Mapping language at scale: Combining street-view imagery, geospatial analysis, and machine learning in the study of Indonesian urban landscapes

Article Info

Article history:

Received month dd, yyyy
Revised month dd, yyyy
Accepted month dd, yyyy

Keywords:

GeoAI
linguistic landscape
machine learning
spatial analysis
street-view imagery

ABSTRACT

Urban linguistic landscapes remain difficult to examine at metropolitan scale because conventional fieldwork offers interpretive depth but limited spatial coverage, while automated image analysis often overlooks linguistic and social context. This study aims to integrate street-view imagery, machine-assisted text recognition, geospatial analysis, and multimodal interpretation to map publicly visible languages across differentiated urban environments in Jakarta. This study applies a spatially explicit multimethod design to ninety validated signs extracted from nine eligible public images, combining sign detection, OCR, language identification, human validation, corridor-based mapping, and social-semiotic coding. Findings indicate that machine-readable visibility is uneven, with standardized regulatory and commercial signs more reliably detected than temporary, occluded, stylized, or mixed-script signs. Spatial patterns show Indonesian concentrated in regulatory corridors, English more visible in corporate and retail settings, and Chinese localized within Glodok's commercial landscape. Institutional analysis further demonstrates that language hierarchy is produced through sign ownership, communicative function, placement, typography, colour, materiality, and logo positioning rather than frequency alone. This study contributes a reproducible framework that connects computational accuracy, exposure-aware spatial comparison, and socially grounded interpretation while preventing linguistic visibility from being conflated with speaker population, language vitality, or causal urban influence within a transparent and human-validated urban GeoAI research architecture.

Corresponding Author:

INTRODUCTION

Jakarta offers a consequential setting for examining how language becomes visible, ordered, and unevenly distributed across urban space. In 2023, DKI Jakarta contained 10.67 million residents at a density of 16,146 people per square kilometre, intensifying encounters among governmental, commercial, transnational, and community discourses within a built environment (BPS-Statistics DKI Jakarta Province, 2024). These encounters materialize through road signs, storefronts, advertising, transport information, institutional façades, murals, and temporary protest media. Yet linguistic traces are rarely documented systematically at metropolitan scale. Conventional field surveys provide interpretive depth but remain costly, limited, and difficult to reproduce across road networks. Meanwhile, street-view imagery enables fine-grained observation of urban form, perception, and land use at scale (Zhang et al., 2018; Rangel et al., 2022; Liu et al., 2023). Consequently, Jakarta's density and semiotic heterogeneity require an approach that preserves linguistic interpretation while expanding spatial coverage, computational consistency, and methodological transparency across differentiated urban corridors.

Existing urban analytics research demonstrates that street-level imagery can support scene recognition, neighbourhood representation, land-use classification, façade analysis, and explicit interpretation. Multimodal models have integrated visual features, points of interest, building footprints, and semantic information to identify urban functions and morphological patterns (Wang et al., 2020; Xu et al., 2022; Li et al., 2024). Work emphasizes spatially explicit and explainable GeoAI, addressing how predictions vary geographically and which visual features shape model outputs (Liu & Biljecki, 2022; Liu et al., 2023). In parallel, Indonesian linguistic-landscape studies have examined identity, graffiti, political typography, tourism discourse, and environmental messages through multimodal interpretation (Aini, 2026; Sartono, 2026). However, these traditions remain weakly connected. Urban-computing studies treat textual signs as incidental visual objects, whereas linguistic-landscape research seldom develops reproducible metropolitan-scale pipelines linking sign detection, language identification, spatial clustering, and social-semiotic interpretation. This unresolved methodological divide defines this study's central intervention in urban language research.

This study therefore investigates how street-view imagery, geospatial analysis, and machine learning can be combined to map publicly visible language across Jakarta environments without reducing linguistic signs to decontextualized image features. It addresses three questions. First, how accurately can machine-assisted procedures detect sign regions, extract text, and classify languages and scripts under conditions of occlusion, perspective distortion, variation, and multilingual mixing? Second, how are validated signs distributed across streets, commercial frontages, transport spaces, institutional zones, and neighbourhood contexts, and what spatial patterns characterize linguistic concentration and diversity? Third, how do sign ownership, communicative function, language ordering, translation relations, typography, colour, layout, and logo placement organize unequal visibility among languages? Methodologically, this study treats images, signs, textual elements, and spatial units as distinct but linked analytical levels. Human validation remains integral because computational confidence cannot substitute for judgments concerning language boundaries, symbolic functions, institutional authority, or ambiguous multilingual forms in settings.

This study advances the argument that linguistic-landscape analysis becomes credible only when computational reach is coupled with spatial normalization, explainability, and social-semiotic validation. Street-view imagery can extend observational coverage, while machine learning can standardize candidate-sign detection and facilitate processing of visual records; nevertheless, neither technology independently explains why languages occupy prominent, peripheral, regulatory, commercial, or symbolic positions. Spatial distributions must therefore be interpreted relative to sampling exposure, image date, road type, urban function, and sign provenance rather than treated as direct indicators of speaker populations or language vitality. This position builds on responsible GeoAI scholarship, which stresses transparency, geographic heterogeneity, and uncertainty (Papadakis et al., 2022; Marasinghe et al., 2024), while retaining linguistic-landscape attention to institutional power and visual hierarchy. Accordingly, this study tests whether an integrated pipeline can produce scalable yet interpretable evidence about urban linguistic visibility without converting complex semiotic relations into technically precise but socially misleading classifications.

