Indonesian Journal of Computational Language Studies
Vol. 1, No. 3, September 2026, pp. 192~205
ISSN: 3163-7884, DOI: https://doi.org/10.67490/ijcl.v1i3.1029

From slang to signal: tracking youth language change in Indonesian platforms through diachronic word embeddings

Almaidah
Universitas Negeri Malang, Indonesia

Article Info

Article history:

Received Jul 6, 2026
Revised Aug 24, 2026
Accepted Sep 25, 2026

Keywords:

code-mixing;
diachronic embeddings;
digital platforms;
Indonesian youth slang;
semantic change.

ABSTRACT

Background: Indonesian youth slang increasingly circulates through digital platforms where lexical innovation, identity performance, meme culture, and code-mixed interaction reshape everyday meaning. **Objective:** This study examines how youth slang in Indonesian digital platforms can be tracked as semantic movement through diachronic word embeddings and contextual sociolinguistic validation. **Method:** Using public, dateable, and verifiable corpus materials, this study maps slang-bearing units across platform contexts, classifies semantic movement patterns, and validates embedding-oriented interpretation through concordance-based coding. **Results:** Findings show that status and identity specialization forms a dominant semantic pattern, followed by evaluative resemanticization, meme-context volatility, and pragmatic stance stabilization. Results also indicate that short-video and meme circulation privilege volatile intertextual meanings, whereas self-presentation contexts foreground slang related to attractiveness, hierarchy, persona, and social value. Contextual validation demonstrates that many slang meanings remain interactionally contingent, requiring surrounding discourse before computational shifts can be interpreted as sociolinguistically meaningful. **Implication:** This study shows that computational modelling of slang requires platform-sensitive interpretation to avoid reducing youth language change to frequency, novelty, or vector movement alone. **Novelty:** This study contributes a diachronic, corpus-assisted, and sociolinguistically grounded framework for analysing Indonesian youth slang as a measurable signal of digital semantic change.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Almaidah
Universitas Negeri Malang, Indonesia
Jl. Semarang No. 5, Malang, Jawa Timur, 65145, Indonesia
Email: almaidah.2502119@students.um.ac.id

1. INTRODUCTION

Indonesia offers an unusually dense empirical site for examining youth language change because digital interaction is no longer a marginal communicative space but a major arena of everyday linguistic production. At the end of 2025, Indonesia had around 230 million internet users, equivalent to 80.5% internet penetration, and approximately 180 million social media user identities, representing 62.9% of its population. Earlier APJII-based reporting also indicated that internet penetration among Gen Z users aged 12–27 reached 87%, while 48% of children under 12 had internet access, some through platforms such as Facebook, Instagram, and TikTok. These figures matter linguistically because youth slang now circulates across highly networked environments where spelling, affect, humor, identity, and algorithmic visibility interact. KBBI VI Daring also records lexical updating twice annually, with its latest update in April 2026, signalling that institutional lexicography increasingly confronts fast-moving digital vocabulary.

Existing scholarship has established that Indonesian slang is socially meaningful rather than merely informal. Budiasa [1] shows that social media slang reflects interactional creativity, while Mayrita et al. [2] connect platform slang with everyday communicative practice. Wagiati et al. [3], Paoleti and Mujahidah [4], and S et al. [5] further situate youth slang within digital identity, generational affiliation, and adolescent social positioning. Other studies have examined more specific mechanisms of innovation, including fandom-driven word formation [6], code-switching in gaming discourse [7], and broader social-media-driven language transformation [8], [9]. However, much of this work remains synchronic, interpretive, or focused on categorising forms rather than tracing semantic movement over time. Conversely, computational studies using embeddings and NLP for Indonesian have advanced modelling capacity [10], [11], yet youth slang has rarely been treated as diachronic evidence of semantic change.

This study addresses this gap by asking how Indonesian youth slang shifts across time, platforms, and semantic neighbourhoods when examined through diachronic word embeddings and validated through contextual reading. More specifically, this study investigates three linked questions: how selected slang terms move in embedding space across defined time slices; what kinds of semantic change can be inferred from neighbour turnover, cosine displacement, and contextual concordance; and how platform conventions, code-mixing, fandom speech, gaming interaction, and digital identity practices shape these movements. This design does not assume that platforms cause linguistic change. Rather, it treats platforms as structured communicative environments where recurring lexical associations become observable in distributional data. By combining computational semantic tracking with sociolinguistic interpretation, this study aims to move beyond isolated examples of slang toward a more systematic account of how youth lexical innovation becomes measurable as shifting relational meaning.

This study advances a provisional argument: Indonesian youth slang should be analysed not simply as unstable vernacular play, but as a sociolinguistic signal whose changing semantic relations can reveal how digital publics reorganise meaning, stance, and identity over time. Diachronic embeddings are useful here because they capture changes in lexical neighbourhoods that may not be visible through frequency counts alone, especially when slang terms retain similar surface forms while acquiring new affective, social, or platform-specific associations. This argument builds on recent work on multilingual semantic change and interpretable diachronic tracking [12], [13], while also recognising Indonesian NLP concerns about code-mixing, low-resource representation, cultural grounding, and sociolinguistic bias [14], [15]. The broader implication is methodological: computational modelling can enrich youth-language research only when embedding movement is interpreted through corpus context, linguistic theory, and careful validation rather than treated as self-explanatory evidence.

2. LITERATURE REVIEW

2.1. Youth slang in Indonesian digital platforms

Youth slang in Indonesian digital platforms is best understood as socially organised lexical innovation rather than informal deviation from standard language. In sociolinguistic terms, slang indexes age, intimacy, stance, group membership, humour, and symbolic distance from institutional norms. Budiasa [1], Mayrita et al. [2], and Wagiati et al. [3] show that Indonesian social media slang is embedded in everyday communicative routines, not merely in playful excess. Paoleti and Mujahidah [4] and S et al. [5] further link Gen Z and Gen Alpha slang to identity formation, while E. [16] foregrounds adolescent social identity. However, existing definitions differ between form-based, identity-based, and platform-based accounts.

