

MODEL OF COMMUNICATIVE IMPACT IN DISINFORMATION MESSAGES BASED ON SPEECH ACT THEORY AND ARTIFICIAL INTELLIGENCE TOOLS

Oleksandr Cherep¹, Yuliia Kaliuzhna², Svitlana Markova³

Abstract. The present article explores the communicative impact of disinformation messages by combining the theoretical framework of speech act theory with the analytical capabilities of artificial intelligence (AI). The study focuses on the structural and pragmatic organisation of disinformation, viewed as a communicative act functioning simultaneously at the locutionary, illocutionary, and perlocutionary levels. The role of emotionally charged and evaluative lexical elements is of particular interest, insofar as these elements function as instruments of psychological influence, thereby shaping the perceptions, emotions, and behaviour of audiences. The objective of the research is to develop a methodological approach for identifying and classifying manipulative intentions within disinformation messages. The present study seeks to integrate speech act analysis with natural language processing (NLP) techniques and large language models (LLMs). The objective is to uncover factual distortions, as well as concealed rhetorical strategies, emotional framing and subtle linguistic manipulations. The methodology underpinning this study is predicated upon a multi-level analytical model that interprets disinformation as a structured communicative act. The framework comprises several stages: compiling a corpus of authentic disinformation texts from social media and propaganda sources; preprocessing linguistic data through tokenisation, lemmatisation, POS-tagging, and syntactic parsing; conducting locutionary analysis of propositional structures and semantic networks; performing illocutionary classification of speech acts (assertives, directives, commissives, expressives) using supervised machine learning; and carrying out perlocutionary analysis to detect sentiment, classify emotions, and identify expressive linguistic devices such as metaphor, hyperbole, epithet, and anaphora. The model is validated using AI tools (BERT, RoBERTa and GPT-based systems), combined with human-in-the-loop verification via fact-checking datasets. The findings show that disinformation relies on a complex interplay of linguistic mechanisms intended to create persuasive and manipulative content. Integrating speech act theory with AI-based linguistic analysis has proven effective in detecting emotional tone, communicative intent and manipulative structures across large volumes of data. In conclusion, this study demonstrates that combatting manipulative communication necessitates a broader approach than mere fact-checking, encompassing an analysis of the pragmatic and emotional aspects of disinformation. The proposed model offers a scalable, systematic approach to detection, thereby enhancing cognitive resilience and improving information security in today's digital landscape.

Keywords: disinformation, speech act theory, artificial intelligence, NLP, language models, emotional manipulation, illocutionary intention.

JEL Classification: C88, D83, O33

1. Introduction

The modern information environment is characterised by the rapid spread of emotional and anxiety-provoking disinformation that distorts perceptions of events. The main challenge lies in the difficulty of detecting

such messages due to their semantic and emotional adaptability. Effective analysis of disinformation requires an in-depth understanding of its linguistic structure, emotional tone and the impact it has on the recipient's consciousness. In this context, artificial

¹ Zaporizhzhia National University, Ukraine (*corresponding author*)

E-mail: cherep2508@gmail.com

ORCID: <https://orcid.org/0000-0002-3098-0105>

² Zaporizhzhia National University, Ukraine

E-mail: kalyuzhnaya.ju@gmail.com

ORCID: <https://orcid.org/0000-0002-3335-6551>

³ Zaporizhzhia National University, Ukraine

E-mail: masvvi@outlook.com

ORCID: <https://orcid.org/0000-0003-0675-0235>



This is an Open Access article, distributed under the terms
of the Creative Commons Attribution CC BY 4.0

intelligence plays a particularly important role, as it can reveal hidden linguistic patterns and manipulative intentions.

Recent advances in natural language processing (NLP) and machine learning mean that artificial intelligence (AI) can now perform increasingly complex language-related tasks, including sentiment detection, discourse analysis and text classification. These capabilities are particularly valuable in the context of disinformation, where the impact often depends not only on factual inaccuracy, but also on emotional framing, subtle rhetorical manipulation and the strategic use of language to achieve perlocutionary effects.

