

# Designing an ecosystem for fake news detection and enhancing resilience to disinformation in the romanian–russian bilingual space

Angela GLOBALA, Ala GASNAȘ

“Ion Creangă” State Pedagogical University of Chișinău,  
Republic of Moldova

globa.angela@upsc.md, gasnas.ala@upsc.md

**Abstract.** *The RealInfo project introduces an integrated information system designed to detect fake news and enhance media literacy skills within the bilingual Romanian–Russian context of the Republic of Moldova. The system combines a modular microservices architecture with an AI/ML engine trained on local media data through fine-tuning of multilingual BERT models. The analysis pipeline ensures comprehensive text preprocessing, classification, and explainability, while a continuous feedback mechanism enables the ongoing improvement of model performance. Interactive educational components complement the technological infrastructure, providing users with resources, exercises, and micro-courses aimed at fostering critical thinking and informational responsibility. Thus, RealInfo demonstrates how artificial intelligence and media education can be integrated into a unified system that strengthens the information resilience of society.*

**Keywords:** Artificial Intelligence, Disinformation, Fake news, Media literacy, Information resilience.

## 1. Introduction

In a global context marked by the acceleration of information flows and the proliferation of manipulative content, combating disinformation has become a major challenge for society, education, and governance. The impact of fake news goes beyond the media sphere and affects social cohesion, public trust, and democratic processes (Cenușa, 2024). In particular, the Republic of Moldova, due to its bilingual Romanian–Russian cultural and linguistic specificity, faces increased vulnerability to disinformation, which requires the development of intelligent tools adapted to the local context. In the case of fake news, the manipulation of emotions becomes a central strategy of influence, while social networks play a major role in amplifying this process by providing a favorable environment for the rapid circulation of content that provokes intense emotional reactions and reduces users’ critical thinking (Cover, Haw & Thompson, 2022).

Media consumers tend to react more strongly to information with emotional content, especially to negative information, which more easily attracts attention and is perceived as more relevant or credible (Soroka, Fournier & Nir, 2019).

In this context, the MIDECOD project responds to these needs by proposing a pedagogical model that combines artificial intelligence technologies with interactive educational methodologies. The model offers users not only the possibility to verify the authenticity of information, but also the opportunity to learn how to evaluate it critically and rationally, thus contributing to the consolidation of a cognitive and educational infrastructure designed to support the informational resilience of society.

## 2. Technological architecture of the RealInfo system

The central digital component of the pedagogical model is the RealInfo information system, an authentic AI-based platform currently under development, which will be launched as a pilot project in March 2026. The system is capable of automatically detecting false content and providing educational feedback to learners. The authors conducted a relevant comparative analysis of existing information systems specialized in fake news detection, fact-checking, and deepfake identification, highlighting the lack of content specific to the national context, which limits the effectiveness of these systems when addressing new or locally relevant topics. Consequently, the need was identified for the development of a hybrid AI-based disinformation detection system that integrates fake news detection (using BERT (RO, RU) models adapted for the Romanian and Russian languages and trained on Moldovan specific narratives), deepfake detection (with images processed through neural networks combined with forensic techniques), and fact-checking (Globa et al., 2025).

The technological dimension includes the improvement of automatic detection algorithms and the integration of functionalities for identifying manipulative content, including deepfakes, while the educational dimension focuses on the development of competences in critical analysis and applied research. Figure 1 presents the five functional layers that constitute the overall architecture of the RealInfo system:

- The *Frontend Layer* brings together the user-facing interfaces, including the web application developed in React, optimized for desktop access, and the browser extension for real-time analysis. The mobile component was intentionally excluded in order to ensure more efficient resource management and increased system performance.
- The *API Gateway* serves as the central access point of the system and is implemented in FastAPI (Jurafsky & Martin, 2025) to ensure high execution speed, scalability, and automatic documentation.
- The *Backend Microservices* are organized according to the single-responsibility principle, with each microservice performing a clearly defined function: the authentication service manages the OAuth2 protocol; the analysis service processes media content; the reporting service collects and structures user feedback; the educational service manages interactive modules and courses.

