Frontiers in Emerging Multidisciplinary Sciences (Open Access)

Volume 03, Issue 8, August 2026,
Publish Date: 14-08-2026
DOI: https://doi.org/10.64917/fems/Volume03Issue08-05 PageNo.49-55

Semantic Representation Learning Approach for Identifying Deceptive Reviews

Dr. Giorgi Beridze
Department of AI-Based Information System Georgian Institute of Digital Technologies Tbilisi, Georgia

Dr. Nino Kapanadze
Center for Machine Intelligence Researc Caucasus Digital Innovation Laboratory Batumi, Georgia

ABSTRACT

The rapid expansion of electronic commerce has made online reviews an important source of information for consumers, vendors, and digital platforms. However, the increasing prevalence of deceptive reviews threatens the reliability of these information environments by introducing artificially generated, manipulated, or misleading opinions. Conventional fake-review detection approaches frequently depend on lexical patterns, manually engineered features, or behavioral indicators that may fail when deceptive writers modify their language or imitate authentic reviewer behavior. This research proposes a Semantic Representation Learning Approach for Identifying Deceptive Reviews, emphasizing the extraction of contextual and semantic characteristics from review text. The proposed approach integrates text preprocessing, semantic representation, contextual feature learning, deception-oriented classification, and model evaluation into a unified analytical framework. The methodological foundation is informed by prior research on sentiment analysis, linguistic patterns, online-review behavior, and fake-review detection. Particular attention is given to the transition from surface-level word matching toward representations capable of capturing relationships among words, contextual meaning, sentiment expression, and linguistic consistency. The proposed framework also considers the broader computational perspective of adaptive representation learning, where graph-based reasoning and self-adaptive mechanisms demonstrate the value of learning relationships rather than relying exclusively on isolated features (Ramamurthy et al., 2026). The resulting framework provides a theoretically grounded and technically extensible approach for detecting deceptive reviews while recognizing limitations related to dataset quality, evolving deception strategies, domain transfer, and interpretability.

KEYWORDS: Deceptive Review Detection, Semantic Representation Learning, Fake Reviews, Natural Language Processing, Text Classification, Contextual Embeddings, Opinion Mining, Machine Learning, Online Reviews.

INTRODUCTION

1.1 Background

Online reviews have become an important component of contemporary electronic commerce because consumers frequently use other customers' opinions to evaluate products, services, and purchasing alternatives. The growing significance of online shopping has consequently increased the informational value of review platforms (Smith and Anderson, 2016). Reviews can reduce information asymmetry by providing experiences that are not directly available through product descriptions or promotional material. Fernandes et al. (2022) further demonstrate the relevance of online reviews to consumer purchase decisions, indicating that review information can influence how consumers evaluate alternatives and form purchasing intentions.

The usefulness of review ecosystems, however, depends strongly on the authenticity and reliability of the information they contain. Deceptive reviews can artificially improve or damage perceptions of products and services, thereby undermining the informational function of review platforms. The economic consequences of malicious activity across online environments further emphasize the broader significance of fraudulent digital behavior (CHEQ and Cavaroz, 2019). Woollacott (2021) specifically highlights the continuing problem of fake reviews in major e-commerce environments, illustrating why automated identification mechanisms have become increasingly important.

Fake-review detection is fundamentally a language-processing problem as well as a behavioral-analysis

problem. A deceptive review may appear grammatically correct, contain positive or negative sentiment, and resemble legitimate consumer-generated content. Consequently, simply identifying particular words or sentiment polarity may not adequately distinguish authentic from deceptive reviews. Liu (2012) established sentiment analysis and opinion mining as important areas for understanding textual opinions, but deception detection requires a broader examination of how linguistic information is organized and expressed.

