

Understanding Fake News through a Quantifiable Linguistic Feature Framework

Yunxi Lin

Putian University

Liang Jiang

Putian University

Shengliang Lin

shengliangl@jnu.edu.cn

Jinan University

Article

Keywords: Fake news, Linguistic feature framework, Explainable artificial intelligence, Machine learning, SHAP, Misinformation

Posted Date: October 13th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10786267/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Understanding Fake News through a Quantifiable Linguistic Feature Framework

Yunxi Lin¹, Liang Jiang^{1*}, Shengliang Lin^{2*}

^{1*}Fujian Key Laboratory of Financial Information Processing, Putian University, Putian, 351100, Fujian, China.

^{2*}School of Journalism and Communication, Jinan University, Guangzhou, 510632, Guangdong, China.

*Corresponding author(s). E-mail(s): ptjliang@ptu.edu.cn;
shengliangl@jnu.edu.cn;
Contributing authors: linyunxi@ptu.edu.cn;

Abstract

Fake news has become a growing concern in the AI-driven digital information environment. However, there is still a lack of a systematic, quantifiable, and interpretable linguistic feature framework for comprehensively characterizing fake news. Drawing on theories from communication, linguistics, and psychology, this study develops a quantifiable linguistic feature framework comprising four analytical dimensions, 24 linguistic features, and 103 fine-grained indicators. The SHAP framework is further combined with multiple machine learning models to evaluate and rank the importance of these linguistic features. Cross-model comparisons identify eight core linguistic features, which can be organized into three broader mechanisms: Attention and Information Intake, Semantic Coherence and Narrative Credibility, and Factivity, Stance, and Group Identity. These features reflect potential mechanisms through which fake news influences information processing and cognitive judgment, suggesting that fake news is not merely a linguistic deviation but a strategically constructed form of discourse that shapes audience perception and interpretation. By developing a systematic, quantifiable, and interpretable framework and identifying the core linguistic features of fake news, this study provides a new analytical perspective for understanding the linguistic mechanisms underlying fake news and offers practical implications for fake news detection, misinformation governance, and media literacy.

Keywords: Fake news, Linguistic feature framework, Explainable artificial intelligence, Machine learning, SHAP, Misinformation

1 Introduction

Fake news is a complex informational phenomenon, which is usually characterized by the deliberate fabrication or distortion of facts. It is typically constructed through specific information frameworks and interpretive pathways and leads audiences to accept manipulated information as credible. The extensive dissemination of fake news can exert profound negative impacts on both individuals and society. Technological innovations such as artificial intelligence and big data can rapidly generate large amounts of fake content and make the features of fake news more difficult to identify. Therefore, identifying the linguistic features of fake news is of critical importance.

Researchers have noted that several linguistic features could be important for detecting fake news, such as linguistic complexity, emotional intensity and subjectivity, semantic coherence, narrative structure, and content verifiability [1–6]. Although prior studies have explored various linguistic features of fake news, there is still a lack of a comprehensive and unified framework capable of capturing the multidimensional characteristics of fake news. Machine learning provides effective tools for feature selection and redundancy reduction.

However, even with interpretable linguistic features, models often struggle to capture the nuanced features of fake news [7, 8]. In particular, previous research has argued that relying on a limited number of features is insufficient for effective fake news detection [9].

Different models may extract and emphasize different features of fake news. As a result, feature attributions across models are often inconsistent, limiting their interpretability and comparability. Thus, interpretive biases may exist in the fake news features selected by models, reflecting model-specific preferences rather than the inherent importance of the features themselves [10, 11].

With an emphasis on feature extraction from fake news, standardizing feature attribution is essential to enable meaningful comparisons across different models [12]. To address the lack of systematic feature representation, this study constructs a multidimensional linguistic feature framework to systematically capture textual linguistic features. We also provide a set of quantifiable indicators for standardized description and comparative analysis.

For feature extraction, this study adopts the LIWC dictionary method and statistical methods for quantitative textual representation, thereby establishing a standardized and comparable feature foundation. Furthermore, different machine learning models often focus on different aspects of linguistic features. This leads to inconsistent feature attributions and limits cross-model comparability. To address this issue, we adopt SHAP as a unified interpretability framework, which provides consistent and comparable feature attribution across different models.

This study focuses on developing a quantifiable, multidimensional linguistic framework for systematically characterizing fake news and identifying its core linguistic features under a unified feature-attribution framework.

Our primary contributions are threefold. First, we construct a quantifiable linguistic feature framework for describing fake news. This framework comprises four analytical dimensions, 24 feature indicators, and 103 fine-grained metrics, which promote the systematic and standardized characterization of linguistic features embedded

in fake news. Second, we employ a SHAP-based feature importance ranking method to quantify the contribution of each linguistic feature across all models considered, generating an interpretable feature importance ranking. Finally, through cross-model comparative analysis, we identify the core linguistic features embedded in fake news, which has the potential to improve our empirical understanding of fake news. The immediate implication is that linguistic feature extraction aids in addressing practical misinformation governance problems, such as fake news identification, content screening, and public opinion regulation, justifying its core role in the technical architecture of online public opinion governance.

2 Literature Review

Based on large-scale social media data, prior research has found that fake news spreads faster, farther, deeper, and more broadly than true news [13]. This diffusion advantage could stem from the capacity of fake news to engage audiences’ cognitive and emotional processes, thereby making it more effective in capturing audience attention [14, 15].

Relevant theories offer important insights into understanding this phenomenon. From the perspective of information communication, such content competition is essentially driven by limited audience attention. Attention Economy Theory posits that, in a communication environment characterized by scarce attention, information production is increasingly centered on competing for audience attention [16, 17]. Against this backdrop, the linguistic features of fake news are often strategically optimized to attract and retain audience attention.

Selective Exposure Theory further suggests that information consistent with individuals’ existing attitudes is more readily accepted [18]. Additionally, heuristic processing mechanisms reduce the likelihood that audiences will systematically verify information [19]. Consequently, fake news often exhibits distinctive linguistic patterns in narrative framing, emotional expression, language tone, and information structure, providing theoretical support for the development of an integrated framework of linguistic features. These cognitive and psychological mechanisms are often embedded in the linguistic expressions of fake news, shaping its overall structure and presentation.

Recently, linguistic analysis has emerged as a key direction in fake news research. Prior studies have demonstrated that linguistic features serve as important criteria for differentiating real news from fake news [19–21]. However, fake news rarely manifests as obvious linguistic errors. Instead, it strategically mimics the discourse conventions of real news [22]. Consequently, differences between fake and real news are more evident in overall expression patterns and combinations of linguistic features than in isolated anomalies. This suggests that systematically cataloging and quantitatively analyzing linguistic features is essential for understanding the linguistic patterns of fake news. Moreover, a broad range of potentially relevant linguistic features remains largely underexplored, while only a limited subset has been widely investigated in prior research.

Machine learning has considerable potential for improving fake news detection, which is fundamentally a task of distinguishing real news from false news [23, 24]. However, detection results alone provide only classification outcomes and do not reveal the

underlying linguistic characteristics of fake news or provide sufficient interpretability [25]. Machine learning methods on their own are therefore limited in identifying deeper associations between textual content and the latent linguistic features of fake news.

More importantly, the interpretability of machine learning models facilitates an understanding of how linguistic features contribute to fake news detection [10, 11]. Different feature importance methods vary substantially in their theoretical assumptions and forms of result representation, which can lead to inconsistent feature attribution outcomes [12]. This limitation weakens the cross-model comparability and interpretability of linguistic features associated with fake news [26].

