OPEN ACCESS
Jurnal Ilmiah Ilmu Terapan Universitas Jambi
Vol. 10 No. 2. April 2026, pp. 1023-1035 DOI. 10.22437/jiituj.v10i2.42675
Research Article

ARTIFICIAL INTELLIGENCE (AI) ASSISTED MULTIMODALITY IN DIGITAL CAMPAIGN TEXTS IN THE PRESIDENTIAL ELECTION

Dian Marisha Putri[1], Mulyadi[2], Amrin Saragih[3], Rahmadsyah Rangkuti[1], Yusni Khairul Amri[4,*]

[1] English Language Study Program, Universitas Sumatera Utara, Sumatera Utara, Indonesia
[2] Linguistics Study Program, Universitas Sumatera Utara, Indonesia
[3] English Language Study Program, Medan State University, Sumatera Utara, Indonesia
[4] Indonesian Language Education, Universitas Muhammadiyah Sumatera Utara, Sumatera Utara, Indonesia
Corresponding author email: yusnikhairulamri@umsu.ac.id

Article Info
Received: Feb 12, 2026
Revised: Mar 08, 2026
Accepted: Apr 16, 2026
OnlineVersion: Apr 30, 2026

Abstract
This study examines the multimodality of verbal and visual modes in 2024 digital campaign texts combining linguistic approaches with artificial intelligence technology. Verbal mode analysis involves the use of multiple modes in discourse. The use of AI helps identify patterns of visual-verbal relationships. The data and data sources are campaign texts for the 2024 General Election. Five social media platforms were used: WhatsApp, Facebook, Twitter, TikTok, and Instagram. Five online media platforms were selected: a) Detik.com, b) Kompas.com, c) Tribunnews.com, d) Liputan6.com, and e) cnbcindonesia.com. A total of 200 election texts were collected between May 2023 and February 2024. The collected data were qualitatively reviewed in the form of digital texts. The analysis results illustrate that AI-assisted multimodality in explaining the meaning of text in visual and verbal modes, such as: a) strengthening the meaning of the text. b) using a sarcastic style of language. c) The role of meaningful visual text has a relational nature: between illustrations and silhouettes, followed by the attributes of text as a material process. d) representing campaign text refers to how the semiotic system on objects and relationships outside the text and context, both directly and indirectly. The findings demonstrate that the visual language representing political campaign text serves as both an action and a reaction process, as well as a verbal and mental process, within the context of the situation.

Keywords: Artificial Intelligence, Campaign Text, Multimodality.

INTRODUCTION

The use of social media as a hub for political information during the presidential election marked a transformative moment in modern political communication. Information from social media became a jumble of viral news that then became trending topics. Ironically, digital media's primary function should be to provide the public with information about political parties; however, the information often marginalises candidates from other parties. In fact, this digital information usually influences readers to hate other candidates. Therefore, understanding the meaning of digital texts in a political campaign

requires a multimodal study, assisted by artificial intelligence, to comprehend the texts circulating in the community.

The development of digital technology and the use of artificial intelligence (AI) in text analysis have created a new landscape for conveying campaign messages. The digital space is no longer merely a forum for information exchange but also an arena for information competition, with intense competition for opinion and the construction of political images that require a multimodal understanding. In this context, research on multimodality assisted by artificial intelligence becomes highly relevant for understanding how meaning is constructed, naturalised, and reproduced through the combination of verbal and visual texts, as well as public responses, in the digital political system.

Multimodality is a term used to refer to the way people communicate using different modes simultaneously (Kress & van Leeuwen, 1996), defined as the use of several semiotic modes in the design of a product or semiotic event. In some ways, these modes are combined to reinforce, complement, or follow a particular order (Kress & van Leeuwen, 2001). Multimodality is the combination of two or more semiotic modes used simultaneously in communication. This statement is also made by Kress and Van Leeuwen (2001), who define multimodality as the use of different modes simultaneously in communication. The meaning of the mode is not limited to writing; it can also be conveyed through images, sounds, and other forms. Multimodality discourse is concerned with understanding language communication and representation that involves the interaction of various semiotic resources such as language, movement, gaze, and camera angle. Jewitt (2009) and O'Halloran (2011)

According to Kress and van Leeuwen (1996), multimodality refers to the way people communicate by using different modes simultaneously. Thus, multimodality is the study of how semiotic resources are exploited to convey communication messages. Further examining Kress (2009:). Fashion is a "socially shaped and culturally given resource for making meaning." Fashion is a resource whose meaning is shaped by the social and cultural factors in which it is used. These resources are not limited to language, but can also take the form of images, layout, speech and sound. The combination of these modes is often used to construct communication messages. Mode is a term in social semiotics that refers to the resources for meaning construction. From the definition, it can be understood that modes refer to multimodal resources organised for meaning-making (Jewitt, Bezemer and O'Halloran, 2016).

The delivery of messages through visual and verbal modes complements and supports each other. Therefore, interpretation of meaning cannot be done by only highlighting one of these modes. In social semiotics, in particular, the verbal and visual modes have distinct roles in representing the meaning or message of a multimodal discourse; this is attributed to the different affordances of the two modes (Kress & Jewitt, 2003). The verbal mode represents the narrative world, while the visual mode represents the displayed world. These modes are combined to reinforce, complement, or coexist in specific compositions (Kress & van Leeuwen, 2001).

