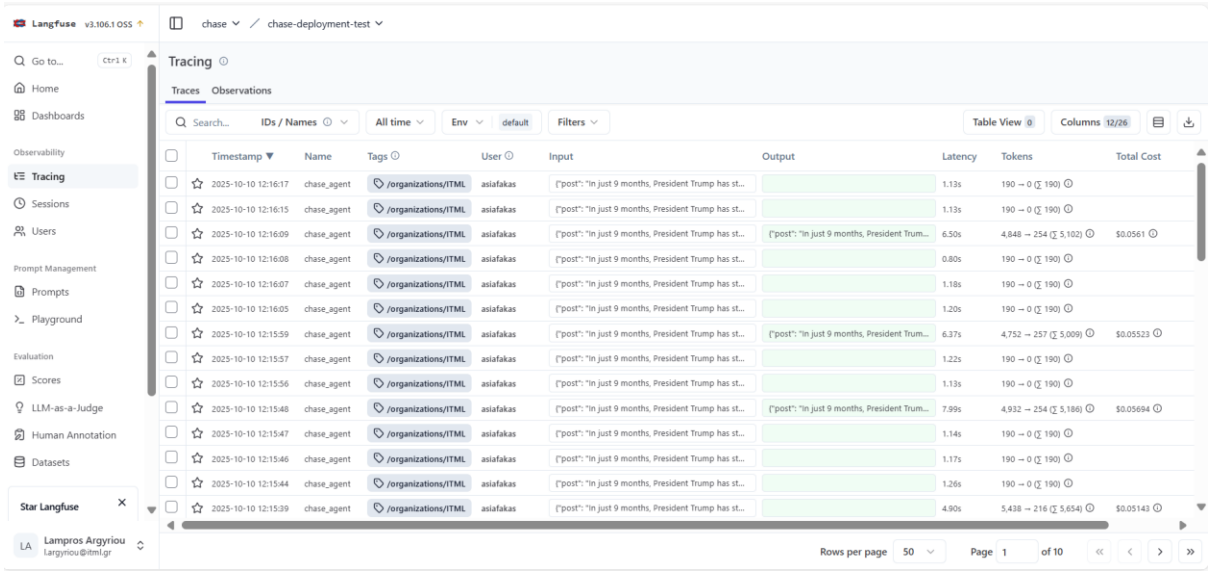

Figura 3: Dashboard di Tracciamento Langfuse che mostra le recenti esecuzioni del workflow agentico CHASE

La **Figura 4** presenta una vista dettagliata di un trace specifico. Il pannello sinistro visualizza il grafo di esecuzione gerarchico, dove ogni nodo rappresenta una chiamata LLM distinta corrispondente agli step nel processo agentico (ad esempio, summarize_post, classify_category, evaluate_classification). Il pannello centrale elenca le metriche dettagliate per ogni chiamata, inclusi latenza e conteggio token, mentre il pannello destro visualizza l'input e l'output strutturati del modello. La parte inferiore della vista include un diagramma di flusso del processo che illustra la progressione logica dall'inizializzazione attraverso gli step di reasoning intermedi fino al completamento.

Attraverso questa integrazione, lo strumento informatico CHASE ora supporta il monitoraggio completo dei processi basati su LLM. Gli sviluppatori possono rivedere le esecuzioni storiche, misurare le tendenze delle prestazioni e correlare prompt specifici con risultati analitici. Questa funzionalità rafforza l'affidabilità e la manutenibilità del componente di text analytics garantendo che tutte le operazioni guidate dal modello siano trasparenti, misurabili e continuamente verificabili durante le fasi di sviluppo e di operazione pilota.

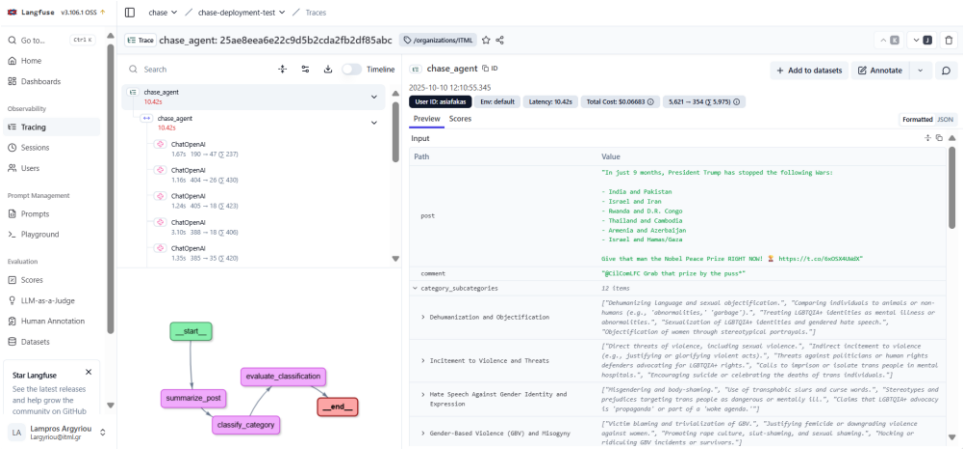

Figura 4: Visualizzazione dettagliata di una traccia Langfuse che rappresenta un'esecuzione del workflow agentico CHASE. La visualizzazione evidenzia il percorso di ragionamento interno del modello e i corrispondenti output analitici

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]

Conclusione

Il completamento di questo deliverable segna la **finalizzazione dello strumento informatico CHASE** come piattaforma completa e completamente integrata. Attraverso fasi di sviluppo successive, ogni componente è stato raffinato, connesso e validato per garantire un'operazione coesa e l'allineamento con gli obiettivi del progetto.

Il risultato è una soluzione stabile e funzionale che combina automazione, interpretabilità e accessibilità all'interno di un framework unificato.

Questo milestone conclude le attività di sviluppo tecnico del WP3 e prepara il terreno per la valutazione e la validazione dello strumento informatico CHASE durante la fase pilota. Il feedback dall'uso nel mondo reale guiderà eventuali perfezionamenti finali e supporterà la sostenibilità a lungo termine della piattaforma oltre la durata del progetto.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or European Commission. Neither the European Union nor the granting authority can be held responsible for them.
[Project Number: 101143159]