Currently, model-agnostic explanation methods such as LIME and SHAP have been widely adopted. LIME primarily explains individual predictions through local linear approximations. In contrast, SHAP places feature contributions within a unified additive interpretability framework, providing a consistent metric basis for cross-model comparisons. SHAP explains feature attribution rather than internal model mechanisms [27, 28]. Existing research indicates that it demonstrates strong consistency in interpretability [29]. In text analysis tasks, SHAP is widely used to generate comparable interpretability units for linguistic features, enabling consistent evaluation of feature contributions across different models [30]. Therefore, as a unified interpretability framework, SHAP is well suited to cross-model comparison and the identification of core linguistic features of fake news. This study adopts SHAP in combination with multiple machine learning models to examine linguistic feature attribution in fake news detection, with particular emphasis on cross-model comparative analysis and the identification of robust core features.

3 Data

3.1 Data Sources

This study uses the Fake and Real News Dataset, which is publicly available on Kaggle. The dataset consists of two parts: fake news and real news, containing approximately 23.5k and 21.4k news samples, respectively. Each news record contains four core fields: title, text, subject, and date. This dataset is considered one of the most widely used public datasets in current fake news research and is frequently employed as a baseline dataset [31].

3.2 Feature Construction

We construct a comprehensive set of linguistic features for fake news based on prior literature on fake news features. The literature draws on journalism, communication studies, linguistics, narratology, and psychology. Specifically, we develop a theoretical framework consisting of four analytical dimensions, 24 secondary linguistic features, and 103 quantifiable tertiary indicators. Table 1 presents the theoretical origins and relevance of the 24 secondary features listed in the first column. We ultimately aim to transform theoretical concepts, which are primarily qualitative, into quantifiable and verifiable linguistic indicators.

Table 1 Fake News Linguistic Features

Feature	Theoretical Origin	Theoretical Implication	Source
Framing	Framing Theory	Determines “how to understand”	[3]
Semantic Coherence	Discourse Semantics	Coherence implies plausibility	[4]
Discourse Coherence	Local vs. Global Coherence	Local coherence outweighs global judgment	[4]
Temporal Inconsistency	Situation Model Theory	Disrupts situational coherence without easy detection	[32]
Metaphor/Narrative Framing	Conceptual Metaphor Theory	Meaning is naturalized through narrative	[3]
Rhetorical Question	Rhetorical Pragmatics	Persuasion rather than inquiry	[33]
Text Structure / Layout	News Discourse Conventions	Form itself signals credibility	[34]
Entity & Quotation Pattern	Epistemic Authority	Authority functions as a linguistic cue	[35]
Emotional Polarity	Affective Valence	Reinforces stance	[5]
Subjectivity	Stance & Evaluation	Reduces perceived objectivity	[36]
Internal Emotional Spread	Emotional Variability	Produces emotional impact	[6]
Emotion Diversity	Affective Complexity	Enhances engagement	[5]
Emotional Dynamics	Narrative Emotion	Sustains attention	[5]
Modality / Epistemic Stance	Epistemic Modality	Shaping Cognitive Certainty	[37]
Lexical Complexity	Cognitive Ease	Easily processed information appears more credible	[38]
Readability	Processing Fluency	Ease of reading implies credibility	[39]
Exaggeration	Appraisal Theory	Amplifies emotional response	[3]
Ambiguity	Strategic Ambiguity	Avoids verification	[37]
Clickbait	Attention Economy	Manipulation occurs at the information entry stage	[32]
Pronoun Usage	Social Identity Theory	Factual judgment shifts to identity judgment	[37]
Pragmatic / Directive Cues	Speech Act Theory	Language guides action	[37]
Veracity Cues	Truth Effect	Truth cues terminate further reasoning	[19]
Factivity Verifiability	Epistemic Certainty	Certainty cues trigger default acceptance	[19]
Semantic Embedding Divergence	Frame Consistency	Rationality substitutes for factual accuracy	[4]

We adopt a top-down strategy to construct the linguistic features of fake news. Figure 1 presents a tree-structured framework for constructing fake news features. We first develop the feature framework from four upper-level perspectives: emotional psychology, narrative structure, linguistic style, and information complexity. These four perspectives have been extensively applied in fake news feature research [1, 13, 24, 40]. According to the theories summarized in Table 1, we further construct several fake news features under each perspective, resulting in a total of 24 features in the middle layer. These features are then further subdivided into 103 individual indicators at the bottom layer.

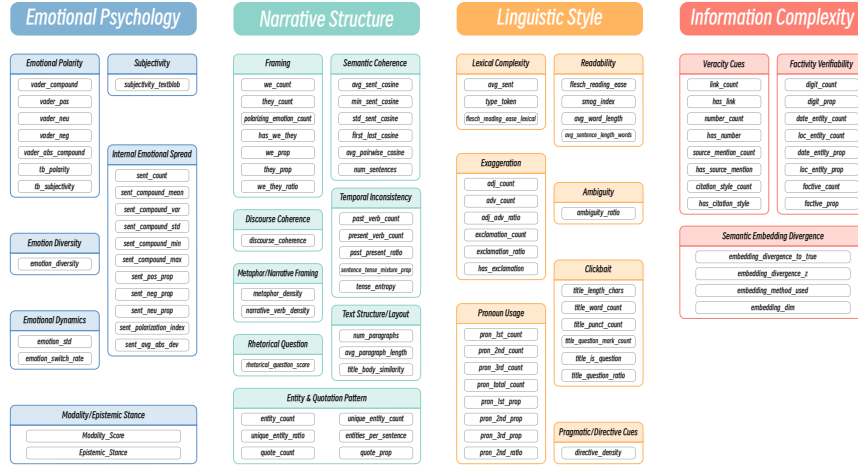


Fig. 1 Three-level quantifiable linguistic feature framework for fake news.

3.3 Feature Extraction

This section describes how we extract 103 fine-grained linguistic indicators from individual news articles and further integrate these indicators into 24 category-level linguistic features. Specifically, the extraction process relies on three main methods: statistical methods, lexicon-based methods, and natural language processing (NLP) techniques.

After extraction, the 103 fine-grained linguistic indicators are standardized using the Z-score method to remove scale differences. Indicators within the same category are then averaged to produce the final representations of these 24 features.

3.3.1 Statistical Method

We adopt statistical methods to compute word frequency, proportional ratios, and text distribution features. These approaches effectively capture basic linguistic patterns at the lexical, sentence, and paragraph levels, covering multiple quantitative indicators such as word and sentence length, lexical diversity, pronoun usage ratios, and readability metrics.

3.3.2 Lexicon-based Method

For the lexicon-based method, we adopt publicly available linguistic resources (e.g., VADER and TextBlob) as well as custom lexicons constructed for this research [41, 42]. The lexicons mainly cover linguistic features such as uncertainty expressions, modality and epistemic stance expressions, factive verbs, directive expressions, metaphorical expressions, narrative verbs, polarizing emotional expressions, pronouns, and tense-related expressions. Furthermore, we employ lexicon-based methods to identify specific

linguistic features in text through lexicon matching, and then compute the word frequency and proportion of each feature.

3.3.3 TF-IDF and sentence embedding Technique

We employ NLP-based techniques to extract linguistic features of fake news by measuring semantic similarities and differences between texts through vector-space and distributed semantic representations. Specifically, news texts are transformed into numerical vector representations using TF-IDF and Sentence-BERT-based semantic embedding methods [43, 44].

3.3.4 Aggregating Linguistic Features

After extracting the 103 fine-grained linguistic indicators, we standardize all indicators using the Z-score method to eliminate dimensional differences and ensure comparability across measures. We then aggregate indicators within the same feature category by averaging their standardized values. Through this process, 103 individual indicators were aggregated into 24 category-level linguistic features, which serve as the final feature representation for fake news analysis.

4 Methodology

To evaluate the fake news linguistic feature set proposed in Section 3, we employ five representative machine learning models and Principal Component Analysis (PCA). PCA is included as a conventional linear benchmark for comparison with the machine learning models.