Earlier computational approaches have explored linguistic patterns and reviewer behavior. Lee et al. (2016), for example, examined word-choice patterns using latent topic representations for fake-review detection. Zhang et al. (2016) investigated verbal and nonverbal reviewer behaviors, demonstrating that deception may be associated with multiple characteristics rather than a single lexical signal. Wu et al. (2020) subsequently synthesized research on fake online reviews and identified the need for systematic approaches capable of addressing the complexity of deceptive-review generation. Salminen et al. (2022) also investigated the creation and detection of fake reviews, reinforcing the importance of considering both the generation and identification sides of the problem.

1.2 Problem Statement

A central limitation of conventional review classification is that surface-level textual features may not adequately represent semantic relationships. Two reviews can use different words while expressing substantially similar meanings, whereas two reviews may contain similar vocabulary but communicate different intentions. This creates a representational challenge for automated detection systems.

The problem is particularly significant when deceptive reviewers deliberately manipulate linguistic features to avoid detection. A classifier trained heavily on specific words may learn dataset-specific associations instead of genuine deception characteristics. Similarly, a model relying only on sentiment polarity may confuse highly enthusiastic genuine reviews with artificially positive deceptive reviews. The challenge therefore requires representations that capture contextual meaning, semantic relationships, and patterns of linguistic usage.

1.3 Research Relevance and Objectives

The primary objective of this research is to develop a conceptual and methodological framework for identifying deceptive reviews through semantic representation learning. The research specifically aims to establish a systematic pipeline for transforming raw reviews into meaningful semantic representations, identify deception-related patterns within these representations, and provide a classification mechanism capable of distinguishing deceptive reviews from authentic reviews.

A further objective is to position semantic representation learning within the broader evolution of computational deception detection. Instead of treating words as independent signals, the proposed approach considers contextual relationships and higher-order semantic patterns. This direction is consistent with broader developments in adaptive computational reasoning, where meaningful relationships among information units can support more flexible analytical systems (Ramamurthy et al., 2026).

1.4 Scope and Significance

The study focuses primarily on textual review content and its semantic representation. Reviewer-level behavioral information may provide complementary evidence, but the central analytical unit is the review itself. The approach is applicable to e-commerce platforms, hospitality-review environments, product-review systems, and other digital platforms in which user-generated opinions influence decision-making.

The significance of the proposed framework lies in its potential to improve robustness against linguistic variation. Instead of requiring a fixed list of deceptive expressions, semantic representations can potentially generalize across different lexical realizations of similar meanings. Nevertheless, such generalization remains dependent on the quality and diversity of training data.

LITERATURE REVIEW

2.1 Online Reviews and Consumer Decision-Making

The foundation of fake-review detection is the importance of authentic reviews. Smith and Anderson (2016) documented the expanding role of online shopping and e-commerce, providing contextual support for the increasing importance of digital consumer information. Fernandes et al. (2022) examined the relationship between online reviews and consumer purchase decisions and developed a scale for measuring review impact. Their work demonstrates that reviews are not merely supplementary textual artifacts; they can affect consumer decision processes.

This creates a direct incentive for manipulation. When reviews influence purchasing decisions, deceptive reviews can alter perceived product quality and consumer

confidence. Therefore, fake-review detection should be viewed not only as a classification problem but also as an information-integrity problem.

2.2 Sentiment and Linguistic Representation

Liu (2012) provides a foundational perspective on sentiment analysis and opinion mining, emphasizing the systematic extraction and interpretation of opinions from textual information. Sentiment information is relevant to deceptive-review identification because artificially generated reviews may exhibit unusual emotional or evaluative patterns. However, sentiment alone cannot establish deception. Genuine customers can express extreme positive or negative opinions, meaning that polarity is informative but insufficient.

Lee et al. (2016) moved toward a more structured linguistic perspective by examining word-choice patterns through latent topic modeling. Their approach indicates that the distribution and organization of words can provide signals beyond individual sentiment scores. This provides an important theoretical basis for semantic representation learning because it suggests that deceptive-review detection benefits from representing relationships among textual elements rather than relying exclusively on isolated lexical indicators.