Disinformation messages are not merely isolated texts; they are deliberate communicative acts that influence the cognitive, emotional and behavioural states of their intended recipients. To understand the structure and intent of these messages, more than a surface-level content analysis is required. This involves unpacking the communicative functions embedded within the language, which are well described by speech act theory. This theory provides a framework for analysing not only what is said (locution), but also why and with what effect it is said (illocution and perlocution, respectively).

Concurrently, artificial intelligence (AI) tools, including large language models (LLMs), offer scalable, data-driven methodologies for the identification of these communicative features across voluminous content repositories. The combination of speech act analysis and AI has been shown to enhance the precision and comprehensiveness of disinformation identification. This synergy forms the foundation of the model proposed in this study.

Speech act theory (J. Austin, J. Searle) established the basis for comprehending how language can function not only as a medium for the transmission of information, but also as an agent for the modification of behaviour (Austin, 1986). In the context of disinformation, it is important to consider the three levels of a speech act: locution (the actual utterance), illocution (the author's intention), and perlocution (the real impact on the recipient). However, current studies of disinformation still lack a systematic analysis of the emotional tone of such messages at these levels. Artificial intelligence tools, particularly NLP models, offer new possibilities for identifying this type of influence.

Recent studies have confirmed that large language models (LLMs), including GPT-based architectures, are capable not only of analysing but also of generating emotionally charged disinformation messages. It has been demonstrated that the tone of prompts exerts a significant influence on the degree of manipulateness exhibited by models' responses, thereby augmenting the capacity of AI to replicate destructive communicative patterns (Vinay, Oehmichen, Agirre, Davis, 2024).

Moreover, a significant research trajectory pertains to the analysis of disinformation networks and the identification of pivotal actors in the propagation of fraudulent messages, predicated on linguistic and behavioural patterns. This enables a comprehensive assessment of disinformation impact not only at the content level but also within the structural dynamics of the information space (Smith, Kao, Mackin, Shah, Simek, Rubin, 2020).

Moreover, recent reviews of the field underscore the theoretical and practical challenges in detecting fake news, including the limitations of existing datasets, methodological constraints, and the necessity of forward-looking research agendas (Harris, Hadi, Ahmad, Alshara, 2024). Concurrently, scholarly debate persists regarding the true extent of generative AI's threat within the context of the misinformation landscape. Some contend that concerns about its impact may be exaggerated (Simon, Altay, Mercier, 2023), while others underscore the heightened risks for journalism, public trust, and democratic discourse (Bell, 2023). Research in a complementary field has demonstrated that AI systems have the capacity to produce persuasive propaganda, thus underscoring the dual-use nature of these technologies in the information domain (Goldstein, Chao, Grossman, Stamos, Tomz, 2023).

The objective of this study is to formulate a methodological approach for the identification of the emotional and manipulative structure of disinformation messages, utilising speech act theory and artificial intelligence tools.

2. Research Methodology

The methodological basis of this study is the integration of speech act theory with artificial intelligence (AI) tools for the purpose of modelling the communicative structure of disinformation messages. The approach is grounded in pragmatic linguistics, where disinformation is interpreted as a structured communicative act comprising three interrelated levels: locutionary, illocutionary, and perlocutionary. Each level is characterised by specific linguistic features and communicative functions.

The research design includes the following key stages:

1. Theoretical framework development. A classification of speech acts based on the works of J. Austin and J. Searle is adapted for analysing disinformation messages. Assertions, directives, commissives and expressions are used as the core illocutionary categories, while verbal expressive devices such as epithets, metaphors, anaphora and hyperbole serve as indicators of the perlocutionary effect.

2. Corpus formation involves compiling a set of real-world disinformation messages from open sources,

such as social media, propaganda sites and messaging platforms. Each message is then manually and/or semi-automatically annotated to identify pragmatic markers and emotional content.

3. Text preprocessing. NLP tools are used for tokenisation, lemmatisation, part-of-speech tagging and syntactic parsing. Stop words and other irrelevant elements are removed to ensure clean input for semantic analysis.

4. Locutionary analysis is the process of identifying the basic propositional structure of messages through the analysis of reference (nouns), predication (verbs) and grammatical cohesion. Semantic networks are utilised for the extraction of central entities and actions.