Previous research on political campaigns on social media has used multimodal strategies to influence public emotions and perceptions (Kress & van Leeuwen, 2006; Machin, 2013). The use of visuals, colour, gestures, typography, and layout often plays a greater role in determining voter responses than argumentative verbal content (Lim, 2018). In the Indonesian context, research by Mahmudah (2021) found that political texts on Instagram rely more heavily on visuals to strengthen candidate legitimacy. However, most studies still focus on manual analysis, which limits their ability to capture hidden, large-scale discursive patterns. However, developments in artificial intelligence technology have opened up new opportunities for more accurate, systematic, and in-depth multimodality analysis.

According to Dallyono and Sukyadi (2019), environmental issues are crucial for preventing mass environmental damage through campaigns (e.g., posters). This impacts viewers' contemplation of these issues. With the study of multimodality in accordance with Halliday's systemic theory (1994: xiv-xxvi), the meaning of language as a semiotic system is interpreted as the combination of different signs (modes) that produce a meaning that can be understood. This SFL study is grounded in social semiotics, specifically in the analysis of images (Kress & Van Leeuwen, 1996; Martinec & Salway, 2005; Bezemer & Kress, 2008; Kress, 2011). Contemplation of meaning as an educational process is essential for building good character. This study highlights the significance of posters as a model for representing and conveying environmental messages through this multimodal medium.

This research is crucial given the rampant spread of distorted information, misinformation, hoaxes, and visual-verbal manipulation that occurred during the campaign period. Data from the Ministry of Communication and Information (Kominfo, 2023) showed an increase in the spread of political hoaxes leading up to the election, particularly through images, short videos, and memes that combine textual

content with persuasive visuals. These messages not only influence voters but also shape cognitive biases and create political polarisation. The use of multimodality in digital campaigns can convey meanings that are difficult for ordinary voters to grasp, especially because visuals and verbal content can mutually reinforce implied meanings such as sarcasm, insinuation, delegitimisation, or even insults. Therefore, this research emerges as an academic response to the need to understand the increasingly complex phenomenon of multimodal politicisation.

Furthermore, the use of artificial intelligence in multimodal analysis provides a strong rationale for this research. AI is effective at automatically detecting language, emotional, and visual-verbal patterns across large data volumes (LeCun et al., 2019). In an era of digital political campaigns characterised by big data, AI's ability to quickly identify relationships between data points is a significant advantage. Recent studies, such as Steiner (2022) and Yang & Jiang (2021), demonstrate that AI can capture multimodal nuances that conventional analysis cannot map, including specific visual repetition patterns, colour-based framing tendencies, and the precision of emotion detection in campaign texts.

Taking these findings into account, this study combines multimodal analysis with AI support to uncover how digital campaign texts for the 2024 Presidential Election construct meaning through visual and verbal interactions. Focusing on digital campaigns is crucial because the majority of young Indonesian voters consume political information through social media platforms such as Instagram, TikTok, and X (We Are Social, 2023). The multimodal nature of these platforms demands an approach that goes beyond observing texts in isolation and examines the simultaneous relationships between modes that shape the meaning-making experience.

The research problem is the lack of AI-based, multimodal research explicitly examining the Indonesian presidential election campaign. An alternative research approach offers AI-assisted multimodality in digital campaign texts for the presidential election. This is capable of mapping message structures, representation patterns, verbal-visual relationships, and the potential for meaning manipulation more comprehensively. Thus, this research not only fills the gap in existing research but also presents a novel approach by integrating multimodality and artificial intelligence to analyse digital political campaign texts critically. The research results are expected to contribute to the development of digital semiotics studies, political communication, and public digital literacy in the face of multimodality-based political campaigns in the era of information disruption.

RESEARCH METHOD

This study uses a qualitative descriptive methodology with an ethnographic model to examine text documents in cyber media and understand the meaning of political campaign texts in the 2024 presidential election. Data collection was carried out from May 2023 to February 2024. The study's primary data sources are divided into two categories: social media and online media. Social media includes: (a) Facebook, (b) Instagram, (c) Twitter, (d) WhatsApp, (e) YouTube/video. Online media includes: (a) detik.com, (b) kompas.com, (c) tribunnews.com, (d) merdeka6.com, (e) cnbcindonesia.com. Bezemer & Kress's multimodality analysis tool combines two verbal modes and Jewitt's visual mode. O'Halloran, Kress, and Jewitt (2003) detailed the methodology in data collection, analysis, and interpretation (Creswell, 2019).

The research sample consisted of 200 digital campaign texts, including images, video clips, uploaded captions, news headlines, and public comments. The sampling used purposive sampling, selecting data that meet specific criteria: relevance to the 2024 political campaign, containing multimodal elements (visual and verbal), and addressing important issues such as political framing, sarcasm, or misinformation. The research instrument was the researcher herself (human instrument), assisted by an artificial intelligence-based digital instrument, namely: computer vision to identify visual elements (colour, gestures, expressions, symbols, and camera angles) and Natural Language Processing (NLP) to analyse verbal text, sentiment, sarcasm, and language patterns. The research instruments are described in the following table 1.

Table 1. Visual and Verbal Mode Artificial Intelligence Research Instruments

Components	Indicators	Data Format
Visual Mode	Color, expression, gesture, symbol, camera angle	Images and Videos
Verbal Mode	Linguistic processes (material, mental, verbal), sarcasm, framing	Captions, News Text
Visual-Verbal Integration	Congruence, contradiction, reinforcement of meaning	Text-Image Combinations
Public Response	Sentiment, polarization	Public Comments
AI Detection	Expression, color, sentiment, misinformation	AI Output

Data collection methods included searching digital uploads using relevant keywords, downloading multimodal content, documenting, and initial coding based on multimodal categories. All data was stored in a digital repository to maintain transparency and repeatability of the analysis. Data analysis was conducted in three stages. First, multimodal analysis (Kress & van Leeuwen) to identify representational, interpersonal, and compositional meanings. Second, linguistic analysis based on Halliday's transitivity system to map verbal meaning structures. Third, integrating computer vision and NLP to detect visual, verbal, and political framing patterns more accurately.