All five machine learning models are trained on real and fake news samples. Feature interpretation results often vary across different machine learning models, which may lead to potential explanatory bias. To address this issue, we adopt the SHAP framework, which supports reliable cross-model comparison, as a unified interpretability tool to eliminate explanatory discrepancies across different machine learning models [27].

4.1 Models and Method

PCA Method. Principal Component Analysis (PCA) is a classical unsupervised linear dimensionality reduction method based on the covariance structure among features. Unlike supervised machine learning approaches, PCA relies on the variance-covariance structure of the feature matrix, without using class labels or prediction targets. As a mathematical dimensionality reduction method, PCA identifies the directions of maximum variance in the data through linear combinations of features [45].

In this study, PCA is applied to the standardized linguistic feature matrix of fake news samples. The number of principal components is automatically determined using a cumulative explained variance threshold of 95%. Feature contributions are estimated based on weighted squared loadings across the retained principal components, where each component is weighted by its explained variance ratio. Since PCA emphasizes

the internal variance structure rather than classification objectives or model interpretability, it is mainly used as an unsupervised comparative baseline for feature analysis.

Logistic Regression Model. Logistic Regression is a linear and interpretable classification model. In natural language processing research, this model is generally regarded as one of the most fundamental and highly interpretable models [46]. Model parameters are estimated by minimizing the cross-entropy loss function as follows:

$$\mathcal{L} = -\frac{1}{N} \sum_{i=1}^N [y_i \log p_i + (1 - y_i) \log(1 - p_i)], \quad (1)$$

where the linear combination of standardized linguistic features is transformed into a probability through the sigmoid function $p_i = \sigma(\beta^\top x_i)$, x_i denotes the feature vector, and $y_i = 1$ if the news is real and $y_i = 0$ otherwise.

XGBoost / LightGBM Models. The detection of fake news often relies on nonlinear relationships and interaction effects among multiple features rather than on isolated feature contributions [32]. XGBoost and LightGBM are typical gradient-boosting algorithms that construct tree ensembles sequentially. Each tree is fitted to reduce the errors of the preceding trees, allowing the models to capture complex nonlinear relationships and high-order interactions among linguistic features [47]. The binary cross-entropy loss defined in Equation 1 is used as the optimization objective.

Random Forest Model. Random Forest is a tree-based ensemble learning method that incorporates randomness during model construction [48]. The final prediction is obtained by aggregating multiple decision trees trained on randomly sampled subsets of data and features. During tree construction, the model selects the feature split that maximally reduces node impurity based on the Gini criterion:

$$G = 1 - \sum_{k=1}^K p_k^2, \quad (2)$$

where p_k denotes the proportion of samples belonging to class k within a node. In this study, Random Forest is used as a comparative model to examine the cross-model consistency of linguistic features and further assess their robustness [11].

MLP Model. As a typical neural network model, MLP can learn hierarchical representations and capture complex nonlinear dependencies in textual data [49]. Deep learning models are particularly effective at modeling semantic patterns, subjective expressions, and feature interactions that are difficult to capture using traditional linear approaches [50]. This advantage plays a critical role in fake news identification, given that disinformation commonly consists of intricate mixes of affective features, subjective expressions, narrative styles, and semantic conflicts.

The MLP takes linguistic features as inputs and captures nonlinear relationships among features through successive nonlinear mappings. The model is trained using the loss function defined in Equation 1.

4.2 Cross-model Feature Attribution and Consistency Analysis

Since the importance of linguistic features varies across different models, we adopt SHAP values to quantify and interpret marginal feature contributions across all models, which is selecting the appropriate implementation method according to the model structure. Specifically, KernelSHAP is used for Logistic Regression, TreeSHAP for tree-based models, and DeepSHAP for the MLP model.

It should be noted that although the methods for computing SHAP values differ, they are based on the same feature attribution principle, share a consistent interpretive logic, and provide interpretation results on a unified scale. Figure 2 presents the detailed analytical process for interpreting the linguistic features of fake news using multiple machine learning models and SHAP-based feature attribution.

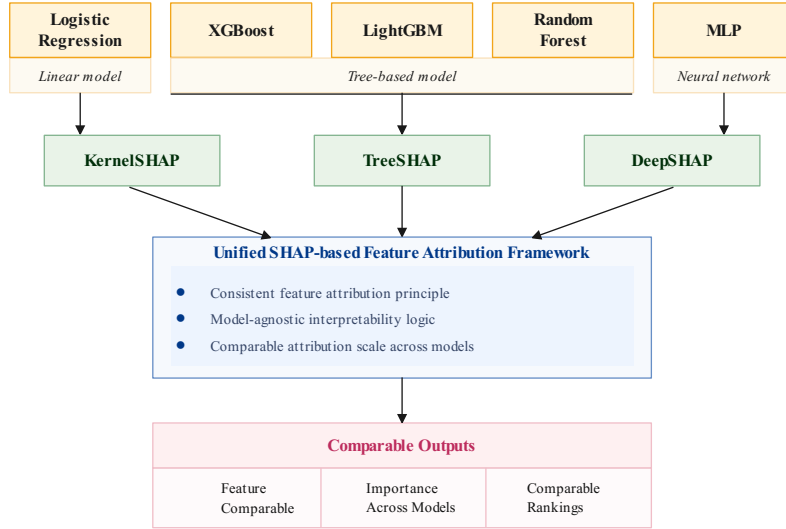


Fig. 2 SHAP-based feature attribution framework.

In this study, SHAP is not adopted for concrete predictive modeling. Instead, it serves as a unified interpretive approach. The primary aim of this paper is not to compare the internal mechanisms of different algorithms, but to examine whether linguistic feature importance remains consistent across models under a unified explanatory scale.

To identify the intrinsic linguistic features in fake news, cross-model consistency analysis is performed at the feature interpretation step. If several features are captured only by chance due to a particular model’s structure, their importance rankings often present significant differences. Nevertheless, our core goal is to construct intrinsic fake news features that do not rely on specific model architectures. Thus, we compare feature importance rankings across different SHAP-based models to identify the inherent features of fake news. Specifically, we adopt correlation analysis to quantify the consistency of linguistic features for fake news among models. High correlation coefficients

imply that the extracted fake news features are nature and independent of particular models or algorithms.

5 Results

5.1 Feature Contrast Between Real and Fake News

For verification of the inherent linguistic features of fake news, the same method is applied to extract features from real news for comparison. To better distinguish the linguistic features of real and fake news, each element in their respective feature vectors is weighted, and the weights are normalized using the softmax function, such that more distinctive features are assigned higher weights.

Subsequently, the cosine similarity is computed via the dot product of the weighted feature vectors to measure the feature differences between real and fake news, while the Euclidean distance is further calculated to quantify their linguistic discrepancy.

As expected, the cosine similarity reaches -1.000 for both 103 and 24 features. This implies that real and fake news exhibit nearly opposite linguistic structures in the feature space. Moreover, the 24-feature representation shows a larger Euclidean distance than the original 103-feature representation. This indicates that the aggregated high-level features can distinguish real and fake news more clearly.

Figure 3 further illustrates the structural differences between real and fake news across two types of feature representation, with 103 original linguistic features shown on the left and 24 integrated features on the right.

The results clearly show that real and fake news exhibit distinct linguistic feature distributions, suggesting that fake news possesses different linguistic characteristics from real news. These findings provide further evidence for the validity of the fake news features presented in Table 1 and indicate that these distinctive features are more strongly embedded in fake news than in real news.