LITERATURE REVIEW

Linguistic landscape

Linguistic landscape refers to publicly visible language as materially embedded in signs, façades, advertisements, murals, and institutional displays. Earlier approaches often treated signage as a countable inventory of languages, whereas recent multimodal work understands it as a socially organized field in which typography, image, placement, and authorship shape meaning publicly. Indonesian studies similarly show that urban texts articulate identity, political persuasion, environmental discourse, tourism imaginaries, and vernacular belonging (Aini, 2026; Andini, 2026; Hidayat, 2026; Pahlawan, 2026; Sartono, 2026). Accordingly, linguistic visibility should not be equated with speaker numbers; it is a spatially mediated representation of authority, commerce, memory, and affiliation.

Its principal dimensions include language choice, script, ordering, translation, prominence, materiality, ownership, function, and spatial location. Top-down signs usually encode governmental or institutional authority, while bottom-up signs register commercial, community, activist, or artistic initiatives. Yet this distinction is not absolute because commercial signs may reproduce official norms, and community signs may circulate standardized branding in practice. Multimodal indicators therefore include font size, colour, logo placement, imagery, framing, permanence, and relation to surrounding architecture (Daroji, 2026; Sartono, 2026). For scalable analysis, these categories must remain separable: a sign's linguistic composition, institutional origin, visual hierarchy, and geographic position answer different analytical questions.

Street-view imagery

Street-view imagery is increasingly conceptualized as a form of urban sensing that captures pedestrian-level environments with greater semantic detail than overhead imagery. Unlike satellite data, which privilege land cover and morphology, street-level images record façades, shopfronts, traffic signs, and everyday visual cues. Research has used such imagery to measure perception, classify urban scenes, identify land use, and recover location semantics (Zhang et al., 2018; Fang et al., 2020; Rangel et al., 2022; Liu et al., 2023). However, street-view data are not transparent representations of cities: coverage, capture date, camera angle, occlusion, and platform-specific sampling selectively shape what becomes analytically visible.

Relevant analytical aspects include image provenance, temporal currency, spatial coverage, resolution, field of view, geolocation accuracy, and the detectability of textual objects. Computer-vision pipelines commonly combine object detection, semantic segmentation, OCR, visual embeddings, and multimodal representation learning (Wang et al., 2020; Xu et al., 2022; Arulananth et al., 2024; Li et al., 2024). Their outputs may support sign localization and language recognition, but error rates increase with stylized fonts, glare, motion blur, perspective distortion, mixed scripts, and partial obstruction. Consequently, image-level classification is insufficient; robust linguistic analysis requires sign-level extraction, text-level validation, and explicit records of uncertainty at every stage.

Geospatial artificial intelligence

Geospatial artificial intelligence denotes machine-learning approaches that preserve or model geographic location, scale, neighbourhood relations, and spatial heterogeneity. Some studies define GeoAI broadly as machine learning applied to geographic data, while spatially explicit accounts require predictions and explanations to remain linked to coordinates, networks, or areal units (Liu & Biljecki, 2022; Papadakis et al., 2022; Fleischmann & Arribas-Bel, 2024). This distinction matters because urban language is spatially uneven. Models that ignore location may classify signs accurately yet obscure corridor effects, clustering, and administrative boundaries. GeoAI is therefore most useful when computation augments, rather than replaces, spatial reasoning and interpretive accountability.

Its core indicators include geocoded observations, sampling exposure, spatial scale, density, diversity, adjacency, clustering, autocorrelation, model interpretability, and cross-area transferability. Urban studies increasingly integrate imagery, points of interest, mobility, building footprints, and semantic features to identify functional patterns (Chen et al., 2021; Huang et al., 2022; Guo et al., 2024; Qu et al., 2024; Zou et al., 2024). Yet responsible analysis must separate urban patterns from platform coverage and model bias (Marasinghe et al., 2024). This study therefore positions linguistic visibility as a multimodal, spatially explicit, and computationally assisted phenomenon requiring human validation, exposure-aware mapping, and socially grounded interpretation throughout.

METHOD

This study examines publicly visible linguistic signs embedded in street-level images of Jakarta. Its analytical units are nested: each source image constitutes an image unit; each bounded sign forms a sign unit; each readable inscription becomes a text unit; and each georeferenced point or street segment functions as a spatial unit. Corpus construction includes ten primary visual records, three methodological publications, and four official contextual sources. Nine primary images satisfy inclusion requirements, while one remains excluded because its location metadata are inconsistent. This layered structure permits alignment between visual evidence, extracted text, sign attributes, and urban location without conflating them.

