CZU: 004.056:37.01 DOI: 10.36120/2587-3636.v44i2.51-60

IDENTIFYING DISINFORMATION NARRATIVES: A SYSTEMIC APPROACH

Dumitru ANDONI, master's student*, https://orcid.org/0009-0000-3561-4975
Nicu ENI, master's student*, https://orcid.org/0009-0009-0765-422X
Angela GLOBA, PhD, Associate Professor
https://orcid.org/0000-0002-2653-0320
Ala GASNAŞ, PhD, Associate Professor
https://orcid.org/0000-0002-7174-7027
* Master's Program in Information Technologies in Education
"Ion Creangă" State Pedagogical University of Chisinau

Abstract. Disinformation narratives represent one of the most challenging threats to the contemporary information ecosystem, particularly in the Republic of Moldova, which has been the target of coordinated influence campaigns documented by DFRLab, Recorded Future, and EUvsDisinfo. Unlike isolated fake news, narratives operate through repetition and coordinated distribution across multiple channels simultaneously, making them difficult to detect through traditional methods. This paper describes the architecture of the RealInfo system, an automatic narrative detection platform built on a hybrid three-layer pipeline: local linguistic analysis based on rhetorical pattern matching, emotional classification through pre-trained models, and narrative analysis through RealInfo LLM - a 14-billion-parameter model running locally on a distributed GPU cluster. The system integrates a semantic similarity module for detecting coordinated narratives across multiple sources. The proposed architecture demonstrates that a multi-layer hybrid approach is superior to any isolated method, while providing complete explainability of the decision.
Keywords: disinformation, narratives, NLP, LLM, automatic detection, Republic of Moldova, propaganda.

IDENTIFICAREA NARATIVELOR DE DEZINFORMARE: O ABORDARE SISTEMICĂ

Rezumat. Republica Moldova se confruntă cu campanii de dezinformare sistematice, în care narativele false nu apar ca afirmații izolate, dar ca structuri discursive repetate coordonat pe mai multe platforme simultan. Articolul prezintă RealInfo, o platformă de detecție automată construită pe trei straturi complementare: analiză lingvistică locală, clasificare emoțională prin modele pre-antrenate și analiză narativă prin RealInfo LLM cu 14 miliarde de parametri, rulat local pe cluster GPU. Un modul suplimentar detectează coordonarea narativelor între surse multiple prin similaritate semantică. Abordarea hibridă multi-strat depășește orice metodă izolată și produce o explicație completă a fiecărei decizii.
Cuvinte-cheie: dezinformare, narative, NLP, LLM, detecție automată, Republica Moldova, propagandă.

1. Introduction

The Republic of Moldova has become one of the main targets of systematic disinformation. In the run-up to the 2024 elections, a single network of 253 coordinated accounts generated over 55 million views in just a few weeks [1], while Recorded Future documented at least six distinct Russian operations that simultaneously targeted the Moldovan information space [2]. The narratives circulated on Telegram (38%), Facebook (27%), clone websites (18%), and WhatsApp (11%). Faced with this reality, fact-checking individual claims is no longer sufficient. A narrative does not operate through a single false statement, but through a story repeated in a coordinated manner across multiple channels. It can take up real facts and place them within a distorted interpretive frame - narrative framing - which makes it far harder to counter than an isolated piece of fake news.

This article describes the identification of narratives with RealInfo, a system built to address this pressing need to counter disinformation [3,4].

2. The disinformation narrative: definition and typology

2.1. *Definition and Conceptual Framework*

The literature distinguishes three types of problematic content according to intent. Misinformation is false information distributed without harmful intent. Disinformation involves a deliberate intent to manipulate. The disinformation narrative is a structure systematically repeated across multiple channels, which organizes false or distorted messages into a coherent story with a precise strategic purpose [5].

Miskimmon, O'Loughlin, and Roselle define strategic narratives as instruments through which political actors shape the shared understanding of reality [6]. Their effectiveness stems precisely from the fact that they do not need to lie entirely; they can take up real facts and place them within a deformed interpretive frame (narrative framing), steering the reader's conclusion in the desired direction [7].