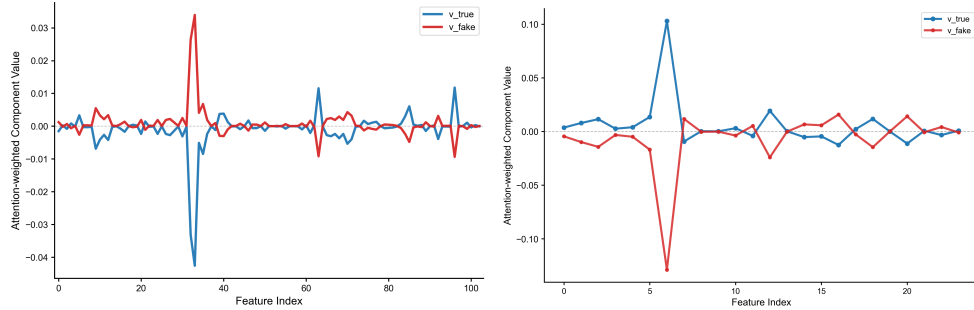


Fig. 3 Comparison of attention-weighted center vectors for true and fake news: (a) original 103 linguistic feature metrics and (b) aggregated 24 linguistic features.

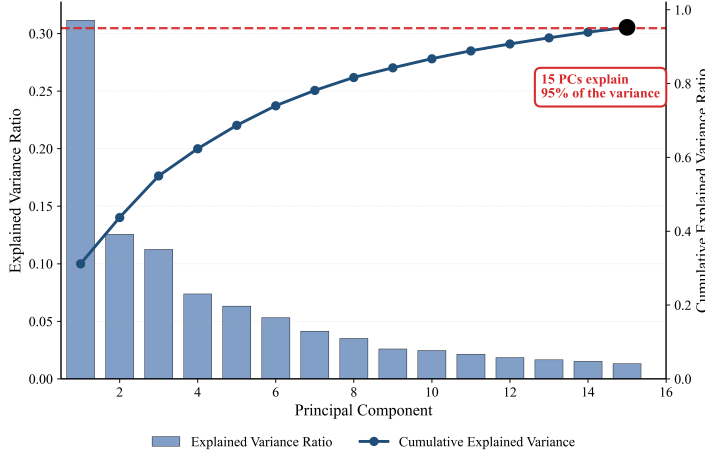


Fig. 4 Explained and cumulative explained variance ratios for PCA.

5.2 Spearman Correlation between PCA and SHAP Rankings

We first examine the linguistic features of fake news using PCA. Figure 4 presents the explained variance ratios of the first ten principal components and the cumulative explained variance ratio. The explained variance ratios of the principal components do not exhibit any significant inflection points, indicating that the variance information is dispersed across multiple principal components and does not form a clear low-dimensional structure. At the same time, the cumulative explained variance results show that 15 principal components need to be retained from the 24 original features for the cumulative explained variance to exceed 95%. Therefore, PCA struggles to achieve effective dimensionality reduction for the proposed linguistic feature set and is not designed to provide interpretable feature importance.

Table 2 presents the correlations between the PCA rankings and the SHAP rankings obtained from the machine learning models. The Spearman rank correlation coefficients are all close to 0. Specifically, the correlation coefficients between PCA and MLP, LightGBM, XGBoost, and Logistic Regression are all negative. Although Random Forest shows a positive correlation, the value is low ($\rho = 0.117$). These results indicate that the feature importance rankings derived from PCA and SHAP are not consistent with each other.

Thus, we can claim that PCA is not suitable as the primary interpretive tool for language features in fake news. One could explain that PCA achieves dimensionality reduction through linear combinations of the original features, weakening the correspondence between the original linguistic features and their semantic interpretation.

Compared with PCA, SHAP reveals the marginal contribution of individual features and preserves their inherent semantic information. Therefore, SHAP offers greater feasibility and practicality in the analysis of feature importance [29, 51].

Table 2 Feature correlation between PCA and SHAP rankings.

Model	Spearman ρ
MLP	-0.069
LightGBM	-0.220
XGBoost	-0.256
Logistic Regression	-0.276
Random Forest	0.117

5.3 SHAP Feature Importance Rankings

Figure 5 shows the SHAP feature importance rankings for five machine learning models. The SHAP values represent the average marginal contribution of each feature to the prediction. A higher value indicates a more significant influence on the model’s decision. Meanwhile, we can observe the most dominant features of fake news and further analyze the overlaps of major features across the five machine learning models. To compare the consistency of importance rankings across different models, we calculate the Spearman rank correlation coefficients for each model.

Table 3 Spearman rank correlations among model feature importance rankings.

	XGBoost	LightGBM	LogReg	MLP	RF
XGBoost	1	0.977	0.749	0.741	0.037
LightGBM	0.977	1	0.703	0.693	-0.013
LogReg	0.749	0.703	1	0.794	0.154
MLP	0.741	0.693	0.794	1	0.208
RF	0.037	-0.013	0.154	0.208	1

Table 3 shows that XGBoost and LightGBM exhibit the highest consistency ($\rho = 0.977$), likely because both models employ similar boosting strategies. Logistic Regression and the tree-based boosting models also show moderate consistency ($\rho \approx 0.7$). In contrast, Random Forest exhibits relatively low consistency with the other models, which may be related to fluctuations in feature importance caused by its random sampling and splitting mechanisms.

Overall, most models demonstrate a relatively high level of consistency in identifying the key linguistic features of fake news. This suggests that the constructed linguistic feature system possesses good structural stability.

5.4 Core Linguistic Feature Identification

To identify the core linguistic features of fake news that remain consistent across different models, we analyze the importance of each feature across multiple machine