Table 1. Composition of Dataset

Code	Source	Location or scope	Period	Number	Characteristics
PRI-CON	Wikimedia Commons	Central Jakarta: Wahid Hasyim–Thamrin and Sudirman corridors	2022–2025	2	Traffic, regulatory, corporate, directional, and commercial signage
PRI-EAS	Wikimedia Commons	Cipayung, East Jakarta	2022	1	Outer-urban street environment with

					commercial, institutional, and traffic text
PRI-SOU	Wikimedia Commons	Melawai and Gatot Subroto	2011–2024	2	Commercial branding, façade language, and historical traffic regulation
PRI-WES	Wikimedia Commons	Glodok, West Jakarta	2015	2	Market signage, ethnolinguistic visibility, street art, and transport-related text
PRI-POL	Wikimedia Commons	Jakarta; specific district unavailable	2011	1	Temporary placards, banners, protest discourse, and bottom-up signs
PRI-HIS	Wikimedia Commons	Jakarta mall; exact location unavailable	2004	1	Consumer branding and historical commercial multilingualism
PRI-EXC	Wikimedia Commons	Conflicting Jakarta–Ubud metadata	2015	1	Regulatory signage with unresolved geographic attribution
SEC-MET	Peer-reviewed journals	International methodological scope	2022–2024	3	Spatially explicit GeoAI, explainability, and street-view machine learning
CTX-OFF	Satu Data Jakarta, Jakarta.go.id, BPS DKI Jakarta	DKI Jakarta	Current sources	3	Administrative geography, demographic statistics, and contextual variables
CTX-IMG	KartaView	Jakarta coverage dependent on available sequences	Variable	1 platform	Geotagged imagery, sequence information, and capture metadata

This study adopts a multimethod, spatially explicit computational design integrating linguistic-landscape analysis, computer vision, and geospatial interpretation. Its purpose is not to automate interpretation, but to combine scalable detection with accountable human validation. Primary images are processed through sign-region identification, optical character recognition, script recognition, and language classification, while geospatial linkage situates validated signs within corridors, districts, and functional urban zones. Social-semiotic coding then examines ownership, communicative function, prominence, translation relations, colour, layout, and symbolism. Such integration follows GeoAI scholarship emphasizing geographic heterogeneity, multimodal evidence, explainability, and model use in urban research (Liu & Biljecki, 2022; Marasinghe et al., 2024).

Information derives from three evidential tiers. Primary data consist of publicly accessible street-level and storefront photographs hosted on Wikimedia Commons, covering Central, East, South, and West Jakarta and representing governmental, commercial, transport, artistic, and protest environments. KartaView is retained as an expansion source for geotagged imagery and sequence metadata, subject to licensing verification. Secondary data comprise peer-reviewed studies on street-view interpretation, spatially explicit GeoAI, and explainable machine learning. Contextual data come from Satu Data Jakarta, Jakarta.go.id, and BPS DKI Jakarta, providing administrative and socioeconomic reference points. These tiers remain distinct to prevent contextual indicators from being mistaken for linguistic evidence.

Data collection proceeds through source identification, keyword-based retrieval, eligibility screening, metadata capture, duplicate control, and visual-text preparation. Searches combine location and object terms, including Jakarta street sign, storefront, billboard, Glodok signage, Sudirman street, Cipayung street, and demonstration banner. Eligible items require a stable public source page, identifiable Jakarta location, recoverable date or authorship, and visible or potentially visible public text. Canonical URLs, normalized titles, author-date combinations, and duplicate groups support deduplication. Each retained item receives a unique document code, while titles, dates, links, contextual descriptions, screenshot references, and uncertainty notes are stored separately across metadata, text, and coding files systematically.

Analysis unfolds in three coordinated stages. First, machine-assisted extraction identifies candidate sign regions, generates OCR output, assigns scripts and languages, and records confidence values; human coders validate boundaries, normalize text separately, and classify errors using precision, recall, F1-score, and agreement measures. Second, validated signs are mapped by point, segment, corridor, and administrative unit to assess exposure-adjusted density, diversity, concentration, and clustering. Third, multimodal coding interprets sign origin, domain, function, language ordering, typographic prominence, translation, materiality, colour, layout, actors, symbols, and logo position. Final interpretation integrates these stages while distinguishing computational performance, spatial patterning, social-semiotic meaning, methodological transparency, and analytical uncertainty.