2.2. *Rhetorical techniques of manipulation*

Studying disinformation content in the Moldovan space, researchers identified nine rhetorical techniques that recur consistently, also found in the international literature on computational propaganda. Each of these techniques operates by exploiting a specific psychological mechanism, and the RealInfo system detects them automatically, as shown in Table 1.

Table 1. Manipulation techniques detected by the RealInfo system [8]

Technique	Description
Appeal to fear	Constructing or amplifying a fear in order to trigger a rapid emotional reaction, ahead of any rational evaluation
False dilemma	Presenting a complex situation as a binary choice, deliberately excluding any middle ground
Ad hominem	Discrediting the source or person instead of countering the argument
Whataboutism	Diverting attention from one issue by invoking another, without answering the original question
Cherry-picking	Selecting real data that supports a desired conclusion while systematically ignoring contrary data
Unverified rumor	Claims attributed to undefined insider sources, impossible to verify
Alarmist headline	The deliberate gap between a sensational headline and the neutral content of the article
Hasty generalization	Extrapolating an isolated incident to the level of a general phenomenon
False authority sources	Attributing statements to fictitious experts or figures who never made the claims in question

2.3. *The life cycle of a narrative*

According to [1], the life cycle of narratives active in the Republic of Moldova has five stages, which explain why detection must track the coordinated pattern rather than the individual article. The cycle begins with the creation of distorted content for maximum emotional impact. This is followed by amplification through coordinated networks of accounts, which create the appearance of organic virality. Real users redistribute the content in the virality stage, followed by normalization, that is, the narrative becomes part of everyday discourse, treated as a self-evident fact. The final impact is measurable in distorted perceptions, voting decisions, and the erosion of trust in institutions [9].

2.4. *The taxonomy of dominant narratives in the Republic of Moldova (2024–2025)*

Based on reports from EUvsDisinfo, DFRLab, WatchDog.md, and Recorded Future, four main categories of narratives active in the Republic of Moldova in 2024–2025 are identified:

- *Geopolitical narratives*, constructed through the false dilemma technique, for example, „Moldova will lose its sovereignty after joining the EU" or „The West is preparing Moldova for war" [10].
- *Economic narratives*, anchored in citizens' real experiences but falsely presented through various manipulation techniques. For example, through the cherry-picking technique, the gas crisis of January 2025 is attributed exclusively to pro-European decisions, omitting the Gazprom context [11].
- *Identity narratives*, which exploit linguistic and religious tensions, including through video deepfakes, etc. [10].
- *Electoral narratives*, which seek to erode trust in the democratic process and in state institutions. For example, „The elections are rigged, voting changes nothing, all candidates are corrupt."

3. The architecture of the RealInfo system

The RealInfo system is designed as a systemic response, not as a tool for individual verification. It operates through a detection pipeline of three modules working in parallel, with all LLM inference running locally on a cluster of three nodes interconnected through the RPC protocol of llama.cpp, without external dependencies [3,12].

In the verification process with RealInfo, each article goes through six asynchronous stages:

(1) extraction of the text from the HTML;

(2) local linguistic analysis, under 50ms, without external calls;

(3) classification of emotional manipulation;

(4) narrative analysis through the RealInfo LLM;

(5) searching the top 20 similar articles and querying the Brave Search API;

(6) aggregating the signals and issuing the verdict.

4. Layer 1. Local linguistic analysis

4.1. *Detection of narrative families*

The *advanced_patterns.py* module compares each text against an ontology of patterns organized into six narrative families (Table 2), each with dozens of key expressions in Romanian and Russian. The *weight* functions similarly to a weighting coefficient: the more specific an expression is to manipulation, that is, the more rarely it appears in neutral texts but frequently in propagandistic/manipulative ones, the higher its weight and the more pronounced its influence on the final score.