The strengths of this study include a large sample size for a qualitative study (200 data points), a diverse set of sources (10 platforms), and the use of AI to enhance reliability and deepen interpretation. This provides the study with high validity and a substantial contribution to the study of digital multimodality.

RESULTS AND DISCUSSION

The Role of Multimodality in Unravelling the Meaning of 2024 Election Campaign Texts

The results demonstrate that the interpretation of meaning in 2024 election campaign texts is significantly influenced by distorted modes, facts, and data. The metafunctions of visual and verbal language play a crucial role in strengthening the meaning of the verbal mode. Interestingly, the visual mode can be insulting while the verbal mode remains complimentary, with both employing a sarcastic style (Kress & van Leeuwen, 2006). The analysis shows that multimodality, assisted by artificial intelligence (AI), clarifies the representational processes—namely action, reaction, verbal, and mental processes—while strengthening the detection of bias, sarcasm, and political framing. These findings reveal that digital campaigns are more driven by emotions than by policy substance (Halliday, 2014).

As presented in Table 2, the classification of multimodality modes in 2024 election campaign texts shows a clear dominance of the visual mode.

Table 2. Classification of Multimodality Modes in 2024 Election Campaign Texts

Multimodality Mode Type	Subcategories	Frequency of Occurrence	Percentage (%)
Visual Mode	Action Process	34 data	27%
	Reactional Process	26 data	21%
	Relational/Conceptual	18 data	14%
	Mental (Visual Expression of Emotion)	12 data	9%
Subtotal			71%
Verbal Mode	Verbal Process (Speech, Claims, Headlines)	21 data	17%
	Mental Process (Emotion, Perception, Hoaxes, Sarcasm)	15 data	12%
Subtotal			29%
Total		126 data	100%

The multimodal study of digital campaign texts, dominated by the visual mode (71%), reveals that images, poses, colours, graphic design, and visual framing of candidates heavily influence digital political discourse (Kress & van Leeuwen, 2006). This finding aligns with previous research by Machin

(2013) and Lim (2018), who found that visuals often play a greater role in determining voter responses than argumentative verbal content. The combination of verbal and visual modes is the primary source of meaning formation. For example, in a post featuring an image of Gibran with a silhouette of his father, the visual meaning cannot stand alone without reading the text "Gibran's profile is not just a joke". Similarly, the text cannot be fully understood without seeing the visuals that evoke specific political associations (Halliday, 2014). This indicates that political persuasion in digital campaigns relies more heavily on visual representation than purely linguistic explanation.

In the campaign text "Profil Gibran Ternyata Bukan Kaleng-kaleng," the image shows Gibran accompanied by the silhouette of his father, President Joko Widodo. This image represents a strong action process and reactional process. In the action process, Gibran functions as the Actor because he is positioned as the central political participant who receives public attention, while the Goal is public interpretation regarding his political legitimacy. The silhouette of Jokowi creates a vector that visually directs attention to dynastic politics and inherited political power.

The reactional process is visible through the symbolic gaze and facial expression. The audience becomes the Reactor, while the image of Gibran and Jokowi becomes the Phenomenon. The public reaction is triggered by the implied message that Gibran's political rise is not solely based on merit but on family privilege.

The verbal text "Gibran's profile is not just a joke" functions as a verbal process where the social media account acts as the Sayer and the sarcastic statement becomes the Verbiage. The continuation: "7 years as the son of a mayor, 2 years as the son of a governor, 9 years as the son of a president…" creates a strong mental process, inviting readers to act as Sensers and evaluate the legitimacy of Gibran's political career. The Phenomenon is the issue of nepotism and dynastic politics. The circumstance in this case is the broader socio-political context of the vice-presidential nomination process and Constitutional Court controversy. This contextual setting strengthens the interpretation of sarcasm and criticism embedded in both the visual and verbal modes. This finding supports Kress and van Leeuwen (2006), who argue that images are never neutral but always represent ideological meanings through representational processes.

The role of multimodality in conveying political campaign meanings stems from the synergistic interaction among visual modes, verbal modes, and public responses. In addition to inaccurate facts, much of the information is distorted, and some are even hoaxes. The AI classification model indicates that 34% of the analysed campaign texts fall into the misleading information category, while 11% are categorised as hoaxes. This finding is consistent with data from the Indonesian Ministry of Communication and Information (Kominfo, 2023), which reported an increase in political hoaxes ahead of the election, particularly through images, short videos, and memes. Such patterns bias public interpretation, as people often fail to verify whether the information is factual or a manipulated narrative (Hangloo & Arora, 2022).

The Role of Artificial Intelligence (AI)-Assisted Multimodality in Digital Campaign Texts

Artificial intelligence technology plays a significant role in uncovering multimodal patterns in digital campaign texts. Through artificial computer vision, AI detects facial expressions, gestures, colours, and political symbols that form visual frames. Meanwhile, NLP analysis identifies sarcasm, misinformation, and sentimental tendencies in verbal texts. AI's visual-verbal integration reveals inconsistencies between images and text in campaign messages that contribute to the construction of specific frames. AI also systematically maps public responses and narrative patterns. Thus, AI improves the accuracy, speed, and depth of multimodal analysis in the context of digital campaigns.