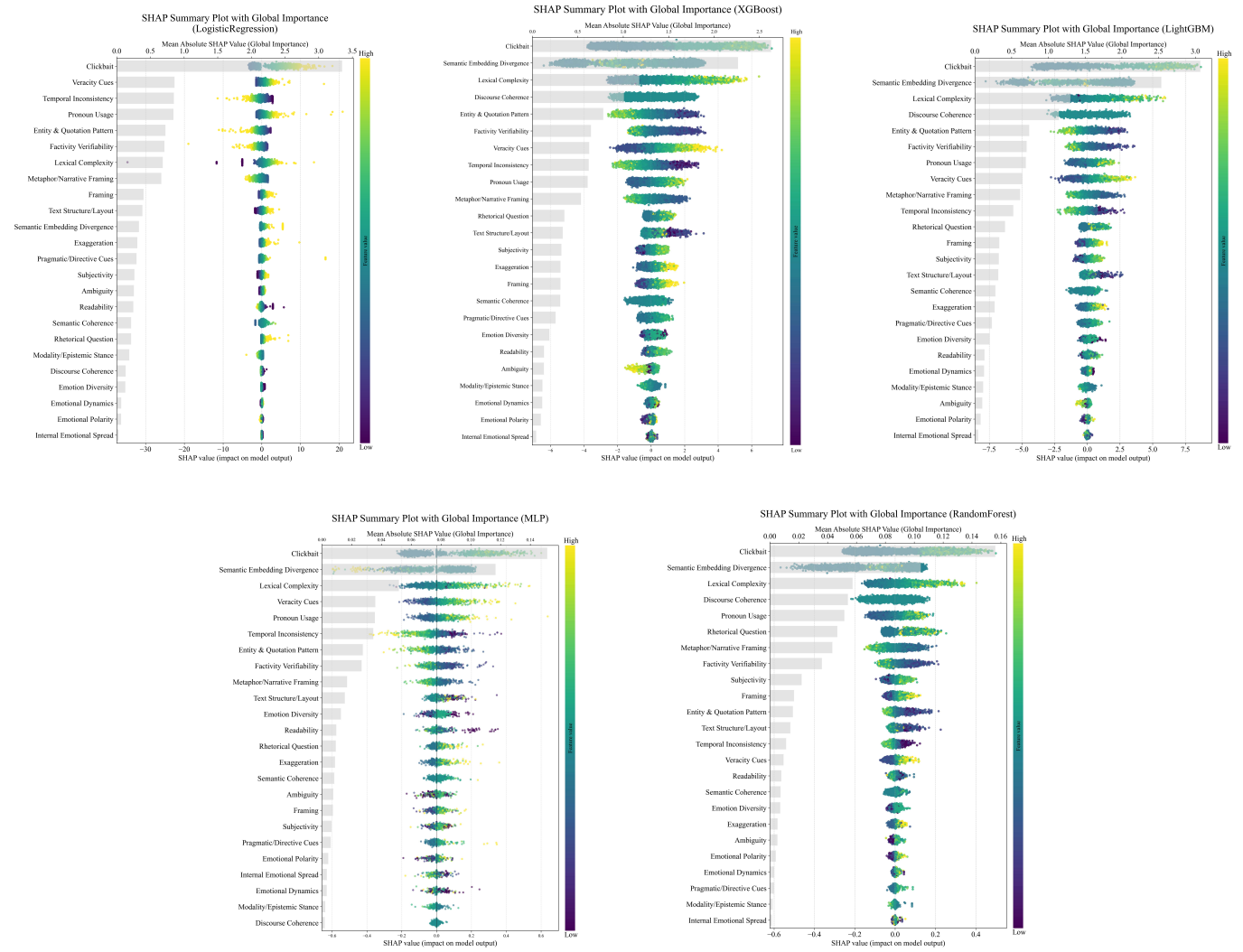


Fig. 5 SHAP summary plots of linguistic feature importance across five machine learning models.

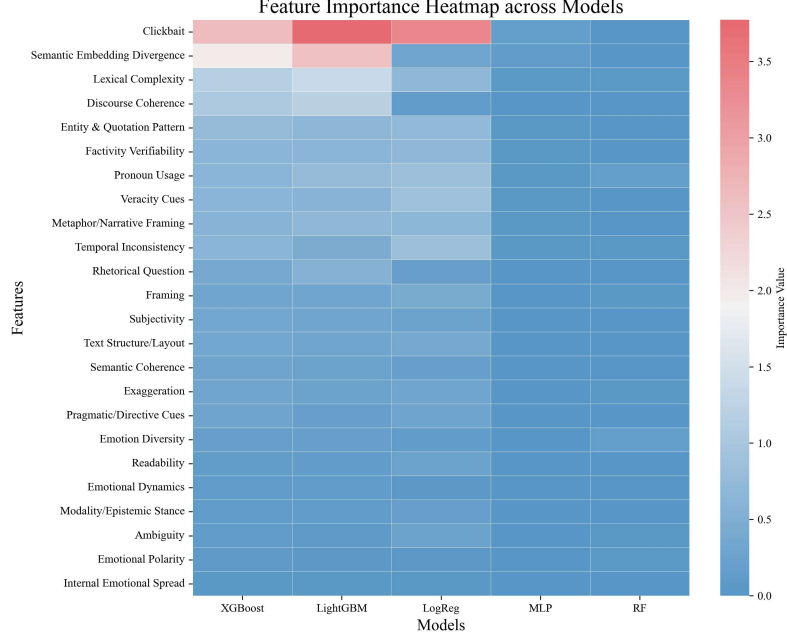


Fig. 6 Heatmap of linguistic feature importance across five machine learning models.

learning models. Together with the results reported in Table 3, we adopt a comparative and interpretive approach to identify linguistic features that demonstrate high and relatively consistent importance across XGBoost, LightGBM, LogReg, and MLP.

Figure 6 presents the mean absolute SHAP values of the 24 linguistic features across five models. The rows represent linguistic features, while the columns represent models. The color gradient indicates the relative importance of each feature, ranging from the most important to the least important (from red to blue).

Based on the results shown in Figure 6, we identify eight core linguistic features of fake news, which can be divided into three categories. Figure 7 illustrates the three-category organization of these eight core linguistic features. The first category corresponds to linguistic style features, capturing the stylistic attributes of fake news, including Clickbait, Pronoun Usage, and Lexical Complexity. The second category reflects information complexity characteristics, consisting of Semantic Embedding Divergence, Factivity Verifiability, and Veracity Cues, which jointly characterize the ambiguity and complexity of news content. The third category comprises discourse and narrative structure features, including Discourse Coherence and Entity & Quotation Pattern, which characterize the textual organization and narrative structure of news content.

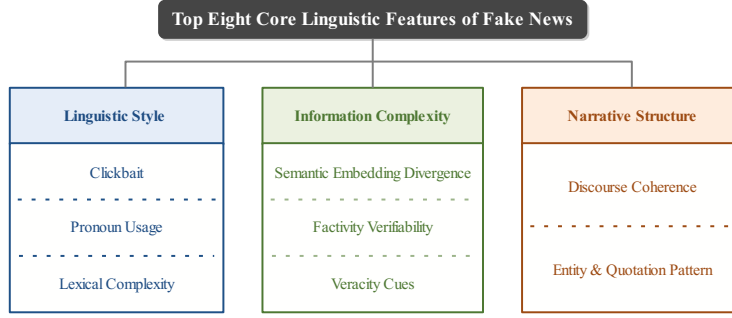


Fig. 7 Top eight core linguistic features of fake news.

6 Discussion

This section discusses how the identified core linguistic features shape the creation and dissemination of fake news from the perspectives of news production, cognition and discourse.

6.1 Linguistics and Manipulation of Fake News

Empirical findings indicate that fake news demonstrates systematic linguistic characteristics rather than random expressions. This reflects consistent communicative goals, recurring linguistic patterns, and the strategic imitation of journalistic conventions. Based on this, the following analysis discusses how fake news misleads audiences through three aspects: communicative purpose, linguistic mechanisms, and the manipulation of norms.

First, despite variations in platforms, contexts, and audiences, fake news adopts similar linguistic manipulation strategies in its dissemination. To attract attention, influence audience judgment, reduce fact-checking efforts, manipulate emotions, and create social conflict [25, 52, 53], fake news consistently employs specific linguistic features to enhance its dissemination and persuasive effects. Previous research has shown that these objectives can be achieved through linguistic and semantic manipulation strategies [21].

Second, the key linguistic features identified in this study are highly consistent across different models, suggesting that these features constitute stable linguistic cues. According to the Elaboration Likelihood Model [54], individuals switch between the central route and the peripheral route during information processing. Audiences tend to rely on linguistic cues to make rapid judgments in environments of information overload and limited attention. Consequently, information processing and judgment are easily guided by these linguistic features, independent of the specific news context.

Third, the production of fake news appears to be increasingly sophisticated and professionalized. Fake news is less likely to manifest as incoherent language or unusual phrasing. Instead, it tends to mimic the format and linguistic features of real news.

This reduces audience vigilance and undermines their ability to assess the authenticity of news [22, 55–57].

6.2 Feature and Mechanisms of Fake News

The eight core features interact synergistically across the entire information processing pipeline, rather than functioning in isolation. Cognitive and discursive theories help to account for these patterns. Additionally, we can explain these patterns from the perspectives of information intake, narrative structure, and judgment.

6.2.1 Attention and Information Intake Mechanisms

Both Clickbait and Lexical Complexity influence audience choices during the information intake stage. This stage determines whether information can move forward to dissemination and comprehension via attention capture and click generation.

Selective Exposure Theory suggests that audiences are not passive recipients in the process of information selection. Instead, they actively engage in selective exposure under conditions of limited attention and cognitive resources. This process takes place prior to deep comprehension, at the stage of information intake [18]. In today’s information-rich digital landscape, attention has become a scarce resource, and competition between information sources is increasingly focused on winning audience attention [16, 17]. Thus, Clickbait uses information gaps as a deliberate strategy. Incomplete and emotionally charged headlines spark curiosity and the urge to click, increasing the likelihood that audiences will select fake news.

In contrast, Lexical Complexity does not simply reflect text difficulty. Instead, it shapes expectations about comprehensibility and informational value, influencing whether audiences continue reading. This process aligns with the psychological mechanism of cognitive ease: information that is easier to process and imposes lower cognitive load may therefore be judged as more truthful or credible [58, 59].