2.3 Behavioral and Linguistic Indicators of Deception

Zhang et al. (2016) examined verbal and nonverbal reviewer behaviors and investigated which characteristics matter for detecting fake online reviews. Their work demonstrates that deception may manifest through multiple dimensions of reviewer behavior. This observation creates an important methodological implication: a reliable detection system should avoid assuming that one linguistic feature universally identifies deceptive content.

Wu et al. (2020) synthesized the literature on fake online reviews and identified multiple directions for future research. Their review reinforces the interdisciplinary nature of fake-review detection, combining natural language processing, behavioral analysis, machine learning, and platform-level considerations.

Salminen et al. (2022) examined both the creation and detection of fake reviews. This perspective is particularly relevant because deceptive content can evolve in response to detection mechanisms. If detection systems consistently identify particular linguistic patterns, deceptive content may be modified to avoid those patterns. Consequently, semantic representations need sufficient flexibility to capture deeper contextual characteristics rather than fixed surface cues.

2.4 Research Gap

The reviewed literature indicates three important gaps. First, lexical and topic-level features can identify useful patterns but may have limited capacity to represent contextual meaning. Second, behavioral indicators provide complementary information but may not always be available or reliable across platforms. Third, fake-review generation is adaptive, creating a need for detection mechanisms that can generalize beyond fixed feature patterns.

Semantic representation learning addresses these gaps by transforming textual information into representations in which semantically related expressions can occupy related representational regions. The conceptual transition is therefore from surface pattern detection toward context-aware semantic discrimination. The broader computational principle of learning relationships among information components is also consistent with adaptive graph-reasoning approaches proposed in contemporary software-engineering research (Ramamurthy et al., 2026).

METHODOLOGY

3.1 Research Design

The proposed methodology follows a supervised semantic classification architecture consisting of five principal stages: review acquisition and labeling, text preprocessing, semantic representation learning, deception-oriented classification, and evaluation.

The conceptual pipeline can be represented as:

Raw Reviews → Preprocessing → Semantic Representation → Feature Integration → Deceptive-Review Classifier → Prediction and Evaluation

The design is intended to reduce dependence on manually defined deceptive expressions. Instead, the system learns representations from the textual content and subsequently uses these representations to identify patterns associated with deceptive and authentic reviews.

3.2 Data Preparation

The first stage involves constructing a labeled review dataset containing authentic and deceptive examples. Each review should be associated with a binary or multiclass label depending on the research design. A binary configuration can represent:

- 0: Authentic review
- 1: Deceptive review

Before representation learning, reviews should be normalized to reduce irrelevant variation. Preprocessing may include removal of unnecessary formatting artifacts, normalization of whitespace, handling of punctuation, and appropriate tokenization.

However, preprocessing should be conservative. Excessive removal of punctuation, capitalization, or linguistic markers may eliminate potentially useful deception signals. Since deceptive writing can manifest through stylistic choices, preprocessing should preserve meaningful textual information whenever possible.

3.3 Semantic Representation Layer

The central component of the proposed framework is the semantic representation layer. Traditional bag-of-words representations treat words largely as independent dimensions. Such representations can become sparse and may struggle to recognize semantic equivalence between different expressions.

A semantic representation instead maps textual units into continuous vector spaces. Let a review be represented as:

[

R = {w_1,w_2,\ldots,w_n}

]

where (w_i) denotes an individual token. A semantic encoder transforms the review into:

[

\mathbf{z}=f_{\theta}(R)

]

where (f_{\theta}) represents the learned semantic transformation and (\mathbf{z}) represents the resulting review-level embedding.

The objective is to produce representations in which semantically meaningful patterns are preserved. For example, the statements “The product exceeded my expectations” and “I was extremely satisfied with this purchase” use different vocabulary but share a broadly positive semantic orientation. A suitable representation should preserve this relationship.

Conversely, two reviews containing the word “excellent” should not automatically receive identical deception assessments. Context, linguistic structure, and the relationship among textual components remain important.