Table 2. Narrative families detected by *advanced_patterns.py*, including weights [13]

Narrative family	Examples of detected expressions (RO/RU)	Weight
Pro-Russian influence	Slavic brother, aggressive NATO, corrupt West, unjust sanctions	2.5
Anti-Western or anti-EU	The EU is destroying us, traditional values under attack, the globalist agenda	2.5
Panic instigator	The end is near, imminent catastrophe, danger is lurking	2.0
Anti-institution	The government is lying, the state destroyed, corrupt judges	2.0
Disinformation	The hidden truth, what they want to hide, the silent conspiracy	2.0
Sensationalism	INCREDIBLE!, BREAKING!, EXCLUSIVE!, URGENT!, VIRAL!	1.5

The *propagandaScore* is calculated according to formula (1) and includes a *co-occurrence factor*, which measures the extent to which several propaganda techniques appear simultaneously in the same text, amplifying this score:

$$propagandaScore = \min\left(100, \sum_i (g_i \cdot f_i) \cdot F_c\right) \quad (1)$$

where g_i is the weight of narrative family i; f_i is the frequency of occurrence of narrative family i; F_c is the co-occurrence factor.

The co-occurrence factor takes the values 1.0 (if 1-2 narrative families are detected), 1.5 (3 families), and 2.0 (4+ families). Scores below 30 indicate normal content; 30-60 a moderate signal; above 60 a high probability of deliberate propaganda.

The basic idea is that deliberate propaganda does not use a single isolated technique; it intentionally combines several techniques at once, for example, emotional appeal + negative labeling + omission of facts. This simultaneous combination is a stronger signal of propagandistic intent than the isolated occurrence of a single technique.

4.2. *Sensationalism and Headline-Text asymmetry*

The *nlp.py* module measures the headline-text gap through two indicators. The *sensationalismScore* quantifies the density of hyperbolic language (a style of expression that intentionally exaggerates reality in order to create a stronger emotional impact than the facts themselves would justify), with a double factor (*title_factor* = 2.0) for the terms in the headline, according to the formula:

$$sensationalismScore = \min\left(100, \frac{NTS}{N_c} \cdot F_t \cdot 1000\right) \quad (2)$$

where NTS is the number of sensational terms; N_c is the number of words; F_t is the *title_factor*.

Headlines are where hyperbolic language is most frequently used, because the headline is the first, and sometimes the only, thing a person reads. A sensational term in the headline has a much greater manipulative impact than the same term in the body of the article, so its contribution to the score is doubled.

The *asymmetryScore* compares the densities in the headline and the body. A headline with a score greater than or equal to 60 associated with a body scoring greater than or equal to 20 activates the CLICKBAIT flag, corresponding to the alarmist headline technique documented by Rashkin et al. [14]. A large asymmetry means that the headline „*promises*" more than the body of the article delivers. In broad terms, this is in fact exactly the definition of clickbait.

5. Layer 2: Emotional classification of the narrative

Manipulation narratives also operate through emotional mechanisms, where fear and anger block critical thinking and drive sharing without verification. People perceive emotionally negative information as more urgent and more credible, which explains why false information spreads six times faster than true information [15].

The API integration module *huggingface_client.py* sends the text to an emotional classification model and obtains normalized scores [0, 1] for four dimensions, correlated with the narrative types in layer 1 (Table 3), adjusting the *propagandaScore* according to the intensity of the signal.

Table 3. Correlation of emotional dimensions with narrative types and impact on the score

Emotional dimension	Threshold	Correlation with narrative type	*propagandaScore* adjustment
Anger	> 0.5	Conflict narratives: external enemies, NATO, EU	+8 points
Fear	> 0.4	Existential threat, catastrophe narratives	+10 points
Anger + fear (combined)	> 0.8	Intense emotional manipulation – strong signal	+15 points
Sadness	> 0.6	Victimization and collective suffering narratives	+5 points
Joy (isolated)	> 0.7	Absence of negative signals – possibly neutral content	−5 points

6. Layer 3: Narrative analysis through the RealInfo LLM

6.1. *Model configuration and the inference mechanism*

The first two layers identify signals, while the third layer interprets them. The RealInfo LLM receives the complete text together with all the previous scores and produces the narrative analysis in natural language. Providing the intermediate reasoning follows the chain-of-thought prompting principle, which significantly improves the accuracy of the final decision [16].