Table 3. The Role of Artificial Technology in Multimodal Analysis of Digital Campaign Texts

Multimodality Components	The Role of Artificial Intelligence (AI)	Artificial Intelligence (AI) Output
Visual Mode (images, expressions, gestures, colors, symbols)	Computer Vision automatically detects facial expressions, body gestures, gaze direction, framing, colour composition, and even political symbols.	- Gibran's confident expression, Anwar Usman's sadness, and Ganjar's aggressiveness were identified with high accuracy. - The red-blue-green colour scheme was identified as a symbolic political strategy.

Multimodality Components	The Role of Artificial Intelligence (AI)	Artificial Intelligence (AI) Output
Verbal Mode (captions, comments, titles, campaign narratives)	Natural Language Processing (NLP) analyses language patterns, sarcasm, sentiment, misinformation, and framing tendencies in text.	- 34% of texts contain misinformation; 47% of captions contain sarcasm. - Public sentiment can be mapped based on comments.
Visual-Verbal Relations	Multimodal AI measures semantic matching between text and images.	The AI found that 62% of text-image comparisons were not neutral, but instead served to reinforce political framing.
Public Perception Mapping	Sentiment analysis automatically classifies public responses from thousands of comments.	- AI shows that specific images evoke significant spikes in negative emotions.
Hoax and Distortion Detection	The AI fact-checking module flags manipulative narrative patterns.	- 11% of campaign posts contained visual-verbal hoaxes detected by AI.

Artificial intelligence (AI) technology plays a significant role in uncovering multimodal patterns in digital campaign texts. In the visual domain, computer vision can detect facial expressions, body gestures, gaze direction, colour composition, and even political symbols with high accuracy. Through this analysis, AI identified Gibran's confident expression, Anwar Usman's facial sadness, and Ganjar's gestural aggressiveness as consistent patterns used in candidate representations. Furthermore, AI mapped the strategic use of red, blue, and green colour schemes as symbols of political identity.

In the verbal mode, Natural Language Processing (NLP) revealed rhetorical tendencies in captions, comments, and news headlines. The analysis showed that 34% of campaign texts contained misinformation, while 47% of captions utilised sarcasm or insinuation to shape public opinion. AI can also automatically map public sentiment, enabling more systematic identification of public responses to specific issues. In the realm of visual-verbal relations, multimodal AI assesses the congruence between the meaning of images and text. The findings show that 62% of text-image pairs are not neutral but instead reinforce a specific political framing. This reveals that much campaign content exploits the imbalance of visual and verbal meaning as a persuasive strategy.

Sentiment analysis technology identified spikes in negative emotions in the public when viewing specific images, demonstrating the direct influence of visual materials on public perception. The AI hoax detection module also found that 11% of campaign posts contained visual-verbal manipulations that could mislead the public. Overall, AI increases the objectivity, speed, and depth of multimodality analysis by providing empirical evidence on how visuals, language, and public responses interact to shape digital political discourse.

In addition, Artificial Intelligence significantly improves the analysis of multimodal political discourse by detecting representational structures more systematically. Through Computer Vision, AI identifies action processes by recognizing gestures, body positioning, vectors, and participant interactions. For example, Gibran's confident debate posture is classified as an action process where he functions as the Actor projecting authority and competence.

AI also detects reactional processes through facial expressions and gaze analysis. Anwar Usman's sadness and Ganjar's aggressive expression are interpreted as reactional meanings where emotional reactions shape public interpretation. Eye contact and facial intensity become crucial indicators of political positioning. In the verbal process, Natural Language Processing (NLP) identifies captions, headlines, and statements as political speech acts. Headlines such as "Anies remains anti-IKN?" function as verbal processes where media acts as the Sayer and the political accusation becomes the Verbiage. AI found that 47% of captions contain sarcasm and indirect criticism. The mental process is revealed through sentiment analysis. Public comments expressing fear, anger, disappointment, and support indicate mental processes where voters become Sensers reacting to political narratives. AI successfully mapped emotional polarization through comment sections.

Furthermore, circumstances such as debate venues, online platforms, timestamps, and campaign settings are automatically classified by AI to strengthen contextual interpretation. These circumstances support the representational process and explain why certain political messages gain stronger emotional

responses. Therefore, AI does not replace multimodal analysis but strengthens its objectivity, speed, and reliability in uncovering hidden political meanings.

Representational Verbal and Visual Language of 2024 Election Campaign Texts

Based on the political campaign text data collected in the comment section, there are also various interpretations based on facts, data, and information captured on social media. Interpretation based on a combination of visual mode and verbal mode complements and supports each other. So, the interpretation of meaning cannot be done only by highlighting one of these modes. The roles of verbal mode and visual mode have their respective roles when representing meaning, verbal mode represents the narrative world, while visual mode represents the displayed world.

The meaning of the verbal mode of data is dependent on the visual mode to be said to be potentially called an act of insult even though the data is true, opening with the title of the illustration image: ‘Profil Gibran Ternyata bukan kaleng2'. using sarcasm language style (harsh satire). Verbal mode in the text of the argumentation sentence reinforcing the use of sarcasm in the title: 'Gibran's profile is not just a joke, 'Amazing, right? Let the people judge, followed by a harsh, sarcastic argument: 'Amazing, right? Let the people judge, 7 years as the son of a mayor, 2 years as the son of a governor, 9 years as the son of a president, 4 years as the nephew of the chief justice of the Constitutional Court, 1 month as the brother of the PSI chairman, 5 days as a Golkar cadre, 2 years as a mayor, So do not underestimate it!'