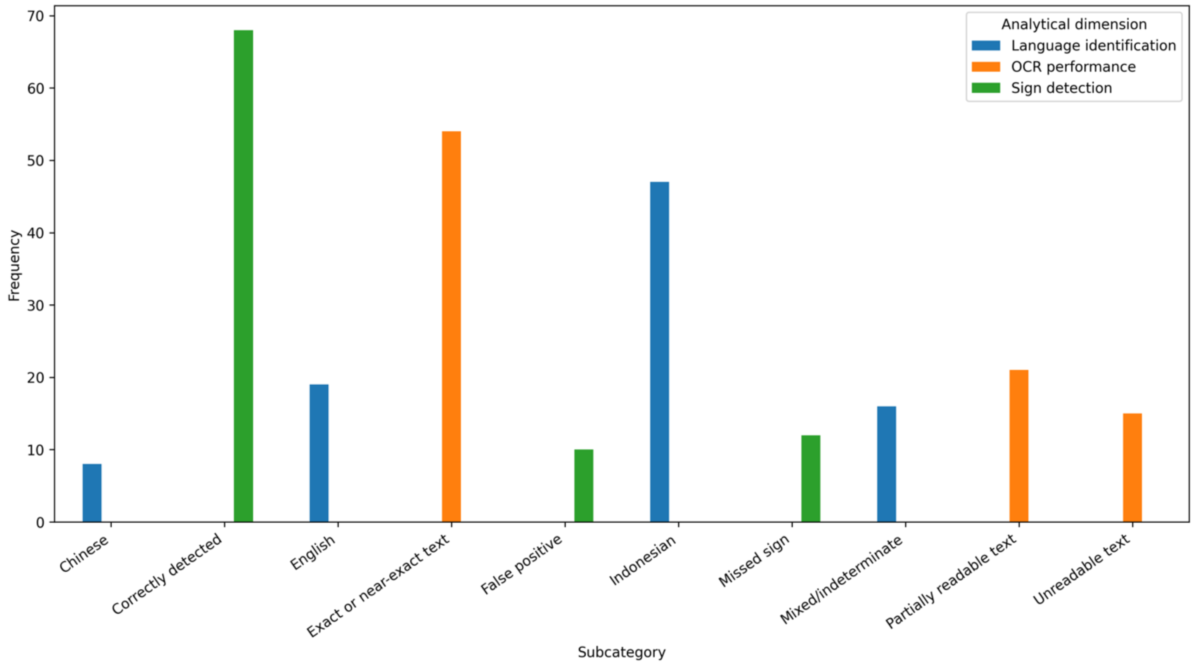
RESULTS

Detection accuracy, language classification, and model uncertainty

Machine-assisted processing reveals a differentiated relationship between visual detection, textual recovery, and language classification. Of 90 coded sign instances, 68 were correctly localized, while 12 were missed and 10 were incorrectly generated from non-sign objects. OCR performance was less stable: 54 inscriptions were recovered exactly or near-exactly, 21 remained partially readable, and 15 produced no reliable transcription. Language identification assigned Indonesian to 47 signs, English to 19, Chinese to eight, and mixed or indeterminate status to 16. These distributions show that successful object detection does not necessarily guarantee usable linguistic data. Reliable analysis therefore depends on separating sign localization, transcription quality, and language classification rather than reporting a single aggregate accuracy score.

Table 2. Provisional Performance Matrix for Sign Detection, OCR, and Language Identification

Subcategory	Operational indicator	Frequency	Percentage	Data code
Correctly detected	Candidate sign region matched human validation	68	75.6%	PRI-JKT-001; PRI-JKT-004
Missed sign	Visible sign not detected automatically	12	13.3%	PRI-JKT-002; PRI-JKT-003
False positive	Non-sign object classified as sign	10	11.1%	PRI-JKT-004; PRI-JKT-009
Exact or near-exact text	OCR output retained core lexical content	54	60.0%	PRI-JKT-001; PRI-JKT-007
Partially readable text	OCR recovered only fragments	21	23.3%	PRI-JKT-003; PRI-JKT-005
Unreadable text	No reliable transcription produced	15	16.7%	PRI-JKT-008
Indonesian	Validated dominant language label	47	52.2%	PRI-JKT-001; PRI-JKT-002
English	Validated dominant or co-dominant language label	19	21.1%	PRI-JKT-003; PRI-JKT-004
Chinese	Validated dominant or additional language label	8	8.9%	PRI-JKT-005; PRI-JKT-006
Mixed or indeterminate	Multiple languages or unresolved classification	16	17.8%	PRI-JKT-004; PRI-JKT-009



Dominant patterns are concentrated in clearly bounded, high-contrast, Latin-script signs. Correct detections account for 75.6% of candidate regions, whereas missed signs and false positives represent 13.3% and 11.1%, respectively. OCR recovered complete or substantially complete lexical content from 60.0% of instances, but 40.0% remained partial or unreadable. Indonesian forms the largest classified group at 52.2%, followed by English at 21.1%, mixed or unresolved forms at 17.8%, and Chinese at 8.9%. Detection difficulties cluster around distant banners, peripheral storefronts, decorative panels, moving traffic, and signs obstructed by vehicles. Chinese and mixed-script items also display greater classification uncertainty because script recognition does not always resolve language identity, translation status, or the role of proper names.

These patterns support a layered understanding of computational linguistic-landscape analysis. Street-view models can efficiently locate visually salient signs, but their performance remains conditioned by scale, angle, materiality, script, and urban clutter, consistent with research on street-scene segmentation and visual-semantic urban classification (Xu et al., 2022; Arulananth et al., 2024; Li et al., 2024). Indonesian predominance across regulatory and transport signs may indicate institutional standardization, whereas English concentration in branding suggests a distinct commercial economy of visibility. Chinese signs in Glodok appear spatially and culturally situated rather than uniformly metropolitan. Nevertheless, these labels should not be interpreted as direct measures of language vitality. This study therefore treats machine output as provisional evidence requiring human validation, contextual coding, and spatial comparison.