- The *AI/ML Engine* is responsible for performing automatic content analysis (Devlin et al., 2019). It integrates a bilingual NLP processor capable of recognizing the language of the text and activating the appropriate deep learning model—BERT-RO for Romanian and BERT-RU for Russian.
- The *Data Storage Layer* is based on a hybrid architecture that combines relational and non-relational databases. Within this structure, PostgreSQL manages structured data, MongoDB is used for storing media metadata, and a vector database such as Pinecone or Weaviate supports semantic search and the identification of similar content.

Given that the project targets interaction with different age groups—from school pupils and university students to older adults, the system is designed to be intuitive and accessible for all users. It is designed with a clear interface, easily interpretable visual elements, a logically structured sequence of steps, and language tailored to the digital competence level of each user group.

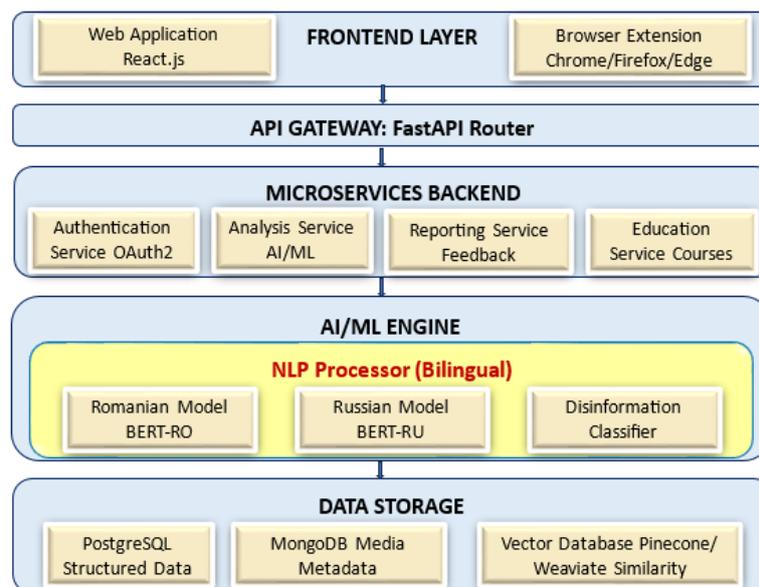


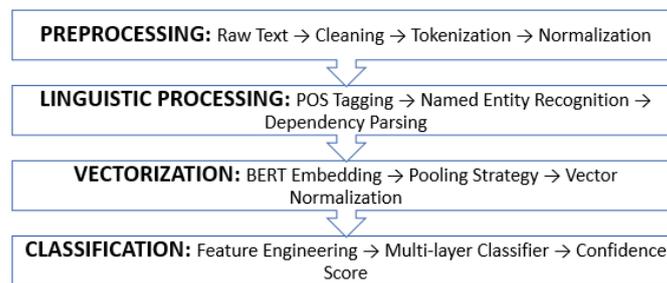
Figure 1. General architecture of the RealInfo system

### 3. Data processing flow

The analysis begins with identifying the language of the text and routing it to an NLP model specific to that language. The NLP processing stage includes tokenization, lemmatization, part-of-speech (POS) tagging, and dependency parsing, all performed in a language-sensitive manner. Subsequently, named entities (persons, organizations, locations, and relevant concepts) are extracted to support contextual understanding. Semantic vectorization converts the text into a dense numerical representation, enabling content evaluation through a hybrid

algorithm that combines AI-generated classification scores with similarity analysis against previously verified content.

**The AI analysis pipeline** of the RealInfo system represents the logical sequence of stages through which textual data are processed, analyzed, and evaluated in terms of content veracity (Figure 2). This flow integrates procedures for linguistic preprocessing, semantic analysis, automatic classification, and contextual validation, all coordinated by a microservice-based orchestration mechanism (Zhou & Zafarani, 2020).



**Figure 2.** AI analysis pipeline

*Preprocessing* serves to standardize the text by removing special characters and applying cleaning rules adapted to Romanian and Russian.

*Linguistic processing* applies part-of-speech (POS) tagging to identify grammatical elements; *Named Entity Recognition (NER)* is used to identify and extract significant entities from the text, such as relevant persons, organizations, and locations, and *Dependency Parsing* analyses syntactic dependencies, highlighting grammatical relations and the hierarchy among the words in the sentence.

*Vectorization* transforms the text into dense numerical representations, using fine-tuned BERT models optimized for the Romanian and Russian languages.