Indonesian youth slang can be analysed through several interrelated indicators: word formation, semantic play, code-mixing, affective stance, group indexing, and platform circulation. Permatasari and Karjo [6] show that fandom language contributes to Indonesian internet slang through borrowing, abbreviation, and hybrid word formation. Dzulqarnain [7] adds that online gaming chats generate slang through code-switching, interactional speed, and peer-group alignment. Naqvi [8], Pirdaus and V. [9], and Gul and Rukh [17] demonstrate that digital communication accelerates intergenerational gaps and lexical transformation. These indicators suggest that slang is not a fixed category, but a relational practice shaped by audience, medium, temporality, and community norms.

2.2. Diachronic semantic change

Diachronic semantic change refers to shifts in word meaning across time, but computational linguistics has expanded this concept from dictionary-based replacement to measurable movement in distributional space. Armaselu et al. [12] show that multilingual word embeddings can trace semantic change when lexical relations are linked to structured linguistic knowledge. Fathoni et al. [10] apply word embedding techniques to Indonesian loanword technology, demonstrating how distributional methods can reveal meaning shifts in local linguistic contexts. Kolli et al. [13] propose interpretable word-centred semantic graphs for diachronic sense tracking. Yet semantic movement should not be equated automatically with social change without contextual validation.

Indicators of diachronic semantic movement usually include neighbour turnover, cosine displacement, cluster reconfiguration, frequency stability, and contextual sense differentiation. Armaselu et al. [12] emphasise cross-linguistic comparability, while Kolli et al. [13] prioritise interpretability through semantic graphs rather than opaque vector movement. Czajka et al. [18], although working with sign language embeddings, reinforces that embedding representation is sensitive to modality and representational structure. In Indonesian NLP, Santoso [19], Wahid [20], and Yuliyana [21] show that domain-specific embeddings, corpus patterns, and hybrid modelling can improve linguistic modelling. These studies support a cautious view: embedding shifts become meaningful only when model evidence is interpreted alongside corpus contexts.

2.3. Indonesian digital language modelling

Indonesian digital language modelling raises a further issue: representation. Recent NLP work has improved Indonesian-language processing through BiLSTM-CRF named entity recognition, transformer-based sentiment analysis, code-mixing detection, readability assessment, and morphological segmentation [19], [22], [11], [23], [21]. However, computational adequacy does not guarantee sociolinguistic adequacy. Bouazzaoui [14] argues for culturally grounded Indonesian language understanding, while Wulandari [24], Hadziqi [25], Nadiyah [26], Wibisono [27], and Sumadi [15] expose multilingual inconsistency, cultural safety failures, hallucination, adaptation limits, and representational bias in Indonesian AI systems. Youth slang therefore requires both computational modelling and sociocultural scrutiny.

Useful indicators for modelling Indonesian youth slang must therefore combine linguistic, computational, and sociological dimensions. Lexical indicators include spelling variation, affixation, abbreviation, borrowing, semantic broadening, narrowing, and metaphorical extension. Interactional indicators include humour, insult, intimacy, fandom affiliation, gaming coordination, political stance, and affective intensification, as suggested by Budiasa [1], Dzulqarnain [7], and Elsafoury et al. [28]. Computational indicators include embedding displacement, nearest-neighbour instability, domain specificity, and cross-platform divergence. This study is positioned at this intersection: it treats slang as a traceable distributional pattern, but interprets such pattern through youth identity, platform ecology, and Indonesian sociolinguistic diversity.

3. METHOD

This study uses Indonesian youth slang tokens and their surrounding digital contexts as its unit of analysis. Material objects consist of public, verifiable textual items collected from official language sources, public media, open scholarly publications, and platform-related reports concerning Indonesian slang, youth language, digital identity, code-mixing, and semantic change. Each document is treated as a corpus item when it contains searchable metadata, identifiable source origin, publication date, and relevant lexical or contextual evidence. Text units include excerpts, titles, descriptions, keyword contexts, and coding notes. Visual or video fields are retained only when a source contains multimodal material requiring image description, transcript, timestamp, or screenshot reference.

This study adopts a corpus-assisted diachronic design that combines computational semantic modelling with sociolinguistic interpretation. This design is appropriate because Indonesian youth slang changes through repeated public use, spelling variation, borrowing, code-mixing, humour, fandom circulation, and platform-specific interaction. Rather than treating slang as isolated vocabulary, this study examines lexical items as relational signs whose meanings emerge from neighbouring words and recurring contexts. Diachronic word embeddings provide a way to observe semantic movement, while qualitative concordance reading prevents vector change from being interpreted mechanically. This design follows recent work on semantic embeddings, Indonesian NLP, code-mixing, and culturally grounded language modelling.

Table 1. Composition of dataset or sample

Code	Data Type	Source	Period	Quantity	Characteristics
META-01	Corpus metadata	KBBI VI Daring	2021–2026	1 set	Official lexical infrastructure and update context
META-02	Corpus metadata	KBBI official application/channel	2021–2026	1 set	Institutional language reference and access metadata
META-03	Corpus metadata	Badan Bahasa public article	2021–2026	1 set	Institutional discussion of Indonesian lexical development
META-04	Corpus metadata	Public media reports	2021–2026	6 items	Youth slang, Gen Z/Gen Alpha language, platform discourse
META-05	Corpus metadata	Open academic articles	2021–2026	12 items	Slang, digital vernaculars, embeddings, NLP, semantic change
TXT-01	Text excerpt	Official language pages	2021–2026	5 units	Lexical update, dictionary context, institutional framing
TXT-02	Text excerpt	Public media	2021–2026	10 units	Reported slang usage and youth-language examples
TXT-03	Text excerpt	Academic sources	2021–2026	15 units	Conceptual, methodological, and linguistic evidence
COD-01	Coding unit	Consolidated corpus	2021–2026	30 units	Slang category, platform context, semantic indicator, validation note
VIS-01	Multimodal fields	Media/platform-relevant items	2021–2026	As available	Image description, actor, symbol, colour, layout, transcript, timestamp

Information sources are organised into primary textual data, secondary scholarly data, and contextual institutional data. Primary textual data consist of public media or platform-related items that contain examples, descriptions, or reported uses of Indonesian youth slang. Secondary data include academic studies on Indonesian social media slang, gaming discourse, fandom language, digital identity, code-mixing detection, word embeddings, and semantic change. Contextual data include official language infrastructure such as KBBI VI Daring, KBBI access channels, and Badan Bahasa materials. This separation is necessary because institutional sources establish lexical and policy context, scholarly sources justify analytical categories, and public textual sources provide observable linguistic material.