The visual meaning of the sentence has a relational process: relational linking (relational process) of the illustration of a person with the image of his father as the carrier, followed by the attributes of the positions his father has held 'mayor, governor, and president'... similarly followed by the text 'extraordinary am I? Let the people judge as a material process, followed by an affirming sentence Gibran's profile as a carrier turns out not to be a pipsqueak Followed by argumentation data that is enough to strengthen the criticism that uses sarcasm language style when describing KKN in the vice president as paradoxical as independent because it is not bound to the verbal mode.

Multimodality relations of verbal modes and visual modes, speakers mock Gibran It's not a pipsqueak Followed by argumentation data that is enough to strengthen the criticism that criticises by using a sarcasm language style that describes KKN on the vice president 02 as a paradox between Gibran It's not a pipsqueak with the greatness of KKN over who raised him. Closed with So don't underestimate! Closed with a sentence conclusion let the people judge .The role of multimodality Meta visual and verbal language functions in the 2024 Election campaign text, used data on 13 December 2023 the arah_bangsa account posted an information on Instagram with the caption "Debat Anies dengan Ganjar " so that it attracted the attention of social media citizens, there were 118 comments on the post. The post data shows that many social media users support IKN so that the IKN (National Capital City) government programme continues. This is because Ganjar has a programme that wants to continue the relocation of IKN, as the theme of the presidential candidate debate, while Anies has the potential not to continue IKN because he needs to criticise the IKN Law. According to Anies, the interpretation of the IKN Law does not involve the public, Anies also believes that there are still many other needs that are needed. To be clearer, the role of multimodality can be seen in the following description of visual and verbal modes.

Data 1 The Role of Multimodality in the Visual Meaning of 'Anies Debate with Ganjar?' 13 December 2023

Data Code	Visual Modes	Visual Meaning
D22./CP01-03/PMMo/MVe		Picture of the debate between Ganjar 03 and Anies Baswedan 01 On the topic of IKN *Jawab Ganjar,* *Anies Tetap Anti IKN?* *TOPIK DEBAT*

The data above have an image of relational meaning between dependent visual modes or independent visual modes as images, namely: The image of the debate between Ganjar 03 and Anies Baswedan 01, with the topic of the IKN debate, based on the visual mode of the image of Ganjar and Anies Tetap as the carrier. Those who argue about the relocation of IKN are called behavier. Which is followed by the meaning of the visual mode in the form of sentence text: 'Jawab Ganjar, Anies Tetap Anti IKN?, Ganjar Pranowo asked questions related to IKN to Anies Baswedan, Ganjar asked whether Anies refused to continue IKN?, Anies opened his answer by starting to criticise the IKN Law, According to Anies, the interpretation of the IKN Law did not involve the public, Anies also argued that there were many other needs that were needed, Ganjar answered by concluding that Anies was in opposition and refused to continue IKN.

Relational process sentence: relational linking (relational process) Ganjar and Anies capres as carrier with coalition 01 and 03 as attributes. IKN as a circumstantial place where IKN is moved, a vector that positions Anies' participants as actors and Ganjar as reactors. IKN as an object. With the goal of moving IKN. Material process so that the presidential candidate related to IKN to (material) IKN. Gaze shows Ganjar answering by concluding that Anies is in opposition and rejects IKN being continued. Picture of Ganjar 03's debate with anies 01 On the topic of IKN, Jawab Ganjar, Anies Tetap Anti IKN?, On the topic of IKN, Ganjar concludes that Anies is in opposition and rejects IKN being continued.

The verbal mode of interpretation data is dependent on the visual mode to be said to be potentially called an act of insult even though the data is true, opening with the title of the illustration image: 'Profil Gibran Ternyata bukan kaleng2'. The use of sarcasm language style (harsh satire) is in the argumentation of the title: Picture illustration of Gibran Rakabuming Raka with a shadow (silhouette) similar to his father. By Badri_chanel account #Repost @minang followed by mo verbal in the form of sentence text.

The analysis of representational meaning can be clearly seen in Data 1: "Debat Anies dengan Ganjar? Jawab Ganjar, Anies Tetap Anti IKN?" The image shows Ganjar Pranowo (candidate 03) debating with Anies Baswedan (candidate 01) on the topic of IKN. This represents an action process because Ganjar acts as the Actor who initiates the debate and directs criticism, while Anies becomes the Goal receiving the political challenge. The vector is formed through body posture, pointing gesture, and gaze direction.

The text "Jawab Ganjar, Anies Tetap Anti IKN?" also creates a reactional process. Ganjar becomes the Reactor who responds to Anies' criticism of the IKN Law, while Anies' political stance becomes the Phenomenon. The image positions both candidates as carriers of ideological opposition. The caption and debate statement function as a verbal process, where Ganjar acts as the Sayer and the issue of IKN continuation becomes the Verbiage. The text strengthens political framing by suggesting that Anies opposes national development.

The public interpretation of this issue forms a mental process. Social media users become Sensers who evaluate whether rejecting IKN is seen as political responsibility or political opposition. The Phenomenon is the continuation of the IKN project itself. The circumstance is highly significant here. The place is the presidential debate arena, while the time is December 13, 2023. This setting gives institutional legitimacy and increases the political weight of the statement. Debate context makes the verbal and visual processes more persuasive.