6.2.2 Semantic Coherence and Narrative Credibility Mechanisms

As reflected in Semantic Embedding Divergence, Discourse Coherence, and Entity & Quotation Patterns, fake news imitates real news styles to appear more credible to audiences.

Semantic Embedding Divergence shows that fake news may appear coherent and plausible on the surface, while its underlying logic departs from objective facts. Previous research suggests that audiences rely on interpretive frames to make sense of information [60]. Thus, a consistent narrative frame may help fake news obscure factual contradictions and appear more credible.

Discourse Coherence embodies structural imitation in narratives. To gain credibility, fake news often imitates the narrative structure and linguistic norms of real news [22]. There are two dimensions of coherence: local and global. The first describes intra-text cohesion, and the second measures the conformity of content to real-world situations [61, 62]. Audiences first establish local coherence when making judgments. As a result, even when content deviates from reality, fake news can still appear credible through local coherence.

Entity & Quotation Pattern captures how authority cues are constructed in fake news. Real news generally builds authority and factuality through named entities, quotation formats and event organization [63]. Fake news adopts the same tactic by featuring experts, institutions, and public figures as authoritative cues. This conveys authority signals and reduces audience skepticism about content authenticity [64].

6.2.3 Factivity, Stance, and Group Identity Mechanisms

Factivity Verifiability, Veracity Cues, and Pronoun Usage reduce audience skepticism and guide audience judgment through factual expressions, veracity cues, and group-oriented pronoun use.

Factivity Verifiability does not refer to the actual truthfulness or verification of information, but rather to the linguistic construction of factuality. To mimic real news, fake news may use high-certainty expressions, thereby reducing audience doubt [57]. Previous research has shown that epistemic certainty can be expressed through specific linguistic forms [65]. As a result, audiences may equate expressions of certainty with factual validity and may therefore be more likely to accept the information as true while conducting fewer follow-up checks on its authenticity.

Veracity Cues draw on pseudo-objective, specific, and technical details, rather than genuine improvements in content actual truthfulness. Such details include precise references to time, location, numbers, or technical terms. When sufficient cues are present, audiences are less likely to engage in systematic processing. Instead, they default to a heuristic mode of processing and perceive the information as credible [19, 66].

Pronoun Usage frames interpretation by linking content to group identity via specific referents. Pronouns function as linguistic resources for identity construction, reinforcing boundaries between in-group and out-group members [67]. This form of identity-based language plays a key role in the spread of fake news [68]. It not only increases audience engagement but may also facilitate dissemination. As a result, Pronoun Usage can shift audience judgment from factual evaluation toward identity-based evaluation, where individuals rely more on group affiliation than on content accuracy.

6.3 Practical Implications

The findings of this study indicate that fake news is not merely random linguistic noise, but is instead characterized by a set of reproducible and systematically identifiable features. This further implies that the analysis of fake news should not be limited to text classification, but also account for the underlying discursive processes and the cognitive shaping of audiences.

Manipulation in fake news often occurs before fact-checking mechanisms are activated. It can shape judgments of content authenticity at the information intake stage. This suggests that the governance of fake news should not rely solely on *post hoc* fact-checking but may benefit from shifting earlier to the stage of information entry. This shift can enhance the recognition of linguistic patterns and increase awareness of potential manipulative mechanisms.

Overall, by identifying inherent linguistic features, this study highlights the underlying structural logic of fake news. These findings suggest that fake news is not

composed of random errors but rather constitutes a communicative phenomenon that can be systematically analyzed, theoretically explained, and effectively governed.

Based on the results above, the practical applications of the identified linguistic features can be further considered in multiple contexts.

One important implication is that these linguistic features provide a foundation for the development of automated tools for detecting fake news. Such tools can be integrated into news or social media platforms to support content moderation. This enables earlier detection of potentially misleading content and reduces the risk of its further spread.

Another important implication concerns media literacy. These features can be incorporated into media literacy initiatives. They enable audiences to recognize recurring linguistic patterns in fake news, which promotes critical awareness. They could also provide professional training for journalists and media practitioners, improving their ability to identify and respond to misleading content.

A further implication lies in public policy and platform governance. These features can also serve as a reference for public policy and platform governance. They enable the development of more fine-grained evaluation frameworks and targeted intervention mechanisms, facilitating a shift from reactive content moderation to proactive risk detection. This contributes to the development of more refined and targeted strategies for fake news mitigation.

7 Conclusion

Based on communication, linguistics, and psychology, we first extract the inherent features of fake news by adopting statistical methods, lexicon-based models, and NLP techniques. We further analyze these features with the combination of SHAP values and machine learning models. The empirical results show that XGBoost, LightGBM, LogReg and MLP have high consistency in feature importance ranking.

Using SHAP analysis, we extract eight core linguistic features of fake news, including Clickbait, Semantic Embedding Divergence, Lexical Complexity, Discourse Coherence, Entity & Quotation Pattern, Factivity Verifiability, Pronoun Usage and Veracity Cues. These features of fake news, which fall into three major categories (information complexity, linguistic style and narrative structure), jointly mislead audiences throughout information acquisition, narrative comprehension and cognitive evaluation.

The core linguistic patterns summarized in this paper can provide a clear reference for the design and optimization of fake news detection algorithms. By focusing on these validated core features, one can reduce feature dimensionality and eliminate redundant variables. This contributes to developing more targeted and effective strategies for fake news mitigation.

This study also suggests that the governance of fake news should pay greater attention to linguistic mechanisms at the information entry stage. This measure can raise public awareness and strengthen people’s vigilance against disinformation, thereby reducing its impact on audience cognition and judgment.

Due to the limitations of the research sample, the generalizability of the findings needs further verification. In future work, we will expand the dataset and explore the dynamic changes of fake news linguistic features and interaction mechanisms.

Data Availability. The dataset analysed during the current study is publicly available in the Kaggle repository, Fake and Real News Dataset (ISOT Fake News Detection Dataset), at <https://www.kaggle.com/datasets/clmentbisailon/fake-and-real-news-dataset>.

References

- [1] Lazer, D. M. J. *et al.* The science of fake news. *Science* **359**, 1094–1096 (2018).
- [2] Rodríguez-Ferrándiz, R. An overview of the fake news phenomenon: From untruth-driven to post-truth-driven approaches. *Media and Communication* **11**, 15–29 (2023). URL <https://www.cogitatiopress.com/mediaandcommunication/article/view/6315>.
- [3] Hamby, A., Kim, H. & Spezzano, F. Sensational stories: The role of narrative characteristics in distinguishing real and fake news and predicting their spread. *Journal of Business Research* **170**, 114289 (2024). URL <https://doi.org/10.1016/j.jbusres.2023.114289>.
- [4] Aich, A., Bhattacharya, S. & Parde, N. Calzolari, N. *et al.* (eds) *Demystifying neural fake news via linguistic feature-based interpretation*. (eds Calzolari, N. *et al.*) *Proceedings of the 29th International Conference on Computational Linguistics*, 6586–6599 (International Committee on Computational Linguistics, Gyeongju, Republic of Korea, 2022). URL <https://aclanthology.org/2022.coling-1.573/>.
- [5] Ghanem, B., Rosso, P. & Rangel, F. An emotional analysis of false information in social media and news articles. *ACM Transactions on Internet Technology* **20**, 1–18 (2020). URL <https://doi.org/10.1145/3381750>.
- [6] Hosseini, A. S. & Staab, S. Emotional framing in the spreading of false and true claims. 15th ACM Web Science Conference 2023 (WebSci '23), pp. 96–106, Association for Computing Machinery (2023). URL <https://doi.org/10.1145/3578503.3583611>.
- [7] Huang, L. Deep learning for fake news detection: Theories and models. *Proceedings of the 2022 6th International Conference on Electronic Information Technology and Computer Engineering (EITCE '22)*, pp. 1322–1326, Association for Computing Machinery (2023). URL <https://doi.org/10.1145/3573428.3573663>.
- [8] Jacovi, A. & Goldberg, Y. Aligning faithful interpretations with their social attribution. *Transactions of the Association for Computational Linguistics* **9**, 294–310 (2021). URL <https://aclanthology.org/2021.tacl-1.18/>.