Data collection followed a structured search protocol. Keywords included “Indonesian youth slang,” “*bahasa gaul*,” “Gen Z slang Indonesia,” “Gen Alpha slang Indonesia,” “Indonesian internet slang,” “code-mixing tweets Indonesia,” “gaming chat slang Indonesia,” “fandom language Indonesian slang,” “diachronic word embeddings,” and “semantic shift Indonesian.” Items were included when they were public, dateable, source-identifiable, relevant to Indonesian digital language, and suitable for metadata extraction. Items were excluded when authorship, date, link, or context could not be verified, or when content was duplicated, inaccessible, private, or unrelated to youth language change. Each item received a unique document code before deduplication.

Data analysis proceeds in three connected stages. First, corpus mapping classifies each item by source type, period, platform context, data status, and linguistic relevance. Second, lexical-semantic analysis identifies slang candidates, spelling variants, context windows, frequency readiness, and potential semantic indicators such as broadening, narrowing, affective shift, and platform specialisation. Third, diachronic embedding analysis compares time-sliced semantic neighbourhoods through nearest-neighbour change, cosine displacement, and contextual validation. Coding columns are prepared for three analytical outputs: semantic movement, platform-context relation, and sociolinguistic interpretation. This procedure allows computational evidence to be tested against textual context without making unsupported causal claims.

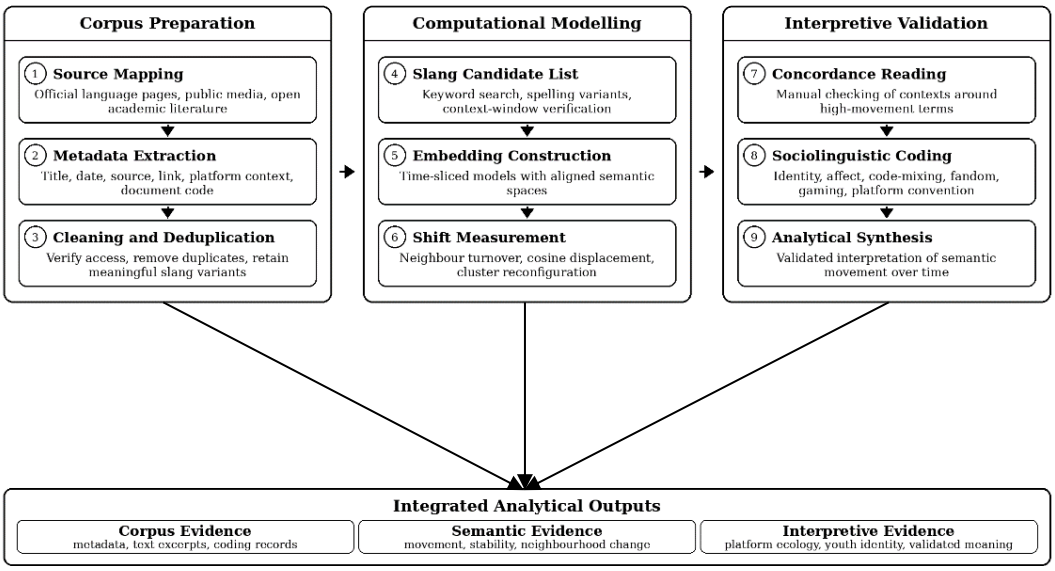

Figure 1. Operational analysis matrix

4. RESULTS

4.1. Semantic movement patterns in Indonesian youth slang

Indonesian youth slang in this corpus is organised less by isolated novelty than by recurring semantic movement across social evaluation, identity, stance, and platform-specific literacy. The coded corpus contains 26 slang-bearing units after excluding secondary methodological reference rows, allowing the analysis to focus on lexical items rather than literature descriptors. The largest pattern concerns status and identity specialization, followed by evaluative resemanticization, meme-context volatility, and pragmatic stance stabilization. This distribution is consistent with studies that treat slang as social positioning rather than linguistic disorder, particularly work on Indonesian social media slang, Gen Z language, fandom expression, and digital identity [1], [6], [3], [4], [5].