*Classification* applies feature engineering techniques, combining semantic embeddings with additional linguistic features and employing multi-layer classifiers to generate the final decision regarding the veracity of the content.

Through this analysis pipeline, RealInfo ensures a comprehensive, automated, and transparent process for detecting fake news, combining artificial intelligence with human expertise.

**The database architecture** of the RealInfo system was designed to meet the complex requirements of an automatic disinformation analysis system that processes large volumes of heterogeneous data, including texts, media metadata, analytical results, and semantic vectors.

*USERS table.* A distinctive element of this structure is the use of the preferences field, stored in JSONB format, which allows the dynamic recording of individual preferences such as interface language, level of detail of reports, or type of notifications.

The *ANALYSES table* represents the core of the operational process and contains data related to individual analysis requests submitted by users. Each

record includes the original text, the source, the detected language, and associated metadata such as the request date, processing time, and analysis status.

The *ANALYSIS\_RESULTS* table stores the results generated by the AI/ML engine for each analysis performed. It includes credibility scores, the classification label (e.g., true, disinformation, neutral), and the associated explanations derived from the internal interpretation of the model, thereby ensuring a clear and transparent presentation of the results for the user and facilitating subsequent verification of algorithmic decisions.

The *VECTOR\_EMBEDDINGS* table is intended for storing dense numerical representations (embeddings) resulting from the semantic processing of texts by the BERT-RO and BERT-RU models. This table enables the system to automatically compare the analyzed texts with a verified knowledge base and to identify similar or potentially derived content, and the integration of this component contributes directly to the performance of the analysis pipeline and the accuracy of the final results.

By combining these tables, the RealInfo system establishes a direct link between the user, the analysis request, the processing result, and the semantic representation of the content.

**The Browser Extension** of the RealInfo system is designed with a modular architecture structured around four main components: Content Script, Background Worker, Popup Interface, and Overlay UI. Its purpose is to provide direct access to the system's functionalities without requiring navigation to the main web platform, while offering an integrated real-time information verification experience:

- *The Content Script* component runs directly in the context of the web page visited by the user and is responsible for analysing its structure through continuous monitoring of the Document Object Model (DOM). The script automatically identifies content fragments that may represent news articles, posts, or headlines with manipulative potential, using a set of predefined rules and simple semantic filters.
- *The Background Worker* represents the logical core of the extension and is responsible for managing communication between the interface component and the server infrastructure. It maintains a secure connection with the RealInfo API, forwarding analysis requests from the Content Script and receiving the results provided by the processing pipeline.
- *The Popup Interface* constitutes the visible element of the extension and provides users with an intuitive control point over the verification process.
- *The Overlay UI* allows the analysis results to be displayed directly on the page in the form of warning banners, coloured highlights, or floating windows. This component provides immediate feedback that is visible in the reading context without interrupting the user experience.

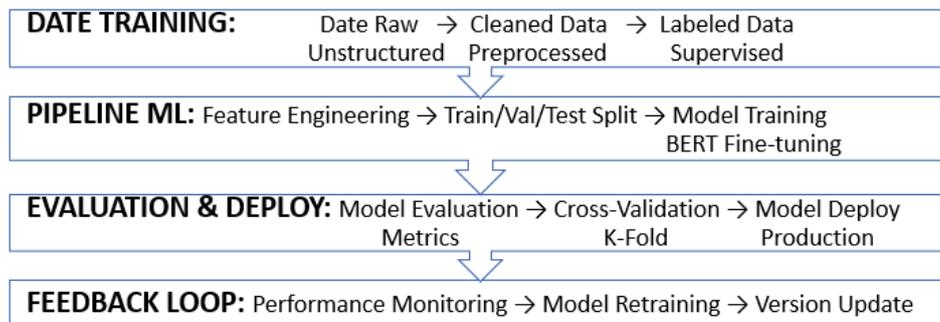
Through the proper interaction of its components, the browser extension combines advanced technology with accessibility and educational impact, while its modular structure allows the addition of new features, such as clickbait

detection or integration with educational platforms, maintaining a safe and smooth user experience.

#### 4. Machine learning pipeline

The AI models within the RealInfo system are developed through a rigorous iterative pipeline (including data collection, cleaning, labeling, training, and validation), with the aim of detecting disinformation adapted to the linguistic and media context of the Republic of Moldova (Figure 3).