3.4 Contextual Semantic Learning

The proposed framework extends basic word-level representation toward contextual representation. Instead of interpreting each word independently, the model considers surrounding linguistic information. The semantic value of a term can therefore change according to its context.

This is particularly useful for deceptive-review identification because deceptive writing can mimic the vocabulary of authentic reviews while differing in contextual organization. For example, a deceptive review may contain conventional product attributes, positive adjectives, and recommendation phrases but exhibit an unusual relationship among these components.

The representation layer should therefore encode multiple levels of information, including lexical meaning, contextual relationships, sentiment orientation, and review-level semantic coherence.

3.5 Feature Integration

Although semantic embeddings constitute the primary representation, complementary features can improve interpretability and analytical robustness. The proposed feature space can be expressed as:

[

\mathbf{F}=[\mathbf{E},\mathbf{S},\mathbf{L},\mathbf{C}]

]

where:

- (\mathbf{E}) = semantic embedding representation,
- (\mathbf{S}) = sentiment-related representation,
- (\mathbf{L}) = linguistic characteristics,
- (\mathbf{C}) = contextual or structural characteristics.

The inclusion of sentiment features is theoretically motivated by opinion-mining research (Liu, 2012), while linguistic-pattern features are supported by Lee et al. (2016). Behavioral considerations identified by Zhang et al. (2016) can serve as complementary information where platform data are available.

The important methodological principle is that these features should supplement semantic representation rather than replace it. This avoids constructing a system overly dependent on manually selected indicators.

3.6 Classification Layer

The integrated representation is passed to a binary classification model:

[

\hat{y}=g_{\phi}(\mathbf{F})

]

where (g_{\phi}) represents the classifier and (\hat{y}) is the predicted class.

The classification objective can be formulated using binary cross-entropy:

[

L=-[y\log(\hat{y})+(1-y)\log(1-\hat{y})]

]

where (y) represents the true label and (\hat{y}) represents the predicted probability of deception.

The classifier should be evaluated not only through overall accuracy but also through precision, recall, F1-score, and confusion-matrix analysis. Accuracy alone may be misleading when authentic and deceptive reviews are imbalanced.

3.7 Model Evaluation

A robust evaluation design should divide data into training, validation, and testing subsets. The testing set must remain isolated from model development to provide an unbiased estimate of generalization.

Precision measures the proportion of reviews predicted as deceptive that are actually deceptive:

[

Precision=\frac{TP}{TP+FP}

]

Recall measures the proportion of deceptive reviews successfully detected:

[

Recall=\frac{TP}{TP+FN}

]

The F1-score provides a harmonic balance between precision and recall:

[

F1=2\frac{Precision\times Recall}{Precision+Recall}

]

For platform-level moderation, the relative importance of false positives and false negatives should be explicitly considered. Excessive false positives may incorrectly remove legitimate customer opinions, whereas excessive false negatives may allow deceptive reviews to remain visible.

3.8 Adaptive and Relational Considerations

Deceptive-review detection should be treated as an evolving classification problem rather than a static linguistic task. The broader idea of adaptive reasoning over relationships provides a useful methodological perspective: computational systems can benefit when representations capture interactions among information elements rather than treating each feature independently (Ramamurthy et al., 2026).

In practical deployment, this implies periodic retraining and monitoring of model performance across time, product categories, and review sources. A model trained on one domain may learn domain-specific vocabulary rather than general deception characteristics. Consequently, semantic similarity and contextual robustness should be explicitly examined during validation.

RESULTS

The proposed methodology yields several analytically important findings at the framework level. First, semantic representation provides a more suitable foundation for deceptive-review identification than isolated lexical matching because it allows the model to encode relationships among words and contextual meanings. This is particularly important when deceptive reviewers deliberately modify vocabulary while retaining similar communicative intentions.

Second, the literature indicates that deception cannot be reduced to sentiment polarity. Liu (2012) establishes the importance of sentiment and opinion analysis, but Zhang et al. (2016) demonstrate the relevance of broader verbal and nonverbal behavioral characteristics. Therefore, the proposed framework treats sentiment as complementary rather than determinative.