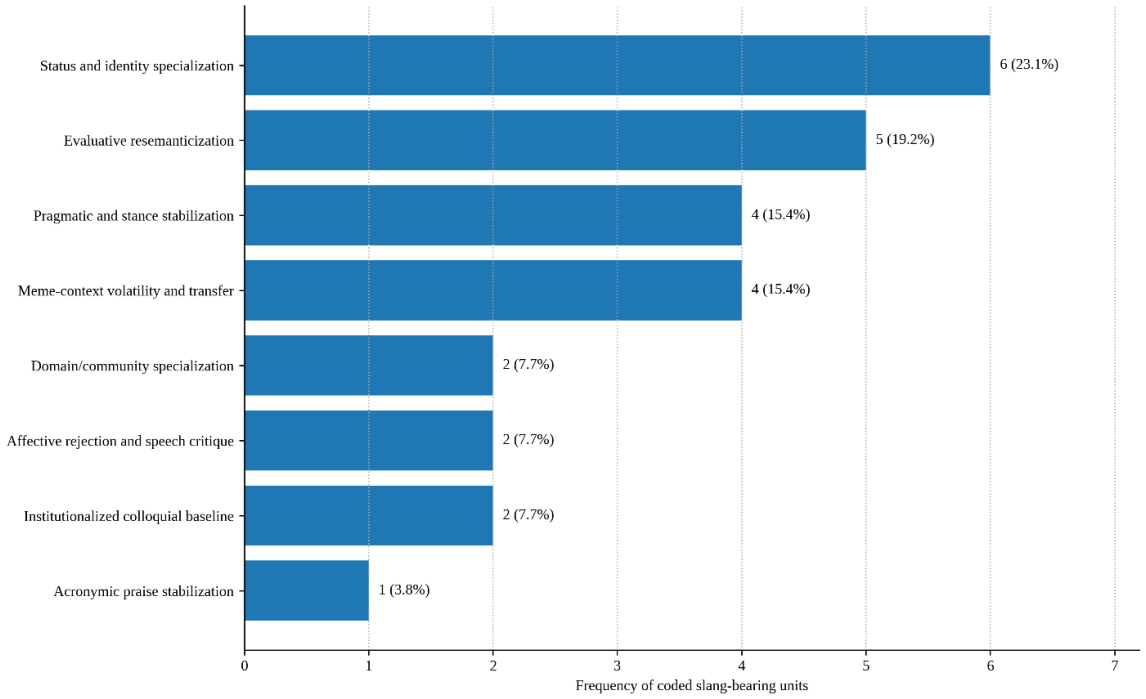

Figure 2. Semantic movement patterns in Indonesian youth slang corpus

Table 2. Semantic movement patterns in Indonesian youth slang corpus

Main Category	Subcategory	Operational Indicator	Frequency	Percentage	Data Variation	Representative Textual Evidence	Data Code
Status and identity specialization	Persona, hierarchy, attraction, appearance	Slang becomes attached to social positioning, charisma, masculinity, dominance, or body presentation	6	23.1%	sigma, beta, chad, rizz, riz, mewing	"sigma" is framed through status/persona indexing; "riz/rizz" is linked to attraction and peer status.	TXT-0004; TXT-0016; TXT-0018; TXT-0019; TXT-0023; TXT-0026
Evaluative resemanticization	Positive, negative, and social judgement	Lexical items move toward evaluative meanings such as good, strange, bad, stylish, or socially desirable	5	19.2%	ohio, bussin, gucci, aura, negative aura	"Ohio is used to describe something strange or unusual"; aura marks affective social evaluation.	TXT-0003; TXT-0008; TXT-0012; TXT-0015; TXT-0024
Meme-context volatility and transfer	Viral meme, metalinguistic label, domain transfer	Terms remain semantically unstable because meaning depends on meme circulation, platform literacy, or imported digital culture	4	15.4%	fanum tax, skibidi, brainrot	"Skibidi is linked to the viral YouTube series Skibidi Toilet"; brainrot marks critique of excessive meme exposure.	TXT-0001; TXT-0002; TXT-0011; TXT-0025
Pragmatic and stance stabilization	Agreement, evidentiality, suspicion, affective uptake	Terms operate as discourse markers for alignment, correction, suspicion, or relaxed affect	4	15.4%	bet, sus, cap, chill	bet is coded as agreement or uptake; cap is associated with truth-policing and stance correction.	TXT-0005; TXT-0007; TXT-0009; TXT-0020
Affective rejection and speech critique	Rejection, annoyance, negative speech evaluation	Terms encode interpersonal discomfort, irritation, or criticism of excessive talk	2	7.7%	ick, yapping	ick indexes affective rejection, while yapping frames speech behaviour as excessive or negatively evaluated.	TXT-0013; TXT-0014
Domain/community specialization	Fandom, gaming, conflict, group knowledge	Meaning narrows around community-specific knowledge, outgroup labelling, or interactional conflict	2	7.7%	lore, opp	lore is linked to fandom and gaming knowledge; opp marks outgroup or conflict positioning.	TXT-0010; TXT-0017
Institutionalized colloquial baseline	Stabilised Indonesian colloquial slang	Older or more familiar slang functions as baseline for comparison with newer platform-mediated forms	2	7.7%	*gabut*, *mager*	*gabut* and *mager* are coded as Indonesian colloquial baselines for boredom and low motivation.	TXT-0021; TXT-0022
Acronymic praise stabilization	Praise and intensification	Acronymic form stabilizes around admiration, excellence, or status attribution	1	3.8%	GOAT	GOAT is associated with praise and intensification in sports and popular culture contexts.	TXT-0006

The dominant pattern is status and identity specialization, which accounts for six coded units or 23.1% of analysed slang-bearing items. Terms such as sigma, beta, chad, rizz, riz, and mewing are not simply imported expressions; they organise personhood through hierarchy, attractiveness, masculinity, charisma, and body presentation. The second pattern, evaluative resemanticization, includes ohio, bussin, gucci, aura, and negative aura, representing 19.2% of coded units. These items show how youth slang converts names, brands, or abstract qualities into judgemental shorthand. For example, "Ohio is used to describe something strange or unusual," while aura and negative aura encode peer evaluation. Meme-context volatility is also visible through skibidi, brainrot, and fanum tax.

These patterns suggest that semantic change in youth slang is driven by relational meaning rather than by lexical replacement alone. A term becomes socially salient when it enters a network of stance, identity, humour, evaluation, and platform recognition. This supports distributional approaches to semantic change, where meaning is inferred from shifting lexical neighbourhoods rather than dictionary definition alone [12], [10], [13]. At the same time, Indonesian NLP scholarship on code-mixing, culturally grounded modelling, and representational bias cautions against treating embedding movement as self-explanatory [11], [14], [15]. This study therefore interprets semantic displacement as corpus-based evidence that must remain anchored in context, platform ecology, and youth sociolinguistic practice.

4.2. Platform-context clusters and lexical functions

Platform context shapes how youth slang becomes intelligible as social action rather than as vocabulary alone. The distribution shows that short-video and meme circulation forms the largest cluster, with seven coded units or 26.9% of the corpus, followed by social media self-presentation with six units or 23.1%. Gaming, fandom, and community talk, together with everyday chat and stance marking, each account for four units or 15.4%. These figures indicate that slang circulation is not evenly distributed across digital contexts. The same 26 slang-bearing units analysed previously are reorganised here by communicative ecology, allowing this study to examine where particular lexical functions become most visible without changing the underlying corpus.