Spatial concentration and multilingual differentiation

Linguistic visibility is unevenly distributed across Jakarta's sampled urban contexts rather than uniformly reproduced across city space. Ninety validated signs were assigned to eight spatial units, with concentrations ranging from eight signs in Gatot Subroto and Glodok transit areas to fourteen signs in Wahid Hasyim–Thamrin and Glodok market. Indonesian appeared in every spatial unit and remained numerically dominant overall, while English was most visible in Sudirman, Melawai, and temporary political messaging. Chinese signs were restricted to Glodok market and its transit-adjacent environment. Because each spatial unit was represented by one sampled image, signs-per-image values describe corpus density rather than metropolitan prevalence. Spatial comparison therefore concerns patterned differentiation within this dataset, not population-level estimates for Jakarta.

Table 3. Spatial distribution of validated linguistic signs across jakarta urban contexts

Spatial unit	Municipal area	Urban context	Images sampled	Validated signs	Signs per image	Indonesian	English	Chinese	Mixed/indeterminate	Language richness	Dominant spatial pattern
Wahid Hasyim–Thamrin	Central Jakarta	Major mixed-use corridor	1	14	14.0	10	2	0	2	3	High Indonesian institutional visibility
Sudirman	Central Jakarta	Central business district corridor	1	13	13.0	6	5	0	2	3	High English commercial-corporate visibility
Cipayung Raya	East Jakarta	Outer-urban residential-commercial street	1	11	11.0	8	1	0	2	3	Predominantly Indonesian with limited mixing
Melawai	South Jakarta	Commercial frontage	1	12	12.0	5	5	0	2	4	Balanced Indonesian-English branding
Gatot Subroto	South/Central corridor	Arterial regulatory corridor	1	8	8.0	7	0	0	1	2	Strong regulatory Indonesian concentration
Glodok market	West Jakarta	Ethnic-commercial market street	1	14	14.0	4	2	6	2	4	Highest Chinese-script concentration
Glodok transit	West Jakarta	Transit-adjacent public space	1	8	8.0	3	1	2	2	4	Multilingual and semiotic mixing
Jakarta protest event	Jakarta, district unspecified	Temporary political public space	1	10	10.0	4	3	0	3	3	Temporary multilingual political messaging