5. Illocutionary analysis involves the use of modal verbs, discourse markers, syntactic structures, and illocutionary force indicators for the classification of speech acts (e.g., assertive, directive) through the utilisation of supervised machine learning models that have been trained on annotated examples.

6. Perlocutionary analysis assesses the emotional and manipulative potential of verbal structures using sentiment analysis and emotion classification models (e.g., BERT, RoBERTa or GPT-based). Lexical items are evaluated using emotion lexicons (e.g., NRC and LIWC) to determine their influence on how recipients perceive them and their potential behavioural effects.

7. Model construction. A layered analytical model is developed, mapping each component of the message to its corresponding speech act level. The model visualises how linguistic elements contribute to the overall communicative and manipulative power of disinformation.

8. Evaluation and validation. The model is evaluated using new message samples to ascertain its precision in detecting emotional tone, communicative intent, and manipulative constructs. The evaluation process involves human-in-the-loop validation and cross-comparison with existing fact-checking datasets.

The methodological innovation resides in the systematic integration of speech act theory with computational linguistic analysis, thereby facilitating the automation of disinformation diagnostics not only at the lexical-semantic level, but also at the level of pragmatic intent. This methodology constitutes the basis for the development of scalable tools within the domains of information security and cognitive resilience systems.

3. Research Results

The dissemination of disinformation can be regarded as a communicative act that has the capacity to influence the perceptions, emotions, and actions of the recipient. The emotional tone of disinformation is a key indicator of its effectiveness. The formation of

verbal elements is understood as a communicative act. The identification of the specifics of communicative intent in message formation is carried out from the perspective of speech act theory (Austin, 1986).

The authors propose identifying the locutionary, illocutionary and perlocutionary components of verbal disinformation and transposing the terminology traditionally applied to specific speech acts into the information environment (see Fig. 1).

A communicative act is a complex phenomenon. According to speech act theory, a speech act can be analysed at three levels. Initially, a speech act can be viewed as the act of saying something. In this aspect, the speech act is treated as a **locutionary act** (from Latin 'locution' – "speaking").

The locutionary act is a complex structure in itself, comprising sound production (phonation), word usage, syntactic arrangement and the denotation of objects through language (reference), as well as the attribution of properties or relations to these objects (predication). Traditionally, the focus of linguistics has been on the locutionary aspect of speech acts. However, the **illocutionary** and **perlocutionary** effects – especially in the context of disinformation tone analysis – have rarely been systematically explored, despite their potential to influence public consciousness both directly and indirectly.

Later, American philosopher (Searle, 1986), a student of J. Austin, in his article *A Taxonomy of Illocutionary Acts*, proposed several categories of illocutionary acts, which the authors also consider relevant for identifying verbal elements in disinformation messages:

- **Assertives** indicate that the speaker commits to the truth of the expressed proposition (for example, to affirm or deny).
- **Directives** show that the speaker attempts to persuade the listener to perform an action (for example, to request, command, permit, or suggest).
- **Commissives** express the speaker's commitment to a future action (for example, to promise, swear, or guarantee).
- **Expressives** convey the speaker's psychological state (for example, to thank, congratulate, apologise, or sympathise).

Morphological units of emotive and evaluative vocabulary—such as **epithets**, **anaphora**, **metaphors**, and **hyperbole** – often manifest in adjectives, nouns, and verbs. It has been demonstrated that these parts of speech most fully convey the pragmatic orientation of a text (Grajs G.P., 1985). The stylistic colouring employed serves to accentuate pivotal elements and direct the viewer's attention towards the desired emotional framing.