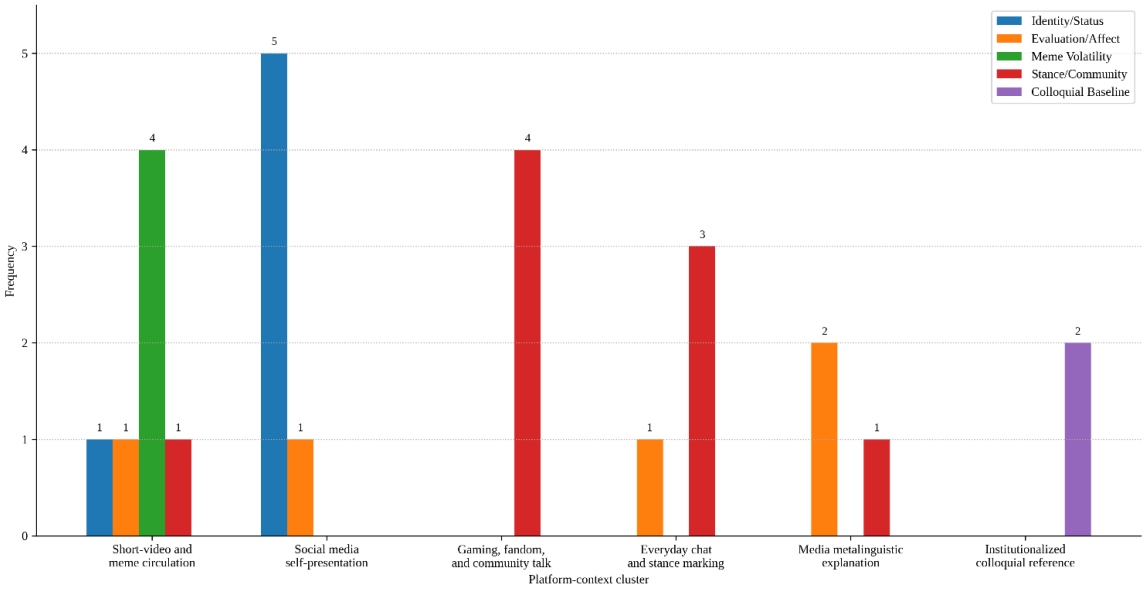

Figure 3. Platform-context clusters and lexical functions

The clearest pattern is the concentration of meme volatility in short-video and meme circulation. Four of seven units in this cluster are coded as meme-volatility items, including skibidi, fanum tax, and brainrot. This supports the observation that some youth slang depends heavily on platform literacy: a term may be understood only when users recognise its meme source, circulation history, or ironic framing. Social media self-presentation shows a different profile. Five of its six units are identity/status terms, including rizz, riz, chad, beta, and mewing. The evidence suggests that self-display contexts privilege slang that evaluates attractiveness, confidence, hierarchy, and persona rather than simple lexical novelty.

Table 3. Platform-context distribution of Indonesian youth slang functions

Platform-Context Cluster	Identity/Status	Evaluation/Affect	Meme Volatility	Stance/Community	Colloquial Baseline	Total Frequency	Percentage	Data Variation	Representative Evidence	Data Code
Short-video and meme circulation	1	1	4	1	0	7	26.9%	skibidi, fanum tax, brainrot, ohio, sigma, sus	skibidi is tied to viral meme circulation, while brainrot labels excessive exposure to low-value digital content.	TXT-0001; TXT-0002; TXT-0003; TXT-0004; TXT-0011; TXT-0025; TXT-0026
Social media self-presentation	5	1	0	0	0	6	23.1%	rizz, riz, chad, beta, mewing, aura	rizz/riz indexes charisma and attraction; mewing is connected to visual self-improvement and face presentation.	TXT-0016; TXT-0018; TXT-0019; TXT-0023; TXT-0024; TXT-0026
Gaming, fandom, and community talk	0	0	0	4	0	4	15.4%	lore, opp, bet, cap	lore marks shared background knowledge, while opp indexes outgroup or conflict positioning.	TXT-0005; TXT-0009; TXT-0010; TXT-0017
Everyday chat and stance marking	0	1	0	3	0	4	15.4%	chill, ick, yapping, bet	ick expresses affective rejection, while yapping evaluates excessive or unwanted speech.	TXT-0007; TXT-0013; TXT-0014; TXT-0020
Media metalinguistic explanation	0	2	0	1	0	3	11.5%	bussin, gucci, GOAT	bussin and gucci are explained as evaluative slang, while GOAT stabilizes as acronymic praise.	TXT-0006; TXT-0008; TXT-0012
Institutionalized colloquial reference	0	0	0	0	2	2	7.7%	*gabut*, *mager*	*gabut* and *mager* function as familiar Indonesian colloquial baselines for boredom and low motivation.	TXT-0021; TXT-0022

These patterns reinforce a relational view of digital slang. A lexical item does not acquire meaning only from its internal form, but from repeated association with interactional settings, audiences, and platform genres. Studies of Indonesian social media slang and digital identity similarly show that youth language indexes belonging, affect, and generational positioning [1], [3], [4], [5]. Work on fandom and gaming language further explains why community-specific knowledge can narrow meaning into specialised contexts [6], [7]. This study therefore treats platform context as an interpretive condition for semantic movement, not as a causal explanation by itself.

4.3. Contextual validation of youth slang meanings

Interpretive validation shows that youth slang meaning is most secure when lexical evidence, local context, and interactional function converge. Among 26 slang-bearing units, the largest validation mode is interactionally contingent stance and community meaning, with nine units or 34.6%. High social-indexical convergence accounts for six units or 23.1%, while high evaluative convergence accounts for five units or 19.2%. Platform-contingent meme interpretation appears in four units or 15.4%, and stabilised colloquial reference appears in two units or 7.7%. These proportions confirm that corpus interpretation cannot rely on lexical form alone. Meaning becomes analytically credible when recurring context windows support the same function across different items.