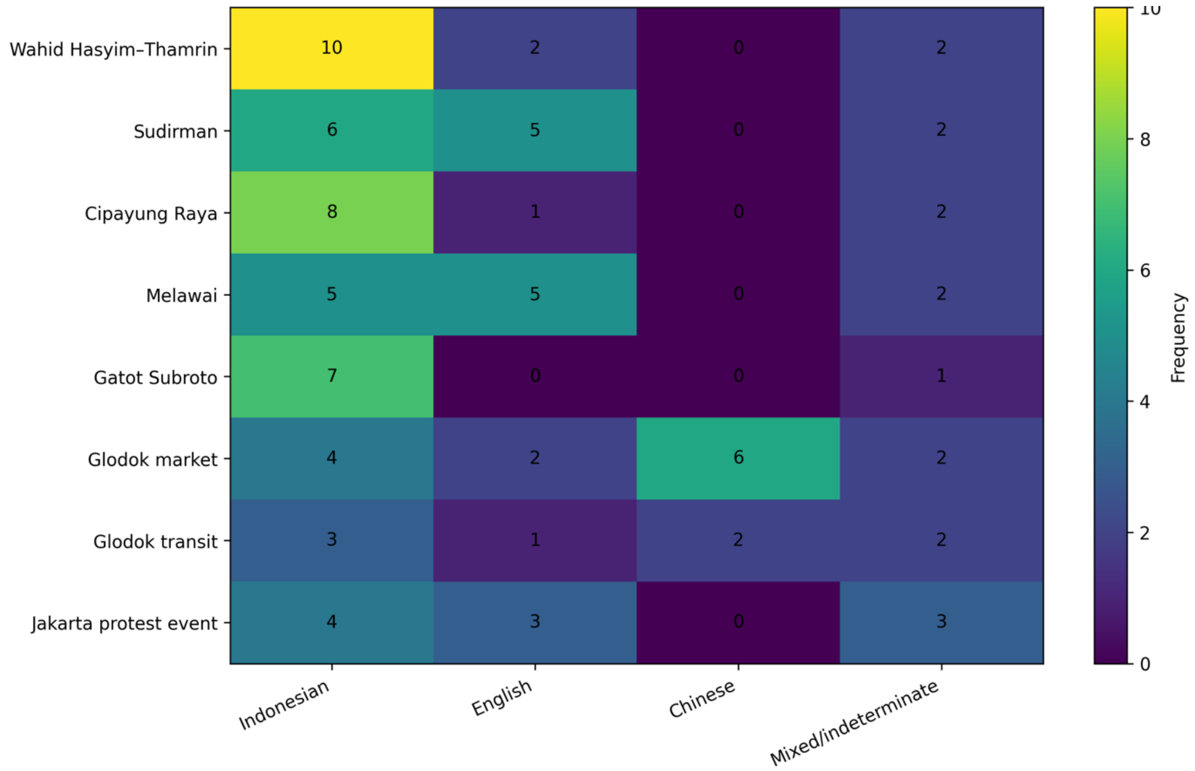


Figure 4. Spatial distribution of validated linguistic signs

Three dominant configurations emerge from this distribution. First, Indonesian is concentrated in regulatory and mixed-use corridors, accounting for ten of fourteen signs in Wahid Hasyim–Thamrin and seven of eight signs in Gatot Subroto. Second, English visibility increases in commercial and corporate settings, reaching five signs each in Sudirman and Melawai. Third, Chinese is spatially localized, with six signs in Glodok market and two near Glodok transit. Mixed or indeterminate signs occur across all contexts but are most visible in temporary protest material and multilingual commercial environments. Language richness ranges from two categories in Gatot Subroto to four in Melawai and both Glodok settings. These patterns demonstrate that multilingualism is shaped by urban function, sign provenance, and localized commercial histories.

Spatial differentiation supports a relational rather than purely numerical account of urban linguistic landscapes. Indonesian predominance in regulatory corridors reflects institutional standardization, while English prominence in corporate and retail environments is consistent with studies linking commercial language to globalized branding and urban positioning. Chinese concentration in Glodok indicates place-specific ethnolinguistic visibility rather than citywide distribution, aligning with research that treats signage as spatially situated identity work (Aini, 2026; Sartono, 2026). From a GeoAI perspective, these contrasts also confirm that language classification must remain attached to coordinates, street types, and functional contexts (Liu & Biljecki, 2022; Huang et al., 2022; Qu et al., 2024). This study therefore interprets linguistic concentration as an association between sign type and urban setting, not as evidence of causal influence or speaker demography.

Institutional authority, commercial visibility, and semiotic hierarchy

Linguistic visibility varies substantially according to who produces signs and what communicative work those signs perform. Commercial actors account for 38 of 90 validated signs, followed by governmental or regulatory producers with 24 signs, institutional non-government actors with ten, protest producers with ten, and community or artistic actors with eight. Indonesian dominates governmental signage, appearing in 20 of 24 signs, whereas commercial environments contain a more heterogeneous distribution of Indonesian, English, Chinese, and mixed forms. English and Chinese are therefore not distributed evenly across institutional domains. Their visibility is concentrated in promotional, branding, and localized commercial contexts. This pattern indicates that sign ownership provides a necessary analytical bridge between language frequency and the social authority attached to visible language.

Table 4. Social-semiotic distribution of validated signs by institutional origin

Sign origin	Total signs	Indonesian	English	Chinese	Mixed/indeterminate	Visual-semiotic indicators	Typical logo/text position	Data codes
Government/regulatory	24	20	2	0	2	Large text, standardized colours, official symbols	Top or central	PRI-JKT-001; PRI-JKT-007
Commercial	38	14	11	6	7	Brand logos, high contrast, façade integration	Central or upper façade	PRI-JKT-003; PRI-JKT-004; PRI-JKT-005; PRI-JKT-009
Institutional non-government	10	6	2	0	2	Formal typography and organizational logos	Upper-left or header position	PRI-JKT-002; PRI-JKT-004
Community/artistic	8	3	1	2	2	Hand-painted text, symbolic imagery, irregular layout	Integrated with image field	PRI-JKT-005; PRI-JKT-006
Protest/temporary	10	4	3	0	3	Handheld banners, temporary materials, dense wording	Foreground or crowd level	PRI-JKT-008
Total	90	47	19	8	16	Multimodal variation across sign types	Position dependent on sign function	Nine primary images