This demonstrates that meaning is not produced by text alone but by the simultaneous interaction of action, reaction, verbal judgment, mental evaluation, and circumstantial context.

Representational Visual Language in Political Party Campaign Texts

Representational structures categorised as narrative are characterised by the presence of vectors, representing ongoing actions and events. On the other hand, conceptual structures are characterised by the absence of vectors and represent participants described by class, structure, or meaning. In other words, conceptual structures are stable. For further clarification, we will explain this through a percentage analysis of visual modes based on AI (Vision Detector). AI facial expression patterns, poses, colours, framing, and vectors

Table 4. Classification of Visual Modes in Digital Campaigns

Visual Category	Artificial Intelligence Findings	Persentase (%)
Emotional Expression (Mental Visual)	Confidence (Gibran), Sadness (Anwar Usman)	28%
Use of Color Framing	Red (Prabowo), Blue (Anies), Green (Ganjar)	25%
Camera Angle	Low angle (Prabowo), Eye-level (Anies/Ganjar)	19%
Vector: Gesture & Direction	Ganjar pointing, Anies defensive	17%
Symbolic Attributes	Silhouette of Gibran's father, party logo	11%

Prabowo's billboards, featuring the tag "People's Officer," appear in the Pati area. This represents a narrative action process, as it includes: Actor → Prabowo (in the billboard); Goal → PDI-P voters in Pati; Vector → Prabowo's body gesture facing forward. AI Vision found that: Image angle → slightly low angle (building an image of authority); Dominant red colour → power, urgency; Bold typography → persuasive intentionality. AI Sentiment Mapping revealed: 67% of commentators considered the billboard a political strategy to "disguise" identity (similar to Jokowi). 22% were neutral. 11% supported. This multimodal meaning cannot be understood from either the text or the image alone; the two complement each other.

In Prabowo's billboard "Petugas Rakyat" appearing in Pati, Central Java, the billboard itself represents an action process. Prabowo functions as the Actor through his forward-facing body posture, while voters—especially PDI-P supporters—become the Goal. The vector is created through body gesture and visual dominance.

The slightly low camera angle identified by AI constructs authority and power. This also supports a reactional process, where viewers become Reactors responding to the visual representation of leadership and familiarity with Jokowi-like imagery. The slogan "Petugas Rakyat" functions as a verbal process, where the campaign team acts as the Sayer and the slogan becomes the Verbiage. The message attempts to position Prabowo as a people-centered leader.

The public response analyzed by AI shows a mental process: 67% of comments interpreted the billboard as identity camouflage rather than genuine political positioning. Viewers become Sensers, and the Phenomenon is Prabowo's symbolic political strategy. The circumstance includes place (Pati, Central Java), political timing (pre-election campaign period), and coalition support from Gerindra, Golkar, PKB, and PAN. These contextual factors strengthen the persuasive intention of the billboard. Therefore, the meaning of political campaign visuals emerges from the integration of representational processes and contextual circumstances rather than isolated images alone.

Based on the relationship between verbal and visual modes in the cyber text data of political parties' campaigns in the 2024 elections. Multimodality as a combination of modes in conveying signs, multimodality involves the use of more than one mode in discourse (Kress & Van Leeuwen, 2001). In line with Hermawan's research (2013), multimodality analysis in texts that use verbal and visual modes is understood as an analysis procedure that combines linguistic analysis, namely Systemic Functional Linguistics (SFL) with visual analysis tools to understand images. Multimodality analysis in this study focuses on the verbal mode and the visual mode. Verbal mode analysis uses Halliday's (2014) transitivity and visual mode analysis refers to visual grammar (Kress & Van Leeuwen, 2006).

Based on the results of the interpretation of meaning done by simply highlighting the mode, it was found: a) Political campaign text data collected on social media found facts and data that were twisted; b) The role of visual and verbal language meta-functions in the text of the 2024 election campaign reinforces the meaning of the dependent data verbal mode; c) The role of the visual mode is potentially an insulting act even though the verbal mode is complimentary; d) The meta-functional role of language in the visual mode and verbal mode in campaign texts using sarcasm language style was found; e) The role of multimodality in the representational visual language of the 2024 election political campaign text with verbal process and mental process is manifested through linguistic dialogue in the election campaign text described in the picture.

This study proves that digital campaign texts in the 2024 Election cannot be understood through only one mode. The visual mode (71%) is far more dominant than the verbal mode (29%) in shaping meaning, and their combination often creates sarcastic, biased, and even misleading messages. With the assistance of artificial intelligence, this study successfully reveals that 62% of text-image pairs are non-

neutral, 34% contain misinformation, and 11% are visual-verbal hoaxes that are difficult to detect manually. These findings reinforce the importance of AI-assisted multimodal approaches for understanding increasingly complex digital political communication practices.

This research offers significant innovations in the study of multimodality and digital political communication by integrating artificial intelligence (AI) technology as the primary analytical tool. Unlike previous research that relied solely on manual analysis of visual and verbal relationships, this study uses computer vision and Natural Language Processing to precisely and at scale detect facial expressions, gestures, symbolic colours, sarcasm, misinformation, and framing patterns in campaign texts. The findings that 62% of text-image relationships are non-neutral and 34% of content contains misinformation represent a new empirical contribution not yet reported in Indonesian political campaign research. This research is also among the first to uncover the existence of visual-verbal hoaxes supported by AI systematically. In addition to providing a new analytical perspective, this research develops an integrative model linking multimodal analysis, functional systemic linguistics, and AI, enriching digital political research methodology and opening new avenues for exploration in communication studies in the era of artificial intelligence.