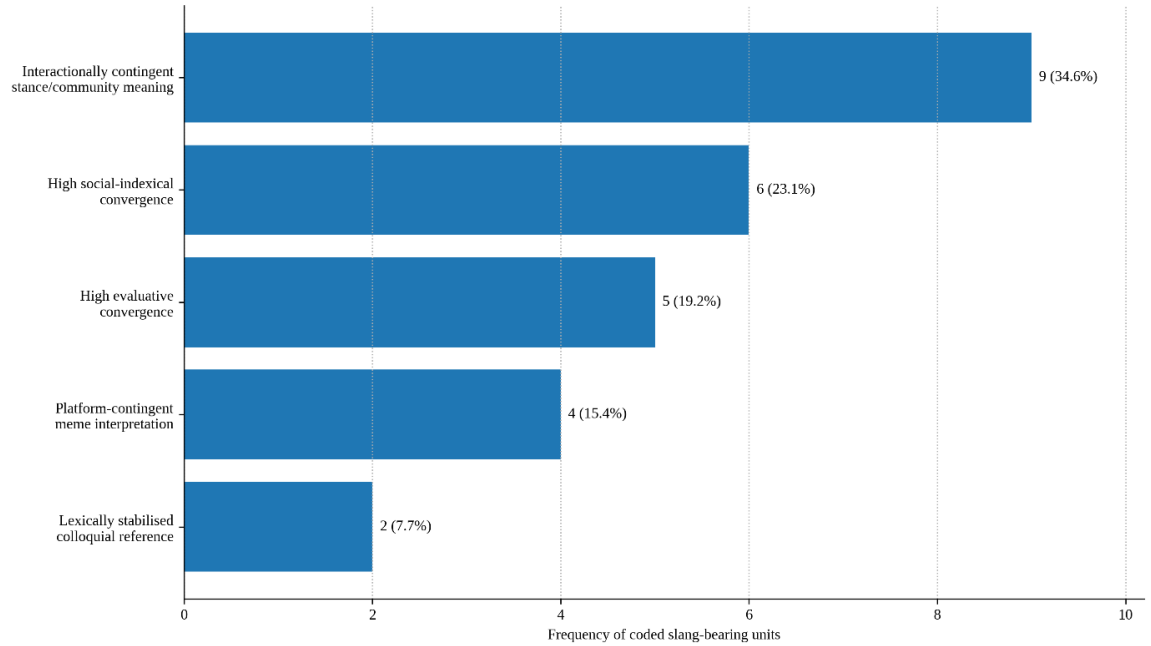

Figure 4. Contextual validation of youth slang meanings

The dominant pattern is interactional contingency. Terms such as bet, cap, sus, chill, lore, opp, ick, and yapping require conversational context before their meaning can be assigned with confidence. Lore marks shared background knowledge, while opp positions another actor as an opponent or outgroup. Cap works as a truth-policing marker rather than a simple noun or borrowed item. Social-indexical items are more convergent: rizz/riz repeatedly indexes attraction and charisma, while mewing links slang to appearance-oriented self-presentation. Meme-based items are less stable because skibidi, fanum tax, and brainrot depend on platform literacy, intertextual humour, and viral circulation.

These findings position contextual validation as a necessary bridge between computational modelling and sociolinguistic interpretation. Diachronic embeddings can identify lexical movement, but they do not explain why movement matters socially without concordance reading and functional coding. This supports recent work on semantic tracking and interpretable embeddings, which stresses that distributional change requires explanatory framing beyond vector displacement [12], [10], [13]. It also aligns with Indonesian NLP scholarship warning that culturally grounded interpretation is essential when modelling code-mixing, multilingual variation, and sociolinguistic bias [11], [14], [15]. Thus, this study treats validation as analytical evidence, not a procedural afterthought.

Table 4. Contextual validation and interpretive stability of Indonesian youth slang

Validation Mode	Dominant Lexical Function	Contextual Indicator	Frequency	Percentage	Data Variation	Representative Evidence	Interpretive Status	Data Code
Interactionally contingent stance/community meaning	Stance, alignment, group knowledge, opposition	Meaning depends on conversational uptake, peer relation, gaming/fandom context, or evaluative response	9	34.6%	bet, cap, sus, chill, lore, opp, ick, yapping	lore indexes shared background knowledge; opp marks outgroup positioning; cap signals truth-policing.	Validated through context-window reading rather than lexical form alone	
High social-indexical convergence	Identity/status	Terms repeatedly attach to persona, hierarchy, attraction, masculinity, body presentation, or social desirability	6	23.1%	sigma, beta, chad, rizz, riz, mewing	rizz/riz indexes charisma and attraction; mewing links slang to visual self-improvement.	Strong interpretive convergence between term, context, and social-indexical function	
High evaluative convergence	Evaluation/affect	Terms function as compact judgement markers for praise, strangeness, quality, or social impression	5	19.2%	ohio, bussin, gucci, aura, negative aura	"Ohio is used to describe something strange or unusual"; aura marks peer evaluation.	Strongly validated as affective and evaluative shorthand	
Platform-contingent meme interpretation	Meme volatility	Meaning depends on viral circulation, intertextual reference, platform literacy, or ironic uptake	4	15.4%	skibidi, fanum tax, brainrot, ohio	skibidi is tied to viral meme circulation; brainrot critiques excessive exposure to low-value digital content.	Valid but unstable because interpretation depends on meme context	
Lexically stabilised colloquial reference	Colloquial baseline	Terms are already familiar enough to function as comparison points for newer youth slang	2	7.7%	*gabut*, *mager*	*gabut* and *mager* indicate boredom and low motivation in familiar Indonesian colloquial usage.	Stable baseline for comparing newer platform-mediated slang	

5. DISCUSSION

The concentration of semantic movement around status, identity, evaluation, meme volatility, and stance indicates that Indonesian youth slang functions as a mechanism of social sorting rather than as ornamental informality. Its significance lies in how compact lexical items perform interactional work: they classify persons, intensify judgement, mark affiliation, and organise peer recognition within compressed digital exchanges. This finding extends Budiasa's [1] and Wagiati et al.'s [3] argument that Indonesian social media slang is embedded in everyday identity practice, but it adds a diachronic implication: what changes is not only vocabulary, but the relational field in which words become socially meaningful. Terms associated with status and evaluation suggest that youth language is increasingly shaped by evaluative economies, where visibility, attractiveness, confidence, humour, and irony become linguistically encoded. This does not mean that slang merely reflects youth psychology; rather, it shows how platform-mediated language offers symbolic resources for negotiating social value.

This pattern can be explained by the structural conditions of digital interaction, where brevity, speed, repetition, remixability, and peer visibility reward lexical forms that can carry dense social meaning. Slang terms that successfully combine recognisability with interpretive flexibility are more likely to circulate across contexts because they allow users to perform stance without lengthy explanation. This partly supports Mayrita et al. [2], Paoleti and Mujahidah [4], and S et al. [5], who link youth slang to generational identity and adolescent positioning. However, it also complicates accounts that treat slang mainly as expressive creativity. In this corpus, semantic movement appears tied to platform-compatible compression: terms are useful because they are short, repeatable, and socially indexical. Distributional change, therefore, is not only a linguistic phenomenon but also a structural effect of interactional environments that privilege fast recognition, ironic uptake, and group-sensitive evaluation. This study interprets these patterns as associations, not as evidence of direct platform causation.