- [9] Gongane, V. U., Munot, M. V. & Anuse, A. D. A survey of explainable ai techniques for detection of fake news and hate speech on social media platforms. *Journal of Computational Social Science* **7**, 587–623 (2024).
- [10] Guidotti, R. *et al.* A survey of methods for explaining black box models. *ACM Computing Surveys* **51** (2018). URL <https://doi.org/10.1145/3236009>.
- [11] Molnar, C. *Interpretable Machine Learning* (Leanpub, 2020). URL <https://christophm.github.io/interpretable-ml-book/>.
- [12] Saarela, M. & Jauhiainen, S. Comparison of feature importance measures as explanations for classification models. *SN Applied Sciences* **3**, 272 (2021).
- [13] Vosoughi, S., Roy, D. & Aral, S. The spread of true and false news online. *Science* **359**, 1146–1151 (2018).
- [14] Ecker, U. K. H. *et al.* The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology* **1**, 13–29 (2022).
- [15] Pennycook, G. & Rand, D. G. The psychology of fake news. *Trends in Cognitive Sciences* **25**, 388–402 (2021).
- [16] Davenport, T. H. & Beck, J. C. *The Attention Economy: Understanding the New Currency of Business* (Harvard Business School Press, Boston, MA, 2001). URL <https://books.google.com.hk/books?id=j6z-MiUKgosC>.
- [17] Webster, J. G. *The Marketplace of Attention: How Audiences Take Shape in a Digital Age* (The MIT Press, Cambridge, MA, 2014). URL <https://doi.org/10.7551/mitpress/9892.001.0001>.
- [18] Zillmann, D. & Bryant, J. (eds) *Selective Exposure to Communication* 1 edn (Routledge, 1985).
- [19] Pennycook, G. & Rand, D. G. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* **188**, 39–50 (2019). URL <https://doi.org/10.1016/j.cognition.2018.06.011>.
- [20] Pérez-Rosas, V., Kleinberg, B., Lefevre, A. & Mihalcea, R. Bender, E. M., Derczynski, L. & Isabelle, P. (eds) *Automatic detection of fake news*. (eds Bender, E. M., Derczynski, L. & Isabelle, P.) *Proceedings of the 27th International Conference on Computational Linguistics*, 3391–3401 (Association for Computational Linguistics, Santa Fe, New Mexico, USA, 2018). URL <https://aclanthology.org/C18-1287/>.
- [21] Moyseyenko, I., Odobetska, I., Kovalenko, A., Mozalov, V. & Rud, O. Analysing fake news through linguistics: Detecting manipulation tactics. *Metaverse Basic*

- and *Applied Research* **4**, 162 (2025).
- [22] Grieve, J. & Woodfield, H. *The Language of Fake News* Elements in Forensic Linguistics (Cambridge University Press, Cambridge, 2023). URL <https://doi.org/10.1017/9781009349161>.
 - [23] Sharma, K. *et al.* Combating fake news: A survey on identification and mitigation techniques. *ACM Transactions on Intelligent Systems and Technology* **10**, 21:1–21:42 (2019). URL <https://doi.org/10.1145/3305260>.
 - [24] Zhou, X. & Zafarani, R. A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys* **53**, 1–40 (2020). URL <https://doi.org/10.1145/3395046>.
 - [25] Aïmeur, E., Amri, S. & Brassard, G. Fake news, disinformation and misinformation in social media: A review. *Social Network Analysis and Mining* **13**, 30 (2023). URL <https://doi.org/10.1007/s13278-023-01028-5>.
 - [26] Li, X., Xiong, H., Huang, S., Ji, S. & Dou, D. Cross-model consensus of explanations and beyond for image classification models: An empirical study. *Machine Learning* **112**, 1627–1662 (2023).
 - [27] Lundberg, S. M. & Lee, S.-I. Guyon, I. *et al.* (eds) *A unified approach to interpreting model predictions*. (eds Guyon, I. *et al.*) *Advances in Neural Information Processing Systems 30*, 4765–4774 (Curran Associates, Inc., 2017).
 - [28] Lundberg, S. M. *et al.* From local explanations to global understanding with explainable ai for trees. *Nature Machine Intelligence* **2**, 56–67 (2020).
 - [29] Wang, H., Liang, Q., Hancock, J. T. & Khoshgoftaar, T. M. Feature selection strategies: A comparative analysis of shap-value and importance-based methods. *Journal of Big Data* **11**, 44 (2024). URL <https://doi.org/10.1186/s40537-024-00905-w>.
 - [30] Madsen, A., Reddy, S. & Chandar, S. Post-hoc interpretability for neural nlp: A survey. *ACM Computing Surveys* **55**, 1–42 (2022). URL <https://doi.org/10.1145/3546577>.
 - [31] Murayama, T. Dataset of fake news detection and fact verification: A survey (2021). URL <https://arxiv.org/abs/2111.03299>. arXiv:2111.03299.
 - [32] Shu, K., Sliva, A., Wang, S., Tang, J. & Liu, H. Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter* **19**, 22–36 (2017). URL <https://doi.org/10.1145/3137597.3137600>.

- [33] Zhang, J., Spirling, A. & Danescu-Niculescu-Mizil, C. Palmer, M., Hwa, R. & Riedel, S. (eds) *Asking too much? the rhetorical role of questions in political discourse*. (eds Palmer, M., Hwa, R. & Riedel, S.) *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 1558–1572 (Association for Computational Linguistics, Copenhagen, Denmark, 2017). URL <https://doi.org/10.18653/v1/D17-1164>.
- [34] Karimi, H. & Tang, J. Burstein, J., Doran, C. & Solorio, T. (eds) *Learning hierarchical discourse-level structure for fake news detection*. (eds Burstein, J., Doran, C. & Solorio, T.) *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT 2019), Volume 1 (Long and Short Papers)*, 3432–3442 (Association for Computational Linguistics, Minneapolis, Minnesota, 2019). URL <https://aclanthology.org/N19-1347/>.
- [35] Qi, P. *et al.* Improving fake news detection by using an entity-enhanced framework to fuse diverse multimodal clues. *Proceedings of the 29th ACM International Conference on Multimedia (MM '21)*, pp. 1212–1220, Association for Computing Machinery (2021). URL <https://doi.org/10.1145/3474085.3481548>.
- [36] Jeronimo, C. L. M., Marinho, L. B., Campelo, C. E. C., Veloso, A. & da Costa Melo, A. S. Characterization of fake news based on subjectivity lexicons. *Journal of Data Intelligence* **1**, 419–441 (2020). URL <https://doi.org/10.26421/JDI1.4-2>.
- [37] Pérez-Montero, A. Analysing linguistic markers on fake news to enhance the explainability of deception detection systems. *Proceedings of the Doctoral Symposium on Natural Language Processing (NLP-DS 2024), CEUR Workshop Proceedings*, Vol. 3797 (2024). URL <https://ceur-ws.org/Vol-3797/paper3.pdf>.
- [38] Hauch, V., Blandón-Gitlin, I., Masip, J. & Sporer, S. L. Are computers effective lie detectors? a meta-analysis of linguistic cues to deception. *Personality and Social Psychology Review* **19**, 307–342 (2015). URL <https://doi.org/10.1177/1088868314556539>.
- [39] Santos, R. *et al.* Calzolari, N. *et al.* (eds) *Measuring the impact of readability features in fake news detection*. (eds Calzolari, N. *et al.*) *Proceedings of the Twelfth Language Resources and Evaluation Conference (LREC 2020)*, 1404–1413 (European Language Resources Association (ELRA), Marseille, France, 2020). URL <https://aclanthology.org/2020.lrec-1.176/>.
- [40] Tandoc Jr., E. C., Lim, Z. W. & Ling, R. Defining “fake news”: A typology of scholarly definitions. *Digital Journalism* **6**, 137–153 (2018).
- [41] Hutto, C. J. & Gilbert, E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. *Proceedings of the International AAAI Conference on Web and Social Media* **8**, 216–225 (2014). URL <https://doi.org/>

10.1609/icwsm.v8i1.14550.