The prompt in *chat_service.py* guides the model through five operations: identifying the anchor claim, verifying citable sources, comparing the headline with the body, classifying the rhetorical techniques, and issuing the verdict with quotations. *The temperature* of 0.2 is practically the industry standard for analysis, classification, and information extraction applications, favoring consistency over creativity, which is essential for factual analysis, that is, the model almost always chooses the word with the highest probability: predictable, consistent, conservative responses. In other words, there is no room for creative interpretations of the facts (*temperature* being a parameter that controls how predictable or creative an LLM's response is).

6.2. *The structured output of the model*

The model returns a structured JSON (JavaScript Object Notation) (Table 4), which ensures integration with the rest of the system and the complete auditing of each decision.

Table 4. The JSON fields returned by the RealInfo LLM and their meaning

JSON field	Type	Content and meaning
narratives	string[]	The list of detected narratives, for example, „pro-Russian influence”, „incitement to fear”
narrativeFraming	string	The dominant narrative, for example, „pro-Russian influence”
credibilityScore	int (0–100)	The credibility score assessed by the LLM, which enters the final aggregation formula
llmExplanation	string	A natural-language explanation of the identified manipulation mechanism
llmRedFlags	string[]	The concrete red flags, for example, „Unspecified sources”, „Extreme generalizations”
manipulativePatterns	string[]	The detected manipulation techniques, for example, „appeal to fear”, „false dilemma”
verdict	string	The final verdict: CREDIBLE, PARTIALLY CREDIBLE, POSSIBLY FALSE, UNVERIFIABLE

7. Detecting narratives through semantic similarity

A single suspicious article may be an editorial error. Ten articles with the same message, published within a few days by seemingly independent sources, constitute a coordinated campaign. The *search_service.py* module detects this difference, transforming RealInfo from a fake news detector into a narrative detector [3, 17].

Each article is converted into a *semantic embedding*. *The semantic embedding* is a numerical representation of the meaning of a text, that is, the transformation of words or phrases into coordinates in a mathematical space, so that texts with similar meaning are close to one another, while texts with different meaning are far apart. In other words, an entire article is converted into a vector of numbers, normalized, usually 384, 768, or 1536 values.

When a new article is analyzed, its vector is compared with all the vectors in the database through cosine similarity (Figure 1), the main method for comparing two embedding vectors, returning the top 20 similar articles. The system detects the same narrative regardless of language, which is essential for the bilingual Moldovan context [17].

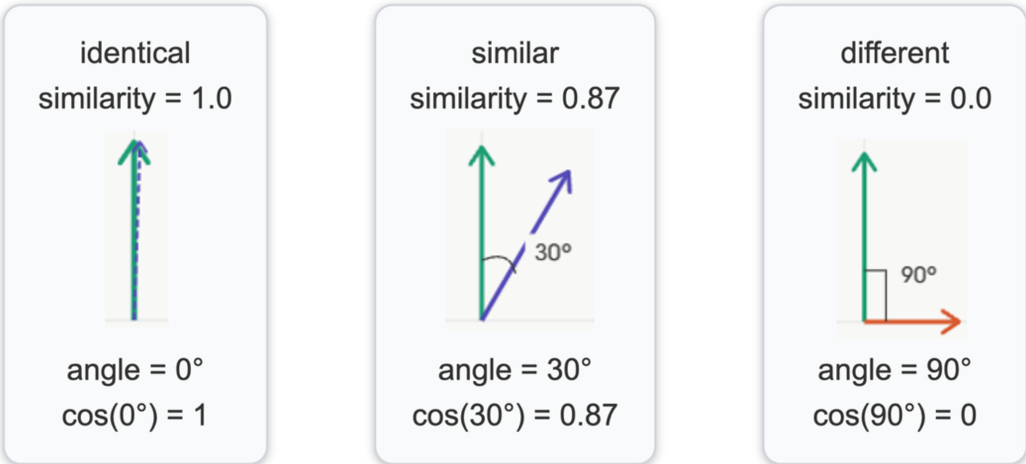

Figure 1. Comparing two embedding vectors through cosine similarity

How is cosine similarity applied in the RealInfo system? When a new article enters the system, its embedding is compared with all the vectors in the database. The result is a score between 0 and 1:

- 0.95 – 1.0: articles almost identical in subject;
- 0.75 – 0.94: articles about the same subject but with a different point of view;
- 0.50 – 0.74: partially related subjects;
- below 0.50: different subjects.