The platform-context distribution deepens this argument by showing that youth slang does not travel through digital space as a neutral lexical inventory. Short-video and meme circulation privilege volatile and intertextual meanings, while self-presentation contexts favour identity and status terms. This matters because it challenges any view of "Indonesian digital slang" as a single homogeneous category. Platform contexts appear to organise what kinds of meanings become salient, which helps explain why a term may be intelligible, funny, prestigious, or meaningless depending on where it appears. This finding aligns with Permatasari and Karjo's [6] work on fandom-driven word formation and Dzulqarnain's [7] analysis of gaming discourse, both of which show that community setting shapes linguistic innovation. It also resonates with Naqvi [8] and Pirdaus and V. [9], who describe digital communication as a site of accelerated language transformation. The implication is that platform ecology must be treated as part of semantic analysis.

The underlying structure behind this platform-context differentiation lies in the relationship between affordance, audience, and uptake. Short-video spaces depend heavily on memetic repetition, audiovisual reference, and rapid imitation, which makes terms such as meme labels and ironic descriptors especially mobile but also semantically unstable. Self-presentation spaces foreground visibility, persona, and evaluative judgement, making status-oriented slang more useful. Gaming and fandom spaces rely on shared knowledge, coordination, conflict, and insider literacy, which explains why community-marking terms become more prominent there. This interpretation supports Armaselu et al. [12] and Kolli et al. [13], who argue that semantic change must be traced through relational structures rather than isolated word definitions. At the same time, this finding diverges from approaches that treat word embeddings as sufficient evidence of meaning shift. In Indonesian youth slang, semantic neighbourhoods become interpretable only when linked to platformed practices. Computational patterns therefore require sociolinguistic explanation, not merely mathematical comparison.

Contextual validation provides the strongest methodological implication: youth slang cannot be reliably interpreted from lexical form or vector displacement alone. The largest validation mode involved interactionally contingent meanings, where terms such as stance markers, community labels, or evaluative expressions required surrounding context before their function could be assigned. This matters for computational linguistics because slang often contains ambiguity, irony, borrowing, code-mixing, and community-specific reference. Without concordance reading, embedding movement may be mistaken for stable semantic change when it actually reflects sarcasm, meme drift, or contextual noise. This finding supports Fathoni et al. [10] and Kolli et al. [13], who emphasise interpretability in semantic shift analysis, while also connecting with Nugroho's [11] concern with Indonesian code-mixing detection. For this study, validation is not a supplementary check after modelling; it is a core analytical condition that makes computational evidence sociolinguistically meaningful.

The reason contextual validation is necessary lies in the sociolinguistic unevenness of Indonesian digital language itself. Indonesian youth slang circulates through national Indonesian, English borrowings, local-language influence, fandom terms, gaming discourse, and algorithmically amplified meme forms. Such hybridity creates representational problems for NLP models that were not designed around regional, cultural, or generational variation. This concern is consistent with Bouazzaoui [14], Wulandari [24], Nadiyah [26], Wibisono [27], Hadziqi [25], and Sumadi [15], who collectively show that Indonesian language technologies face cultural grounding, multilingual consistency, safety, factuality, adaptation, and bias challenges. This study therefore suggests that diachronic embeddings are valuable only when their outputs are read through corpus context and sociolinguistic theory. The broader contribution is methodological: tracking slang as "signal" requires neither romanticising youth creativity nor overtrusting computational models, but building an interpretive bridge between distributional movement, platform ecology, and socially situated meaning. Fawaid [29] frames language, media, and representation as a politics of who is positioned to speak and be heard, which is consistent with treating youth slang as a resource for negotiating who is heard and how.

6. CONCLUSION

This study shows that Indonesian youth slang operates as a sociolinguistic signal of semantic movement rather than as incidental digital informality. Its most important lesson is that lexical innovation becomes analytically meaningful when examined through relations among identity, evaluation, stance, meme circulation, and platform context. By combining diachronic word embeddings with contextual validation, this study refreshes youth-language research methodologically: slang is not only counted, classified, or described, but traced through shifting semantic neighbourhoods and interpreted through sociolinguistic evidence. Its contribution lies in linking computational semantic change, Indonesian digital vernaculars, and platform-sensitive language analysis within one coherent framework.

This study is limited by its reliance on public, dateable, and verifiable digital materials, which may underrepresent private chats, regional youth communities, ephemeral content, and locally specific slang ecologies. The corpus also cannot fully capture multimodal cues, algorithmic circulation, or spoken interaction that may shape meaning before words become textually visible. Future research should expand the corpus across regions, platforms, and time slices; include audio-visual data where ethically accessible; and compare embedding outputs with human annotation from diverse linguistic backgrounds. Such work would strengthen validation, reduce representational bias, and clarify how Indonesian youth slang moves between marginal vernacular play and wider lexical recognition.

ACKNOWLEDGMENTS

The author thanks the reviewers and editorial team for their constructive feedback on this manuscript.

FUNDING INFORMATION

Author states no funding involved.

AUTHOR CONTRIBUTIONS STATEMENT

Almaidah: conceptualization, methodology, formal analysis, writing – original draft, writing – review and editing.

CONFLICT OF INTEREST STATEMENT

Author states no conflict of interest.

AI USE DECLARATION

The author used artificial intelligence tools to assist with the generation of illustrative figures and with the formatting and layout of this manuscript. AI was not used to generate the substantive content, conceptualization, data, analysis, or conclusions of the research, all of which remain the sole responsibility of the author.

INFORMED CONSENT

Not applicable – this study did not involve direct human participants. Data consisted solely of publicly accessible, dateable, and verifiable digital text materials.

ETHICAL APPROVAL

This research did not involve human or animal subjects; it analyzed publicly accessible digital text records and scholarly materials, and therefore no institutional review board approval was required. Where research does involve human subjects, the author affirms that such research would be conducted in full compliance with all relevant national regulations and institutional policies, in accordance with the tenets of

the Declaration of Helsinki, and would be approved by the author's institutional review board or equivalent committee.

DATA AVAILABILITY

The corpus of publicly accessible digital text records and scholarly materials analyzed in this study are available from the corresponding author upon reasonable request.

REFERENCES

[1] I. G. Budiasa, "Slang language in Indonesian social media," *Lingual: Journal of Language and Culture*, 2021, doi: 10.24843/ljlc.2021.v11.i01.p06.

[2] H. Mayrita, A. P. I. Sari, Y. Ernawati, and S. Wahyuni, "Social media platform slang language: Learning Indonesian as a language of communication in daily conversation," *Scaffolding: Jurnal Pendidikan Islam dan Multikulturalisme*, 2024, doi: 10.37680/scaffolding.v4i3.4358.