- [42] Aljedaani, W. *et al.* Sentiment analysis on twitter data integrating TextBlob and deep learning models: The case of US airline industry. *Knowledge-Based Systems* **255**, 109780 (2022). URL <https://doi.org/10.1016/j.knosys.2022.109780>.
- [43] Salton, G. & Buckley, C. Term-weighting approaches in automatic text retrieval. *Information Processing & Management* **24**, 513–523 (1988). URL [https://doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0).
- [44] Reimers, N. & Gurevych, I. Inui, K., Jiang, J., Ng, V. & Wan, X. (eds) *Sentence-BERT: Sentence embeddings using siamese BERT-networks*. (eds Inui, K., Jiang, J., Ng, V. & Wan, X.) *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, 3982–3992 (Association for Computational Linguistics, Hong Kong, China, 2019). URL <https://doi.org/10.18653/v1/D19-1410>.
- [45] Jolliffe, I. T. *Principal Component Analysis* 2 edn. Springer Series in Statistics (Springer, New York, NY, 2002).
- [46] Jurafsky, D. & Martin, J. H. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition, with Language Models* 3rd edn (Stanford University, 2026). URL <https://web.stanford.edu/~jurafsky/slp3/>. Online manuscript released January 6, 2026.
- [47] Friedman, J. H. Greedy function approximation: A gradient boosting machine. *Annals of Statistics* **29**, 1189–1232 (2001).
- [48] Breiman, L. Random forests. *Machine Learning* **45**, 5–32 (2001).
- [49] Goldberg, Y. *Neural Network Methods for Natural Language Processing* 1 edn. Synthesis Lectures on Human Language Technologies (Springer, Cham, 2017).
- [50] LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).
- [51] Kitanovski, A., Toshevskaa, M. & Mirceva, G. Distilbert and roberta models for identification of fake news. *Proceedings of the 46th MIPRO ICT and Electronics Convention (MIPRO)*, pp. 1102–1106, IEEE (2023). URL <https://doi.org/10.23919/MIPRO57284.2023.10159740>.
- [52] Denniss, E. & Lindberg, R. Social media and the spread of misinformation: Infectious and a threat to public health. *Health Promotion International* **40**, daaf023 (2025). URL <https://doi.org/10.1093/heapro/daaf023>.

- [53] Celliers, M. & Hattingh, M. Hattingh, M. *et al.* (eds) *A systematic review on fake news themes reported in literature.* (eds Hattingh, M. *et al.*) *Responsible Design, Implementation and Use of Information and Communication Technology*, Vol. 12067 of *Lecture Notes in Computer Science*, 223–234 (Springer International Publishing, Cham, 2020). URL https://doi.org/10.1007/978-3-030-45002-1_19.
- [54] Petty, R. E. & Cacioppo, J. T. *Communication and Persuasion: Central and Peripheral Routes to Attitude Change* 1 edn. Springer Series in Social Psychology (Springer, New York, NY, 1986). URL <https://doi.org/10.1007/978-1-4612-4964-1>.
- [55] Ireton, C. & Posetti, J. (eds) *Journalism, Fake News & Disinformation: Handbook for Journalism Education and Training* UNESCO's Series on Journalism Education (UNESCO, Paris, 2018).
- [56] Tandoc Jr., E. C., Thomas, R. J. & Bishop, L. What is (fake) news? analyzing news values (and more) in fake stories. *Media and Communication* **9**, 110–119 (2021). URL <https://www.cogitatiopress.com/mediaandcommunication/article/view/3331>.
- [57] Egelhofer, J. L. & Lecheler, S. Fake news as a two-dimensional phenomenon: A framework and research agenda. *Annals of the International Communication Association* **43**, 97–116 (2019). URL <https://doi.org/10.1080/23808985.2019.1602782>.
- [58] Alter, A. L. & Oppenheimer, D. M. Uniting the tribes of fluency to form a metacognitive nation. *Personality and Social Psychology Review* **13**, 219–235 (2009). URL <https://doi.org/10.1177/1088868309341564>.
- [59] Reber, R., Schwarz, N. & Winkielman, P. Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review* **8**, 364–382 (2004). URL <https://doi.org/10.1207/S15327957PSPR0804.3>.
- [60] Goffman, E. *Frame Analysis: An Essay on the Organization of Experience* (Harper & Row, New York, NY, 1974).
- [61] van Dijk, T. A. *Text and Context: Explorations in the Semantics and Pragmatics of Discourse* (Longman, London, 1977).
- [62] van Dijk, T. A. *Macrostructures: An Interdisciplinary Study of Global Structures in Discourse, Interaction, and Cognition* 1 edn (Routledge, New York, NY, 1980). URL <https://doi.org/10.4324/9780429025532>.
- [63] van Dijk, T. A. *News As Discourse* 1 edn (Routledge, New York, NY, 1988). URL <https://doi.org/10.4324/9780203062784>.

- [64] Cialdini, R. B. *Influence: Science and Practice* 5 edn (Pearson Education, Boston, MA, 2009).
- [65] Chafe, W. L. & Nichols, J. (eds) *Evidentiality: The Linguistic Coding of Epistemology* Advances in Discourse Processes (Ablex Publishing Corporation, Norwood, NJ, 1986).
- [66] Hasher, L., Goldstein, D. & Toppino, T. Frequency and the conference of referential validity. *Journal of Verbal Learning and Verbal Behavior* **16**, 107–112 (1977). URL [https://doi.org/10.1016/S0022-5371\(77\)80012-1](https://doi.org/10.1016/S0022-5371(77)80012-1).
- [67] Hamed, L. A. A. & Behnam, B. Linguistic devices of identity representation in english political discourse with a focus on personal pronouns: Power and solidarity. *Journal of English Language Pedagogy and Practice* **13**, 205–229 (2020). URL <https://doi.org/10.30495/jal.2020.677911>.
- [68] Chen, K., Jin, Y. & Shao, A. Science factionalism: How group identity language affects public engagement with misinformation and debunking narratives on a popular q&a platform in china. *Social Media + Society* **8**, 1–15 (2022). URL <https://doi.org/10.1177/20563051221077019>.

Acknowledgements. This work was supported by the National Social Science Fund of China (Grant No. 22FXWA002).

Author Contributions Statement. Y.L. wrote the main manuscript text and conducted the data analysis. S.L. revised the manuscript. L.J. prepared the figures and tables. Y.L., S.L., and L.J. jointly developed the linguistic feature framework. All authors reviewed the manuscript.

Additional Information. Competing Interests The author(s) declare no competing interests.

Figure Legends. Figure 1. Three-level quantifiable linguistic feature framework for fake news.

Figure 2. SHAP-based feature attribution framework.

Figure 3. Comparison of attention-weighted center vectors for true and fake news.

(a) Original 103 linguistic feature metrics. (b) Aggregated 24 linguistic features.

Figure 4. Explained and cumulative explained variance ratios for PCA.

Figure 5. SHAP summary plots of linguistic feature importance across five machine learning models.

Figure 6. Heatmap of linguistic feature importance across five machine learning models.

Figure 7. Top eight core linguistic features of fake news. “

Table Legends. Table 1. Fake news linguistic features.

Table 2. Feature correlation between PCA and SHAP rankings.

Table 3. Spearman rank correlations among model feature importance rankings.