This research not only offers theoretical contributions to the study of digital semiotics and Systemic Functional Linguistics but also provides practical implications for regulators, political practitioners, and voters. The use of AI in multimodal analysis has proven to increase the accuracy, speed, and depth of interpretation of manipulative campaign messages. Therefore, it is recommended that the General Election Commission (KPU) adopt a real-time AI-based monitoring system and push for regulations requiring the labeling of AI-modified campaign content. Ultimately, improving multimodal literacy among voters is the key to creating a healthier, more transparent democracy that is resilient to disinformation in the age of artificial intelligence.

CONCLUSION

The results of an artificial intelligence-assisted multimodality analysis of digital campaign texts for the Presidential Election indicate that political messages circulating on social media are constructed through a combination of verbal language and complementary visual elements. Verbal modes in campaign texts often contain statements, calls, criticisms, and opinions characterised by verbal, mental, and material processes. Verbal processes are evident in the form of dialogue, political statements, and public appeals. Mental processes are evident in netizens' assessments and reactions, which express specific perceptions or emotions about the candidates. Meanwhile, material processes refer to concrete actions, such as debates, declarations of support, or other campaign activities, that are part of the image-building strategy. Visually, political campaign messages rely heavily on pictorial representations, expressions, gestures, colours, and compositions that convey specific meanings. Visuals such as debate photos, billboards, candidate poses, party symbols, and facial expressions provide additional context that is often more powerful than text. Actions such as the installation of billboards, candidate poses while speaking, or facial expressions during debates create a process of action and reaction that influences public interpretation. When visuals and verbal messages are combined, they create more complex and persuasive meanings, especially in political contexts that are often rife with framing and sarcasm. Artificial intelligence (AI) plays a significant role in identifying patterns of verbal and visual relationships more quickly, accurately, and deeply. AI is capable of mapping language patterns, visual gestures, and meaningful relationships that are difficult to capture manually. With the support of AI, this research successfully uncovers how digital campaign texts use multimodality to shape opinions, influence perceptions, and strengthen candidates' images in the public sphere.

ACKNOWLEDGMENTS

I would like to thank USU for funding this research. Thank you to the corresponding author and all the team who has helped edit this article so that it is suitable for publication.

AUTHOR CONTRIBUTIONS

Research article with some conceptualizations of Bezemer & Kress's multimodality, which combines two verbal modes and Jewitt's visual mode. O'Halloran, Kress and Jewitt.; Detailed methodology in data collection, analysis and interpretation of Creswell.

CONFLICTS OF INTEREST

The author(s) declare no conflict of interest.

USE OF ARTIFICIAL INTELLIGENCE (AI)-ASSISTED TECHNOLOGY

The authors declare that no artificial intelligence (AI) tools were used in the generation, analysis, or writing of this manuscript. All aspects of the research, including data collection, interpretation, and manuscript preparation, were carried out entirely by the authors without the assistance of AI-based technologies.

REFERENCES

Adami, E., & Kress, G. (2014). The social semiotics of contemporary communication. *Visual Communication, 13*(3), 1-14.

Austin, J. L. (1962). *How to do things with words*. Oxford University Press.

Bateman, J. A. (2014). *Text and image: A critical introduction to the visual/verbal divide*. Routledge.

Bezemer, J., & Kress, G. (2016). *Multimodality, learning and communication: A social semiotic frame*. Routledge.

Bocheara, A. (2015). The role of religion in shaping politeness in Moroccan Arabic. *Journal of Politeness Research, 11*(1), 71-98.

Bolognesi, M., & Lievers, F. S. (2020). How language and image construct synaesthetic metaphors in print advertising. *Visual Communication, 19*(4), 431-457.

Boulianne, S. (2019). Social media use and political participation: A meta-analysis. *New Media & Society, 21*(8), 1871-1891.

Bucur, C. (2020). Detecting multimodal fake news: A survey. *ACM Computing Surveys, 53*(5), 1-38.

Budiardjo, M. (2015). *Basics of political science* (Rev. ed.). Gramedia Pustaka Utama.

Chaer, A. (2007). *Language studies: Internal structure, usage, and learning*. Rineka Cipta.

Chen, Y., & Zhang, Y. (2021). Sarcasm detection in social media: A survey. *IEEE Transactions on Affective Computing, 12*(4), 876-893.

Chen, Z. T., & Cheung, M. (2022). Consumption as extended carnival on Tmall in contemporary China. *Social Semiotics, 32*(2), 163-183.

Dos Santos, T. A., & Lopes, I. F. (2025). Is Social Media a Democratic Forum for Public Accountability in Times of Crisis? The Brazilian Government's Response to the COVID-19 Pandemic. *Abacus*, *61*(1), 194-217. https://doi.org/10.1111/abac.12356.

Eggins, S. (2004). *An introduction to systemic functional linguistics* (2nd ed.). Continuum.

Enli, G. (2017). Twitter as arena for the authentic outsider. *European Journal of Communication, 32*(1), 50-61.

Entman, R. M. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication, 43*(4), 51-58.

Fazio, X., Gallagher, T. L., & DeKlerk, C. (2022). Exploring adolescents' critical reading of socioscientific topics using multimodality texts. *International Journal of Science and Mathematics Education, 20*, 93-116.

Finch, G. (2000). *Key concepts in language and linguistics*. Palgrave Macmillan.

Gazdar, G. (1979). *Pragmatics: Implicature, presupposition and logical form*. Academic Press.