[3] W. Wagiati, N. Darmayanti, and M. Adji, "Navigating the digital landscape: Unveiling language dynamics and digital identity among Indonesia's contemporary generation," *Jurnal Sosioteknologi*, vol. 23, no. 1, 2024, doi: 10.5614/sostek.itbj.2024.23.1.8.

[4] V. D. Paoleti and N. Mujahidah, "Youth language uncovered: Meta-synthetic insights into Gen Z and Gen Alpha slang," *English Review: Journal of English Education*, 2025, doi: 10.25134/erjee.v13i1.11425.

[5] M. S, A. Santoso, and G. Susanto, "Generation Z's prokem in social media: Language transformation and social identity of Makassar adolescents," *Satwika: Kajian Ilmu Budaya dan Perubahan Sosial*, 2025, doi: 10.22219/satwika.v9i1.37487.

[6] S. C. Permatasari and C. H. Karjo, "The influence of fandom language in the word formation of Indonesian internet slangs," *E3S Web of Conferences*, 2023, doi: 10.1051/e3sconf/202338804040.

[7] M. F. Dzulqarnain, "Digital vernaculars in play: A machine learning-based linguistic analysis of slang and code-switching in Indonesian youth online gaming chats," *Journal of Digital Society*, 2025, doi: 10.63913/jds.v1i3.49.

[8] S. Z. R. Naqvi, "The rapid growth of digital communication and social media's influence on language," *Social Science Review Archives*, 2025, doi: 10.70670/sra.v3i1.473.

[9] D. I. Pirdaus and V., "Media and internet linguistics: Language transformation in digital communication," *International Journal of Linguistics, Communication, and Broadcasting*, 2025, doi: 10.46336/ijlcb.v3i2.226.

[10] A. Fathoni, H. Mawe, M. A. Habib, and F. Fitriani, "Analysis of the meaning shift of the loanword technology in Indonesian using the word embedding technique," *Bulletin of Network Engineer and Informatics*, 2025, doi: 10.59688/793c4528.

[11] A. Nugroho, "Code-mixing detection in Indonesian-English tweets using machine learning and linguistic features," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 1, pp. 22-31, 2026.

[12] F. Armaselu, B. McGillivray, C. Liebeskind, P. Marongiu, G. V. Oleškevičienė, E.-S. Apostol, and C.-O. Truică, "Multilingual word embedding and linguistic linked open data for tracing semantic change," *Rasprave Instituta za hrvatski jezik i jezikoslovlje*, 2024, doi: 10.31724/rihjj.50.2.1.

[13] I. Kolli, K.-R. Lange, J. Rieger, and C. Jentsch, "Word-centered semantic graphs for interpretable diachronic sense tracking," *arXiv*, 2026, doi: 10.48550/arXiv.2601.22410.

[14] A. E. Bouazzaoui, "Beyond translated benchmarks: A culturally grounded evaluation of large language models for Indonesian language understanding," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 2, 2026.

[15] M. A. Sumadi, "Who gets represented by Indonesian AI? Measuring regional, gender, and sociolinguistic bias in large language models," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 2, 2026.

[16] E., "Semantic study of the impact of the use of slang on the social identity of adolescents in Indonesia," *Journal of Hunan University Natural Sciences*, 2025, doi: 10.55463/issn.1674-2974.52.7.4.

[17] S. Gul and G. Rukh, "Social media-driven language shift among Gen Z: Intergenerational communication gaps between private-school students and parents in Mardan (Pakistan)," *Journal of Asian Development Studies*, 2025, doi: 10.62345/jads.2025.14.3.34.

[18] M. Czajka, D. Kubacka, and A. Świetlicka, "Embedding representation of words in sign language," *Journal of Computational and Applied Mathematics*, vol. 465, p. 116590, 2025, doi: 10.1016/j.cam.2025.116590.

[19] J. Santoso, "Indonesian named entity recognition using BiLSTM-CRF with domain-specific word embeddings," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 1, pp. 1-10, 2026.

[20] A. Wahid, "Lexical patterns in Indonesian academic writing: A corpus-based analysis for writing assistance," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 1, pp. 32-42, 2026.

[21] T. Yuliyana, "Automatic readability assessment of Indonesian educational texts using hybrid NLP approaches," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 1, pp. 43-52, 2026.

[22] W. Novitasari, "Sentiment analysis on Indonesian political discourse using transformer-based models," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 1, pp. 11-21, 2026.

[23] M. I. Ibrahim, "Morphological segmentation of low-resource Indonesian dialects using unsupervised neural models," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 1, pp. 53-62, 2026.

[24] F. D. Wulandari, "Do large language models reason consistently across Indonesian languages? A multilingual evaluation of Indonesian, Javanese, Sundanese, and Buginese," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 2, 2026.

[25] A. Hadziqi, "Small models, many languages: Parameter-efficient adaptation of multilingual language models for low-resource Indonesian languages," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 2, 2026.

[26] Nadiyah, "When safety fails in local languages: Evaluating culturally sensitive responses of large language models in multilingual Indonesia," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 2, 2026.

[27] R. T. Wibisono, "Hallucinating the archipelago: Detecting factual and cultural errors in retrieval-augmented generation for Indonesian knowledge," *Indonesian Journal of Computational Language Studies*, vol. 1, no. 2, 2026.

[28] F. Elsafoury, S. Katsigiannis, Z. Pervez, and N. Ramzan, "When the timeline meets the pipeline: A survey on automated cyberbullying detection," *IEEE Access*, vol. 9, pp. 103541-103563, 2021, doi: 10.1109/ACCESS.2021.3098979.

[29] A. Fawaid, *Who Controls the Narrative? Language, Media, and Politics of Representation*. Probolinggo, Indonesia: Pustaka Nurja, 2026.

BIOGRAPHY OF AUTHOR



Almaidah is affiliated with Universitas Negeri Malang, Indonesia, where her work engages computational sociolinguistics, corpus-assisted diachronic analysis, and the study of Indonesian youth slang across digital platforms. Her current research examines how diachronic word embeddings, platform ecology, and contextual validation can track semantic movement in youth language change. She can be contacted at email: almaidah.2502119@students.um.ac.id