By constructing conceptual images of objects, news or messages, nouns enable the recipient to make generalised associations. Verbs, with their dynamic and expressive nature, influence and prompt the recipient,

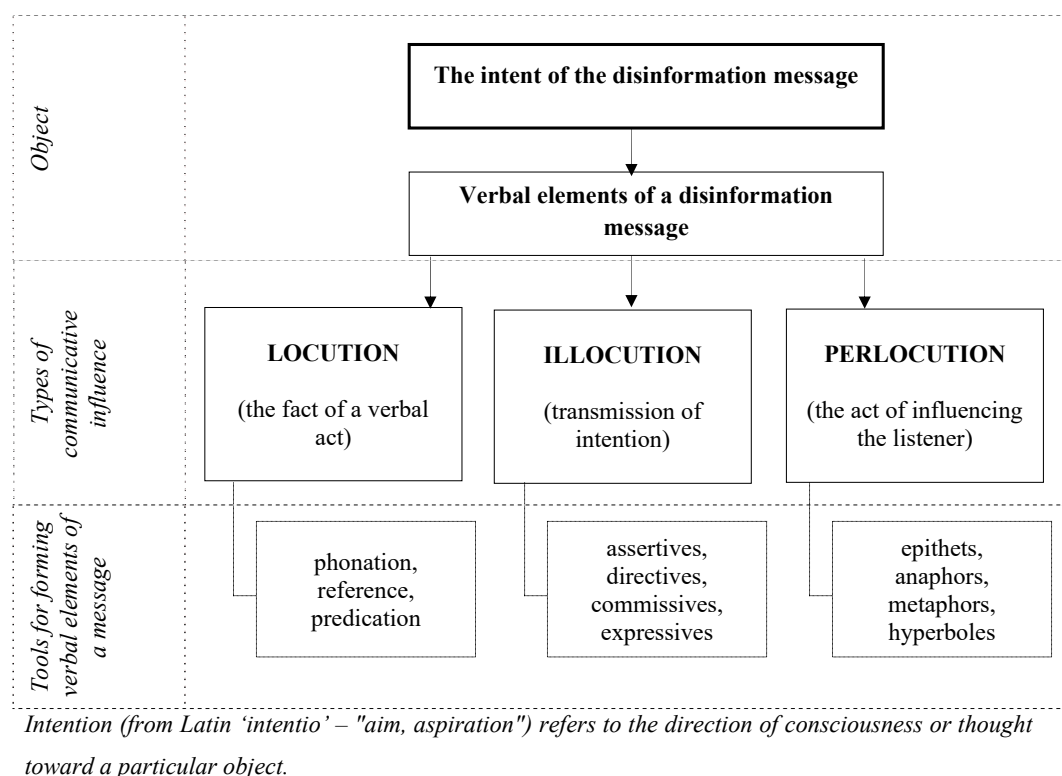


Figure 1. Methodological approach to the formation of verbal elements in disinformation messages

Source: developed by the authors

thus enhancing the perlocutionary effect. Adjectives provide imagery, emotional tone and stylistic richness, enabling the author to influence perception and imagination (Grishhenko, 2007). Due to their high degree of subjectivity and emotionality, adjectives often function as tools of manipulation and are central to achieving a disinformative impact.

From a linguistic perspective, effective communication is defined as the transmission of the author's illocutionary intention, resulting in a tangible perlocutionary effect.

The effectiveness of a message's perception depends on the composition and structuring of its components, the presence of a representational image and an underlying idea or persuasive aim. The communicative function of informational appeals can be realised through the use of specific words (e.g., addressing the recipient directly) and grammatical constructions (e.g., using specific pronouns or imperative constructions).

It is crucial to analyse the internal structure of verbal elements and their communicative context. Two strategic errors may arise: paying insufficient attention to the characteristics of the intended audience (which can result in communication failure), or adapting the language too much (which can lead to the distortion of linguistic norms). Tailoring verbal elements to the intended audience can enhance the communicative impact. The interpretation of an information message

often varies depending on the subject's personal perception.

From a communicative standpoint, verbal elements are components of a speech situation. Here, extralinguistic factors such as the topic, the intended audience and the author's intention influence the selection of linguistic tools, and consequently the genre and stylistic form of expression. Thus, **illocutionary force** (authorial intention) manifests in news texts through specific codes – verbal, structural, and non-verbal.

The effective use of content-related components in verbal structures, especially those with strong locutionary and perlocutionary dimensions, has been demonstrated to enhance overall communication. This approach enables the message to reach its intended audience quickly and efficiently, while shifting its tone from neutral information to targeted disinformation.

In emotionally charged contexts involving disinformation, **artificial intelligence (AI)** tools, particularly those based on **natural language processing (NLP)**, are vital analytical instruments. These technologies enable the automated analysis of textual content to detect emotional content and potential manipulation.