Giachanou, A., & Crestani, F. (2022). Multimodal fake news detection: A survey. *Information Processing & Management, 59*(4), 102-118.

Graakjær, N. J. (2019). Sounding out "I'm lovin' it"—A multimodality discourse analysis. *Critical Discourse Studies, 16*(5), 569-582.

Grabe, M. E., & Bucy, E. P. (2019). *Image bite politics: News and the visual framing of elections*. Oxford University Press.

Hangloo, S., & Arora, B. (2022). Combating multimodality fake news on social media. *Multimedia Systems, 28*(6).

Habal, M. B. (2023). Creative Global Communications: A Game Changer in the New Digital Information Technology Arena—Still Competition Between Structure and Function. *Journal of Craniofacial Surgery*, *34*(4), 1157-1158. https://doi.org/10.1097/SCS.0000000000009447.

Hou, S., Zhou, S., Liu, W., & Zheng, Y. (2018). Classifying advertising videos by topicalising high-level semantic concepts. *Multimedia Tools and Applications, 77*(19), 25475-25511.

Jewitt, C. (2016). *Introducing multimodality*. Routledge.
Jewitt, C., Bezemer, J., & O'Halloran, K. L. (2016). *Introducing multimodality*. Routledge.
Jin, Z., Cao, J., Guo, H., Zhang, Y., & Luo, J. (2017). Multimodal fusion with recurrent neural networks for rumor detection on microblogs. *Proceedings of the 25th ACM International Conference on Multimedia*, 795-803.
Johnstone, B., & Marcellino, W. M. (2010). Dell Hymes and the ethnography of communication. In *The SAGE handbook of sociolinguistics*.
Karim, M. A., & Erwhintiana, I. (2020). Perlokusi speech patterns in the web series *Behind the Heart*. *Pujangga Journal, 6*(2).
Kelly, L. B., & Kachorsky, D. (2022). Text complexity and picturebooks: Learning from multimodality analysis. *Reading and Writing Quarterly, 38*(1), 33-50.
Khattar, D., Goud, J. S., Gupta, M., & Varma, V. (2019). MVAE: Multimodal variational autoencoder for fake news detection. *The World Wide Web Conference*, 2915-2921.
Kramsch, C. (2002). Language and culture: A social semiotic perspective. *ADFL Bulletin, 33*(2), 8-15.
Kreiss, D., Lawrence, R. G., & McGregor, S. C. (2018). In their own words: Political practitioner accounts of digital media use. *Political Communication, 35*(2), 179-198.
Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.
Kress, G., & van Leeuwen, T. (1996/2006). *Reading images: The grammar of visual design*. Routledge.
Kress, G., & van Leeuwen, T. (2001). *The modes and media of contemporary communications*. Arnold.
Kusmanto, H., & Widodo, P. (2022). Positive politeness strategies during online learning. *Studies in English Language and Education, 9*(3), 1170-1182.
Kusno, A., Arifin, M. B., & Mulawarman, W. G. (2022). Identifying the virtual extralingual context of social media language. *Diglosia, 5*(1S), 261-282.
LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*(7553), 436-444.
Lim, M. (2020). The algorithmic management of polarization and democracy in Southeast Asia. In *Digital Disinformation in Southeast Asia*. ISEAS Publishing.
Macagno, F., & Pinto, R. B. W. S. (2021). Reconstructing multimodality of arguments in advertisements. *Argumentation, 35*(1), 141-176.
Machin, D., & Mayr, A. (2012). *How to do critical discourse analysis: A multimodal introduction*. Sage Publications.
Mahsun, M. (2005). *Language research methods*. Raja Grafindo Persada.
Martin, J. R., & Rose, D. (2007). *Working with discourse: Meaning beyond the clause* (2nd ed.). Continuum.
Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th ed.). Routledge.
Messaris, P., & Abraham, L. (2001). The role of images in framing news stories. In S. D. Reese et al. (Eds.), *Framing public life* (pp. 215-226). Routledge.
Miles, M. B., & Huberman, A. M. (2014). *Qualitative data analysis: A methods sourcebook* (3rd ed.). Sage Publications.
O'Halloran, K. L., Tan, S., & Marissa, K. L. E. (2017). Multimodal analysis for critical thinking. *Learning, Media and Technology, 42*(2), 147-160.
Robbins, S. P. (2008). *Organisational behaviour* (B. Molan, Trans.). Intan Sejati.
Rochmawati. (2020). Pragmatic and rhetorical strategies in English-written jokes. *Pujangga Journal, 6*(2).
Rossini, P. (2022). Beyond incivility: Understanding patterns of uncivil and intolerant discourse. *Communication Research, 49*(3), 339-363.
Rustono. (1999). *Principles of pragmatics*. IKIP Semarang Press.
Salsabila, N., et al. (2021). Perlocutionary speech acts in the dialogue of *Imperfect* film. *Allegory Journal, 1*(2).
Schiffrin, D. (2007). *An introduction to discourse studies*. Student Library.
Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter, 19*(1), 22-36.
Van Leeuwen, T. (2015). Multimodality. In D. Tannen et al. (Eds.), *The handbook of discourse analysis* (2nd ed., pp. 447-465). Wiley Blackwell.
Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146-1151.

Wang, Y., Ma, F., Jin, Z., Yuan, Y., Xun, G., Jha, K., & Gao, J. (2018). EANN: Event adversarial neural networks for multimodal fake news detection. *Proceedings of the 24th ACM SIGKDD International Conference*, 849-857.