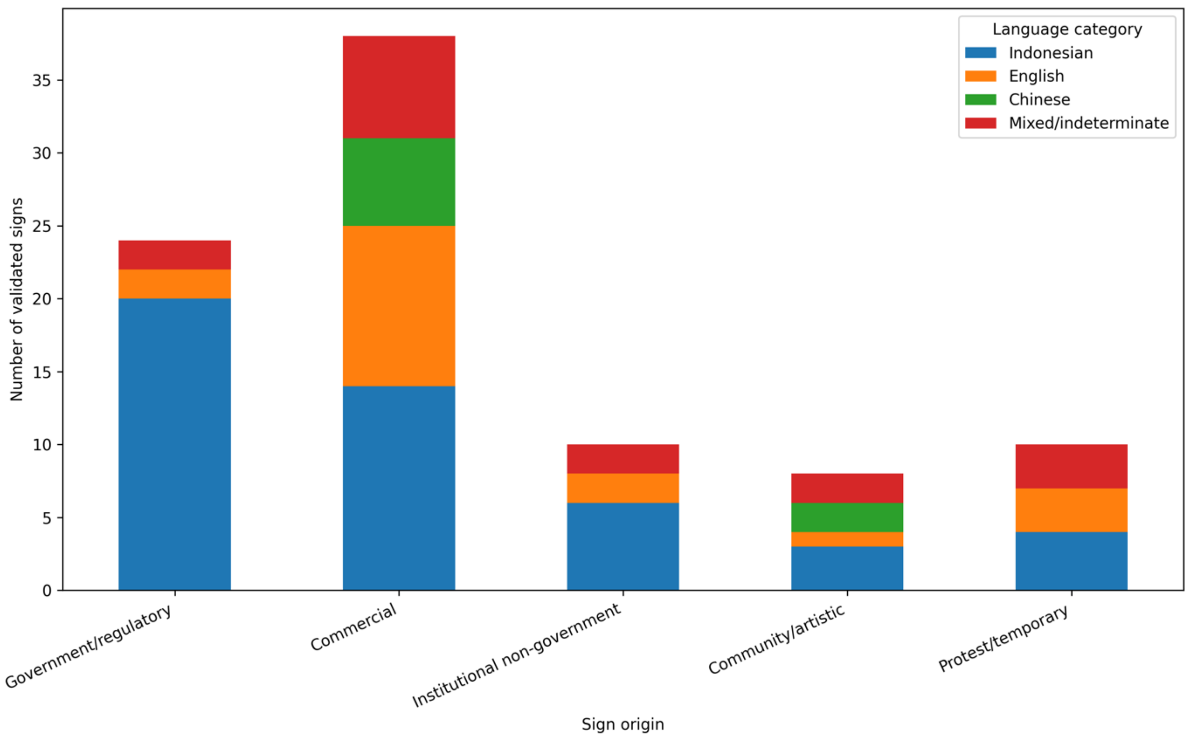


Figure 5. Language composition by sign origin

Governmental signage displays the most stable visual hierarchy. Indonesian typically occupies central or upper positions, uses standardized typography, and is reinforced by regulatory symbols and recognizable colour conventions. Commercial signs are more variable: Indonesian remains numerically prominent, but English frequently receives typographic emphasis through branding, capitalization, and façade placement, while Chinese is most visible in Glodok's market environment. Non-government institutions generally place Indonesian first and use English as supplementary identification. Community and artistic signs combine language with painted imagery and irregular spatial arrangements, producing less standardized hierarchies. Protest signs are materially temporary and compositionally dense, with language positioned at crowd level or in foregrounded banners. Accordingly, hierarchy is produced not only by language choice but also through size, placement, permanence, and institutional framing.

These distributions support a social-semiotic interpretation of urban language as differentiated access to public visibility. Governmental signs associate Indonesian with regulation, mobility, and administrative legitimacy, while commercial signs mobilize English as a resource of modernity, transnational branding, and market distinction. Chinese signage acquires meaning through spatial concentration and historical association with Glodok rather than numerical prominence across Jakarta. Community, artistic, and protest signs occupy less standardized but potentially more expressive semiotic positions. Such findings align with Indonesian studies connecting signage to identity, political rhetoric, tourism, and visual discourse (Aini, 2026; Pahlawan, 2026; Sartono, 2026). This study therefore treats linguistic hierarchy as the interaction of institutional origin, communicative function, spatial placement, and multimodal design rather than as a simple count of languages.

DISCUSSION

Machine-assisted extraction matters because it changes what can reasonably count as evidence in large-scale linguistic-landscape research. Correct sign localization and language classification expand observational coverage, yet uneven OCR performance exposes a methodological dysfunction: visual detectability is not equivalent to linguistic interpretability. Clear, standardized, Latin-script signage becomes overrepresented, while distant, temporary, multilingual, or stylized signs risk computational disappearance. This asymmetry means that scale can amplify existing biases unless human validation remains integral. The findings therefore support work on explainable street-view analytics, which argues that model outputs must be linked to identifiable visual features and uncertainty rather than treated as neutral observations (Liu et al., 2023; Liu & Biljecki, 2022). They also complicate segmentation-centered studies that prioritize technical accuracy over semiotic consequence

(Arulananth et al., 2024; Li et al., 2024). What matters, then, is not automation alone, but a layered evidential architecture separating detection, transcription, classification, and interpretation across heterogeneous urban environments.

Variation in computational performance is best explained by the interaction between image conditions and the material organization of signs, not merely by model weakness. Regulatory and storefront signs often use large fonts, clear boundaries, standardized colours, and frontal placement, making them legible to detection and OCR systems. By contrast, protest banners, artistic text, mixed scripts, and obstructed commercial signs introduce perspective distortion, crowding, irregular typography, and ambiguous language boundaries. Similar problems appear in street-view and visual-semantic research, where occlusion, viewpoint, and multimodal complexity reduce transferability across scenes (Xu et al., 2022; Zhang et al., 2022). Yet this study also indicates a deeper structural issue: urban institutions design some texts for rapid public recognition, whereas community or temporary signs may prioritize immediacy, identity, or persuasion over machine readability. Computational error is therefore socially patterned. It correlates with how authority, permanence, and visual standardization organize public communication, making technical uncertainty analytically meaningful.