One key direction is the identification of emotional tone. Pretrained language models such as **BERT**, **RoBERTa**, or **GPT-like architectures** can classify

text into emotional categories including fear, anger, anxiety, outrage, compassion, etc. This allows messages that could cause mass anxiety or social tension through emotional manipulation to be detected.

The next step involves detecting **manipulative linguistic patterns** typical of disinformation content, including:

- Excessive use of expressive vocabulary;
- rhetorical structures;
- stylistic exaggerations (hyperbole, metaphors, epithets);
- emotional appeals through contrast or personalisation (e.g., "enemy-us," "traitor-hero");
- use of evaluative adjectives that impose subjective tone without evidence.

In addition, AI can help identify the **illocutionary intentions** of messages: assertions, calls to action, expressions of emotion, etc. This is achieved via analysis of syntactic structures, modal verbs, intention markers, and discourse formulas. Thus, AI is capable not only of detecting disinformation but also of interpreting the **communicative intent** of its author.

To ensure high accuracy, analytical systems combine:

- Specialised emotion lexicons (e.g., NRC Emotion Lexicon, LIWC);
- corpora of annotated disinformation texts;
- audience response analysis modules (e.g., comments, engagement, sharing).

In practice, such tools are implemented as:

- Automated fact-checking instruments;
- content filtering or tagging modules on social media platforms;
- analytical support tools for information security professionals, journalists, and researchers.

Algorithm for Disinformation Message Analysis Based on Speech Act Theory and AI Tools

1. The text is preprocessed through tokenisation, noise and tag removal, and word normalisation.
2. The intent of the disinformation message is determined by identifying the types of communicative impact, key lexemes, and frames, such as triggers of fear, conflict, enemy image, conspiracy, or calls to action.
3. The tools involved in the formation of verbal elements are identified at three levels – locution (phonation, reference, predication), illocution (assertives, directives, commissives, expressives),

Table 1

Examples of disinformation message analysis based on speech act theory and AI tools

Fake news	AI tools used	Algorithm/module	AI-determined intent of a disinformation message	Possible impact on the audience
"Hundreds of people died after vaccination in India. This is the result of a global experiment on humanity!" (spread on social media, 2021)	NLP model analyses semantics and tone: finds words with fear ("died", "experiment"), rhetorical structures	◆ Sentiment Analysis ◆ Named Entity Recognition ◆ Text Classification	◆ Emotion: fear, anger ◆ Illocution: assertive, expressive ◆ Manipulation: conspiracy theory, generalisation	Fear of vaccination, undermining of trust in the healthcare system
"Ukrainian refugees receive more aid than locals!" (propaganda sites, EU, 2022)	AI performs lexical and emotional analysis: highlights comparisons, emotionally colored adjectives ("more", "dishonest"), emphasis on conflict between groups	◆ Emotion Detection ◆ Text Framing Detection ◆ Topic Modeling	◆ Emotion: envy, indignation ◆ Illocution: directive ◆ Manipulation: social opposition	Provocation of hatred, social tension between refugees and locals
"The US army created the virus in a laboratory – documents confirm this!" (Russian disinformation, 2020–2021)	Model looks for fake structures: no evidence, references to "documents", exaggeration, "conspiracy" semantics, classifier of typical fakes is applied	◆ Fake News Classifier ◆ Evidence Checker ◆ Semantic Similarity Search	◆ Emotion: suspicion, anger ◆ Illocution: assertive ◆ Manipulation: pseudo-facts, conspiracy, fear	Distrust of international partners, change of geopolitical loyalties
"Europe is on the verge of collapse: Ukrainians have destroyed the economy!" (anonymous Telegram channels, 2022–2023)	AI identifies exaggeration, hostile rhetoric, emotional tension; detects anonymous source as a marker of unreliability	◆ Hyperbole Detector ◆ Hate Speech Detection ◆ Source Credibility Scoring	◆ Emotion: fear, anger ◆ Illocution: directive, expressive ◆ Manipulation: hyperbole, xenophobia	Increasing anti-Ukrainian sentiment, radicalisation of part of the audience

Source: compiled by the authors

and perlocution (epithets, anaphora, metaphors, hyperbole).