The training *data* come from a mixed corpus of media sources from the Republic of Moldova, reflecting the thematic, stylistic, and ideological diversity of the local information space and enabling the models to learn the specific features of Romanian and Russian, including forms of code-mixing (Nițu & Dascălu, 2024).



**Figure 3.** Machine Learning Pipeline

The *machine learning pipeline* (ML pipeline) is modular in structure and integrates components for feature engineering, fine-tuning, and adaptive evaluation.

*Model evaluation* is carried out on a separate dataset that is not used during training, by applying metrics specific to disinformation detection. These include: Precision – the proportion of correct labels out of all positive predictions; Recall – the ability of the model to detect all existing disinformation cases; and F1-Score – the balance between precision and recall.

The operational *feedback loop* continuously monitors model performance in the production environment. The results obtained from real-world use – including corrections provided by experts or trained users are collected and integrated into the periodic retraining process, thereby continuously improving the accuracy and adaptability of the models.

Through this training and evaluation architecture, the RealInfo system goes beyond the role of a simple automatic detector, becoming an adaptive learning ecosystem grounded in the collaboration between artificial intelligence and human expertise. The proposed model ensures not only algorithmic performance but also scientific and ethical sustainability in the process of countering disinformation.

The educational component of the RealInfo system extends the technological functionality of the system through a pedagogical dimension intended to support education for informational resilience, addressing diverse age groups and different levels of digital competence.

## 5. Conclusions

The RealInfo system demonstrates the potential of integrating artificial intelligence with media education into a unified approach capable of addressing the complex challenges of disinformation within the socio-linguistic context specific to the Republic of Moldova. The system creates favorable premises for the further development of research focused on multilingual and multimodal disinformation detection models, as well as for the integration of the RealInfo platform into educational programs and public media literacy initiatives. Thus, RealInfo is not only a technological innovation but also a concrete step towards a more informed, critical, and resilient society in the face of informational manipulation.

## Acknowledgements

This article is produced as part of the scientific research project „The methodology of development of an Integrated Model for the Detection and Counteraction of Disinformation in the Republic of Moldova: A Technological and Socio-Educational Approach” (MIDECOD), with the code 25.80012.0807.65SE.

## REFERENCES

- Cenușa, D. (2024) Russian Narratives and Resilience to Disinformation in Moldova. *Russian Analytical Digest*. 313, 6-10. <https://doi.org/10.3929/ethz-b-000673162>.
- Cover, R., Haw, A. & Thompson, J. D. (2022) *Fake News in Digital Cultures: Technology, Populism and Digital Misinformation*. Emerald Publishing, Bingley Book March.
- Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the NAACL-HLT, June 2-7, 2019, Minneapolis, Minnesota USA*. Association for Computational Linguistics. pp. 4171–4186. doi: <https://doi.org/10.1073/pnas.1908369116>.
- Globa, A., Gasnaș, A., Andoni D. & Eni, N. (2025) Dezinformarea ca fenomen național complex: perspective societale și informatice. *Acta et Commentationes. Științe ale Educației*. 3(41), 44-56. doi: <https://doi.org/10.36120/2587-3636>.

Jurafsky, D., Martin, J. H. (2025) *Speech and Language Processing (Third Edition draft)*. [https://web.stanford.edu/~jurafsky/slp3/ed3book\\_aug25.pdf](https://web.stanford.edu/~jurafsky/slp3/ed3book_aug25.pdf) [Accessed: 12<sup>th</sup> October 2025].

Nițu, M. & Dascălu, M. (2024) Natural Language Processing Tools for Romanian—Going Beyond a Low-Resource Language. *Interaction Design and Architecture(s) Journal - IxD&A*. 60, 7–26. doi: <https://doi.org/10.55612/s-5002-060-001sp>.

Soroka, S., Fournier, P. & Nir, L. (2019) Cross-national evidence of a negativity bias in psychophysiological reactions to news. *Proceedings of the National Academy of Sciences*. 116(38), 201908369. pp. 18888 – 18892.

Zhou, X., Zafarani, R.-A. (2020) Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Computing Surveys (CSUR)*. 53(5), 1–40. <https://doi.org/10.1145/3395046>.