The system can thus identify whether a new article propagates the same type of narrative as articles already classified as propaganda, even if it uses completely different words. A narrative-coordinate is confirmed when at least three articles with a cosine score > 0.85, from different sources, appear within a maximum of 72 hours. The 0.85 threshold separates coordinated campaigns from parallel journalistic coverage of the same event, which usually generates scores between 0.60 and 0.80 [17].

8. Aggregating the signals and the final verdict

All the signals are brought together in the *credibilityScore*, calculated through formula (3). The logic is penalizing: each article starts from 100 and loses points proportionally to the narrative signals detected.

$$credibilityScore = 100 - (propagandaScore \cdot 0{,}30) -$$

$$-(sensationalismScore \cdot 0{,}15) - \big((anger + fear) \cdot 0{,}20\big) - \big(nr_{llmRedFlags} \cdot 4\big) -$$
$$-(asimetryScore \cdot 0{,}10) + nr_{cited\_sourses} \cdot 3) \qquad (3)$$

The score is calculated on a scale from 0 to 100. Very high scores (above 70) are capped at a maximum of 15 points. The reason is that an extremely high score does not necessarily guarantee that all the answers are completely valid; it only indicates that an absolute minimum of the honesty standard has been reached. The score produces the verdict:

- ≥ 70 → CREDIBLE; few signals reported, a relatively stable and consistent profile;
- 40 – 69 → PARTIALLY CREDIBLE; moderate signals;
- < 40 → POSSIBLY FALSE; strong signals of a manipulative narrative; the clearest configuration of a confirmed narrative combines: *credibilityScore* < 40, a cluster of at least seven articles with a score > 0.85, and external sources that contradict the anchor claim, generating the verdict: POSSIBLY FALSE with an active narrative [3];
- UNVERIFIABLE; insufficient information to issue a verdict.

9. Conclusions

No isolated method detects propaganda narratives in their full complexity. Local NLP analysis offers speed but remains vulnerable to paraphrasing. Emotional classification adds a relevant psychological dimension, yet cannot identify the narrative itself. Large language models cover both limitations but cannot on their own detect coordination between sources. Only the integration of the three layers together with the semantic similarity module ensures both accuracy and explanatory transparency, which is an essential condition for the system to be accepted by decision-makers.

The campaigns documented in the Republic of Moldova are systematic and well-funded [2]. The technical response must reach at least the same level of sophistication as the attack it counters.

Future development directions target three priorities: expanding the ontology of patterns through unsupervised clustering, integrating the analysis of distribution networks on Telegram and Facebook, and fine-tuning the language model on a corpus specific to the Republic of Moldova, with the aim of reducing the processing latency to under three seconds per article.

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.

Bibliography

1. CALMIS, D., OLARI, V., GIGITASHVILI, G. Malign interference in Moldova ahead of presidential election and European referendum. In: *Digital Forensic*

Research Lab (DFRLab), October 18, 2024. Available: https://dfrlab.org/2024/10/18/malign-interference-moldova/ [accessed 2026-01-15].