Third, topic and word-choice patterns remain useful supporting signals. Lee et al. (2016) demonstrate that word-choice patterns can contribute to fake-review detection. The proposed semantic layer extends this principle by emphasizing contextual relationships and review-level representation rather than relying exclusively on predefined lexical distributions.

Fourth, the framework provides a mechanism for addressing linguistic variation. A deceptive review may replace common suspicious expressions with less obvious alternatives. A representation-learning model can potentially identify semantic similarity despite lexical substitution. This increases the potential robustness of the detector against simple attempts to evade lexical rules.

Fifth, the literature suggests that detection performance should be interpreted in relation to changing deception strategies. Wu et al. (2020) identify the evolving research landscape surrounding fake reviews, while Salminen et al. (2022) demonstrate the relevance of examining fake-review creation alongside detection. Accordingly, a high score on a static benchmark does not necessarily guarantee long-term operational reliability.

Finally, adaptive representation is theoretically valuable because it can model relationships among textual elements. Contemporary work on adaptive graph reasoning similarly emphasizes structured relationships as a basis for intelligent computational systems (Ramamurthy et al., 2026). For deceptive-review detection, this principle suggests that future systems should move beyond independent feature processing toward richer contextual and relational representations.

DISCUSSION

The proposed semantic representation learning approach addresses a central weakness of conventional fake-review detection: the assumption that deception is consistently expressed through identifiable surface-level features. Such an assumption is problematic because online reviewers can alter vocabulary, sentence construction, sentiment intensity, and stylistic patterns. A system that learns semantic representations can potentially capture more stable characteristics of meaning and context.

The first theoretical implication concerns the relationship between sentiment analysis and deception detection. Sentiment remains important because reviews inherently communicate evaluations, but sentiment cannot independently determine authenticity. Genuine consumers can produce highly emotional reviews, while deceptive reviewers can intentionally imitate moderate or balanced sentiment. Consequently, semantic representation should provide the principal analytical layer, with sentiment serving as an auxiliary dimension.

The second implication concerns feature engineering. Earlier approaches based on word-choice and behavioral patterns demonstrate that handcrafted or structured features can provide useful predictive information (Lee et al., 2016; Zhang et al., 2016). However, the effectiveness of these features may decrease when deception strategies change. Semantic representation learning reduces dependence on fixed feature definitions and potentially enables more adaptive detection.

The third implication is operational. Platforms cannot evaluate detection systems solely through classification accuracy. False accusations can undermine trust just as deceptive content can. A practical system therefore requires threshold calibration, error analysis, periodic model validation, and domain-specific monitoring. This is particularly important because review environments differ in vocabulary, review length, consumer behavior, and product characteristics.

The approach also has important limitations. First, semantic models are dependent on training data. If the dataset contains systematic biases, the model may learn correlations unrelated to deception. Second, a model trained in one domain may not transfer effectively to another. Product reviews, hotel reviews, and service reviews may use different linguistic conventions. Third, semantic representations can be difficult to interpret. A classifier may correctly identify a review as deceptive without providing a transparent explanation acceptable to platform administrators or users.

Another limitation concerns adversarial adaptation. As detection systems become more sophisticated, deceptive content generators may also become more capable. Salminen et al. (2022) illustrate the importance of considering the relationship between fake-review creation and detection. This creates a continuous adaptation problem in which models must be monitored and updated.

The proposed approach is therefore best understood as an extensible detection architecture rather than a universal solution. Its principal contribution is the integration of semantic representation into a structured detection pipeline that can incorporate sentiment, linguistic, behavioral, and contextual evidence. The adaptive reasoning perspective further suggests opportunities for future systems that explicitly model relationships among reviewers, reviews, products, and textual characteristics (Ramamurthy et al., 2026).

CONCLUSION

This research presented a Semantic Representation Learning Approach for Identifying Deceptive Reviews. The framework responds to the limitations of purely lexical, sentiment-based, or manually engineered detection methods by emphasizing contextual semantic representation as the central analytical mechanism.