4. Reliability is assessed through source verification, reference checking, and fact-checking.

5. The potential impact is evaluated by analysing the emotional, social, and political harm.

4. Conclusions

The study has shown that disinformation messages have a complex structure of communication based on the realisation of locutionary, illocutionary and perlocutionary speech acts. The verbal content of the message is the primary tool of influence, with lexical units bearing emotional and evaluative connotations amplifying the psychological effect of the information and shaping how it is perceived by the audience.

Artificial intelligence, particularly natural language processing (NLP) technologies, provides an effective means of detection:

- The emotional tone of messages (e.g., fear, anger, anxiety, compassion);
- types of illocutionary intentions (e.g., directives, assertions, emotional expressions);
- key semantic constructions that constitute manipulative content.

This approach allows for a more in-depth analysis of disinformation texts, examining not only their content, but also identifying their pragmatic goals and potential influence on mass consciousness. Integrating speech act theory with AI creates new possibilities for automatically detecting and mitigating harmful informational influences, thereby strengthening the resilience of the information environment.

Consequently, the utilisation of artificial intelligence in the analysis of disinformation signifies a promising trajectory, facilitating the integration of linguistic, psychological, and technological dimensions in the development of systems designed to counter emotionally charged manipulative influence.

References:

- Austin, J. L. (1986). Word as action. *New in foreign linguistics. Theory of Speech Acts*, 17, 22–129.
- Searle, J. (1986). What is a speech act. *New in foreign linguistics: theory of Speech Acts*, 17, 151–169.
- Grajs G. P. (1985). Logic and Speech Communication. *New in foreign linguistics: linguistic pragmatics*, 16, 217–237.
- Grishhenko, A. I. (2007). Sources of the emergence of expressive ethnonyms (ethnofolisms) in modern Russian and English: etymological, motivational and derivational aspects. *Materials of the International Conference in Memory of L.V. Nikolenko and Y.P. Soloduba «Active processes in modern vocabulary and phraseology»*. (pp. 40–52). Yaroslavl: TOV «Remder».
- Borisova, E. G. (2001). Perlocutionary Linguistics and its Teaching to Philology Students. *Bulletin of Moscow University*, 1, 115–133.
- Vinay, R., Oehmichen, A., Agirre, E., & Davis, B. (2024). Emotional Manipulation Through Prompt Engineering Amplifies Disinformation Generation in AI Large Language Models. *ArXiv preprint arXiv:2405.15923*. Available at: <https://arxiv.org/abs/2403.03550>
- Smith, S. T., Kao, E. K., Mackin, E. D., Shah, D. C., Simek, O., & Rubin, D. B. (2020). Automatic Detection of Influential Actors in Disinformation Networks. *ArXiv preprint arXiv:2010.11920*. Available at: <https://arxiv.org/abs/2005.10879>
- Harris, S., Hadi, H. J., Ahmad, N., & Alshara, M. A. (2024). Fake News Detection Revisited: An Extensive Review of Theoretical Frameworks, Dataset Assessments, Model Constraints, and Forward-Looking Research Agendas. *Technologies*, 12, 222. DOI: <https://doi.org/10.3390/technologies12110222>
- Simon, F. M., Altay, S., & Mercier, H. (2023). Misinformation reloaded? Fears about the impact of generative AI on misinformation are overblown. *Harvard Kennedy School (HKS) Misinformation Review*. DOI: <https://doi.org/10.37016/mr-2020-127>
- Bell, E. (2023, March 3). Fake news, ChatGPT, truth, journalism, disinformation. *The Guardian*. Available at: <https://www.theguardian.com/commentisfree/2023/mar/03/fake-news-chatgpt-truth-journalism-disinformation>
- Goldstein, J. A., Chao, J., & Grossman, S., Stamos, A., & Tomz, M. (2023). Can AI write persuasive propaganda? *SocArXiv*. DOI: <https://doi.org/10.31235/osf.io/fp87b>

Received on: 14th of August, 2025

Accepted on: 02th of October, 2025

Published on: 29th of October, 2025