2. RECORDED FUTURE. INSIKT GROUP. *Russian influence assets converge on Moldovan elections.* 2025. Available: https://www.recordedfuture.com/research/russian-influence-assets-converge-on-moldovan-elections [accessed 2026-01-15].
3. GLOBA, A., GASNAȘ, A. Designing an ecosystem for fake news detection and enhancing resilience to disinformation in the romanian–russian bilingual space. In: *International Conference on Virtual Learning*, ISSN 2971-9291, ISSN-L 1844-8933, vol. 21, pp. 555- 562, 2026. https://doi.org/10.58503/icvl-v21y202646
4. GLOBA A., GASNAȘ, A. Analyzung of the disinformation phenomennon in the Republic of Moldova from a social and technological perspective. In: *Proceedings of the 32nd International Conference on Applied and Industrial Mathematics. Communications in Education*. September 18-21, 2025, Bucharest, România. pp. 142-153.
5. WARDLE, C.; DERAKHSHAN, H. Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking. Strasbourg: Council of Europe, DGI(2017)09, 2017.
6. MISKIMMON, A., O'LOUGHLIN, B., ROSELLE, L. Strategic Narratives: Communication Power and the New World Order. New York: Routledge, 2013. 240 p. ISBN: 978-0-415-71760-1. DOI: 10.4324/9781315871264
7. POMERANTSEV, P. This Is Not Propaganda: Adventures in the War Against Reality. New York: PublicAffairs, 2019. 256 p.
8. GASNAȘ, A., GLOBA, A., ANDONI, D., ENI, N. Mecanisme de dezinformare și impactul acestora asupra cetățenilor Republicii Moldova. In: *Acta et Commentationes. Științe ale Educației*, nr. 4(42), 2025, pp.39-51. ISSN 1857-0623, E-ISSN 2587-3636. https://doi.org/10.36120/2587-3636.v42i4
9. OSCE. Moldova's parliamentary elections were competitive but campaign marred by cyberattacks, illegal funding and disinformation, 2025. Available: https://www.oscepa.org [accessed 2026-02-01].
10. EUvsDisinfo. *Rapoarte de monitorizare a dezinformării în spațiul moldovenesc, 2024–2025*. Available: https://euvsdisinfo.eu [accessed 2026-02-04].
11. THE KYIV INDEPENDENT. *Russia fueling Transnistria energy crisis for propaganda, destabilization, Moldova says.* Ianuarie 2025. Available: https://kyivindependent.com/russia-is-fueling-transnistria-energy-crisis-for-propaganda-destabilization-moldova-says/ [accessed 2026-02-04].
12. GASNAS, A., GLOBA, A., ANDONI, D., ENI, N. Model conceptual și arhitectura tehnologica pentru sisteme de recunoaștere automata a știrilor false. În: *Proceedings of the fifth International Scientific Conference „Inter /transdisciplinary approaches*

in the teaching of the real sciences, (STEAM concept)" dedicated to the 95th anniversary of the founding of the Faculty of Physics, Mathematics and Information technologies. Chisinau, Republic of Moldova, October 31 – November 01, 2025. p. 345-349.
13. Russian Foreign Propaganda in Occupied Ukraine, Moldova, and Belarus. In: *Russian Analytical Digest*, nr. 313, 2024. Zurich: ETH. https://doi.org/https://doi.org/10.3929/ethz-b-000673162 [accessed 2026-02-12].
14. RASHKIN, H. et all. Truth of Varying Shades: Analyzing Language in Fake News and Political Fact-Checking. In: *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing,* p. 2931–2937, Copenhagen, Denmark. Association for Computational Linguistics. DOI:10.18653/v1/D17-1317 [accessed 2026-03-05].
15. VOSOUGHI, S., ROY, D., ARAL, S. The Spread of True and False News Online. *Science*. 2018, 359(6380), pp. 1146–1151. https://doi.org/10.1126/science.aap9559
16. WEI, J. et all. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In: *Advances in Neural Information Processing Systems* (NeurIPS). 2022, vol. 35, pp. 24824–24837. https://doi.org/10.48550/arXiv.2201.11903
17. ZHOU, X., ZAFARANI, R. A. Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. In: *ACM Computing Surveys*. 2020, 53(5), pp. 1–40. https://doi.org/10.1145/3395046.

Received / *Recepționat:* 22.05.2026
Accepted / *Acceptat*: 29.06.2026
Email: globa.angela@upsc.md
gasnas.ala@upsc.md
andonidumitru1@gmail.com
eni.nicu99@gmail.com