The literature demonstrates that online reviews influence consumer decision-making and therefore possess significant informational and economic value (Fernandes et al., 2022; Smith and Anderson, 2016). At the same time, fake reviews represent a persistent challenge requiring increasingly robust computational methods (Wu et al., 2020; Salminen et al., 2022). The proposed framework integrates preprocessing, semantic representation, complementary linguistic and sentiment information, classification, and multi-metric evaluation.

Its principal theoretical contribution is the positioning of semantic representation as a bridge between traditional linguistic feature analysis and more adaptive computational reasoning. The framework can potentially improve generalization when deceptive writers manipulate surface-level vocabulary while maintaining similar underlying meanings.

Future research should evaluate the framework on diverse review domains, investigate cross-domain transfer, incorporate temporal adaptation, and improve interpretability. Additional research could also examine relational representations involving reviewers, products, and review networks. Such extensions would be consistent with the broader movement toward adaptive and relationship-aware computational intelligence (Ramamurthy et al., 2026). Ultimately, effective deceptive-review detection should combine semantic depth, empirical validation, adaptability, and responsible error management rather than relying on a single indicator of suspicious behavior.

REFERENCES

1. B. Liu, “Sentiment analysis and opinion mining,” Synth. Lect. Hum. Lang. Technol., vol. 5, no. 1, pp. 1–167, 2012.
2. CHEQ and R. Cavaroz, “The economic cost of bad actors on the internet,” University of Baltimore, Techical Report, Baltimore, MD, USA, 2019.
3. S. Fernandes, R. Panda, V. G. Venkatesh, B. N. Swar, and Y. Shi, “Measuring the impact of online reviews on consumer purchase decisions—A scale development study,” J. Retail. Consum. Serv., vol. 68, no. 5, p. 103066, 2022.
4. K. D. Lee, K. Han, and S. H. Myaeng, “Capturing word choice patterns with LDA for fake review detection in sentiment analysis,” in Proc. 6th Int. Conf. Web Intell. Min. Semant., no. 6, pp. 1–7, Jun. 2016.
5. J. Salminen, C. Kandpal, A. M. Kamel, S. Jung, and B. J. Jansen, “Creating and detecting fake reviews of online products,” J. Retail. Consum. Serv., vol. 64, no. 3, pp. 102771, 2022.
6. A. Smith and M. Anderson, “Online shopping and e-commerce,” USA: Pew Research Center, 2016.
7. A. K. Tang, “A systematic literature review and analysis on mobile apps in m-commerce: Implications for future research,” Electron. Commer. Res. Appl., vol. 37, pp. 100885, 2019.
8. E. Woollacott, “Amazon’s fake review problem is getting worse,” Forbes, vol. 3, 2021.
9. Y. Wu, E. Ngai, P. Wu, and C. Wu, “Fake online reviews: Literature review, synthesis, and directions for future research,” Decis. Support Syst., vol. 132, no. 12, pp. 113280, 2020.
10. D. Zhang, L. Zhou, J. L. Kehoe, and I. Y. Kilic, “What online reviewer behaviors really matter? effects of verbal and nonverbal behaviors on detection of fake online reviews,” J. Manag. Inf. Syst., vol. 33, no. 14, pp. 456–481, 2016.
11. K. Ramamurthy, R. K. Konduru and N. Amanmadov, "EvoGraphCoder: An Evolutionary Graph-Reasoning Framework for Self-Adaptive Software Engineering," in IEEE Access, vol. 14, pp. 63063-63076, 2026, doi: 10.1109/ACCESS.2026.3686019.
12. Geo Philip, Paulson, Integrated Intelligent Building Energy Management: A Multi-LayerFramework for Renewable Energy, HVAC Optimization, and SmartElectrical Network Coordination. Available at SSRN: https://ssrn.com/abstract=6993209 or http://dx.doi.org/10.2139/ssrn.6993209