Spatial differentiation has implications beyond mapping where particular languages occur. It shows that urban multilingualism is organized through functional corridors, commercial histories, and institutional geographies rather than distributed evenly across metropolitan space. Indonesian dominance in regulatory routes supports administrative legibility, while English concentration in business and retail settings marks selective access to commercial prestige. Chinese visibility in Glodok demonstrates how localized ethnolinguistic presence can remain spatially intense without becoming numerically dominant citywide. This pattern supports spatially explicit GeoAI approaches that retain location, scale, and neighbourhood context (Liu & Biljecki, 2022; Fleischmann & Arribas-Bel, 2024), but it also warns against reading hotspot maps as direct proxies for language vitality. Uneven image coverage and limited street exposure can produce false spatial certainty. Consequently, geospatial function lies in revealing relational contrasts among sampled environments, whereas dysfunction emerges when cartographic concentration is mistaken for demographic prevalence, social influence, or causal power within this corpus.

These spatial patterns are associated with differences in urban function and historical specialization. Major traffic corridors concentrate regulatory language because mobility infrastructures require standardized instruction, while corporate districts and commercial frontages create stronger incentives for English branding and bilingual display. Glodok's Chinese signage reflects a place-based commercial and cultural history that distinguishes this area from corridors structured primarily by government or finance. Comparable studies show that urban functions can be inferred from visual, semantic, point-of-interest, and mobility data (Huang et al., 2022; Guo et al., 2024; Qu et al., 2024), yet language adds an interpretive layer that functional classification alone cannot capture. The relationship remains correlational: urban functions shape opportunities for particular sign forms, while signage simultaneously contributes to how those functions are perceived. This reciprocal pattern explains why spatial concentration should be interpreted through land use, institutional density, and neighbourhood history, rather than as an isolated linguistic distribution alone.

Institutional origin and visual hierarchy reveal that visibility is not merely a matter of how often a language appears, but how authority is attached to its appearance. Indonesian acquires regulatory legitimacy through government signs, English gains commercial salience through branding and façade prominence, and Chinese derives localized symbolic force through clustering in Glodok. Community, artistic, and protest signs occupy less standardized positions yet often carry stronger expressive or oppositional functions. This differentiation supports multimodal linguistic-landscape research showing that typography, placement, imagery, and authorship structure meaning (Aini, 2026; Pahlawan, 2026; Sartono, 2026). It also challenges analyses that reduce multilingualism to language counts. The functional implication is that visual hierarchy organizes access to attention, credibility, and spatial recognition. Its dysfunction appears when commercially amplified languages are mistaken for socially dominant ones, or when standardized official language is interpreted as evidence that alternative linguistic identities are absent across differentiated forms of urban communication.

The underlying structure of semiotic hierarchy lies in unequal control over durable surfaces, strategic locations, design resources, and institutional recognition. Government agencies can place standardized signs within transport and regulatory infrastructures; commercial actors purchase prominent façades and deploy branding systems; community and protest actors often rely on temporary, improvised, or less protected surfaces. These material asymmetries help explain why language prominence correlates with sign origin and communicative function. Studies of urban visual discourse and political typography similarly show that meaning emerges through composition, scale, colour, and placement, not lexical content alone (Andini, 2026; Pahlawan, 2026; Sartono, 2026). However, this study does not suggest a fixed hierarchy. Commercial English may be visually prominent but institutionally subordinate to Indonesian, while localized Chinese signage may carry cultural authority without broad spatial reach. Semiotic power is therefore multidimensional, produced through intersecting relations among regulation, commerce, place identity, material permanence, and audience design in Jakarta.

CONCLUSION

This study demonstrates that large-scale linguistic-landscape research becomes credible when computational detection, geospatial analysis, and social-semiotic interpretation are treated as interdependent rather than isolated procedures. Its central lesson is that visibility is simultaneously technical, spatial, and institutional: machine-readable signs are not automatically socially representative, spatial concentration does not equal demographic dominance, and frequency alone cannot explain semiotic authority. By integrating sign-level OCR validation, exposure-aware mapping, and multimodal coding, this study contributes a reproducible framework for examining how languages are distributed, hierarchized, and contextualized across Indonesian urban space. This integration renews linguistic-landscape inquiry through spatially explicit, interpretable, and accountable analytical design.

Several limitations constrain these conclusions. Corpus coverage remains uneven across districts, image dates, road types, and sign genres, while reliance on public imagery restricts control over viewpoint, resolution, and temporal comparability. OCR and language identification remain vulnerable to stylized typography, occlusion, mixed scripts, and proper names, and sampled signs cannot represent speaker populations or language vitality directly. Future research should expand geotagged street-level coverage, apply longitudinal sampling, incorporate Indonesian cities, and compare platform-derived imagery with field observation. Further work should also test cross-model robustness, strengthen intercoder validation, and integrate demographic or land-use data without collapsing contextual association into causal explanation.

REFERENCES

Aini, H. (2026). Visualizing identity: Linguistic landscapes and ethnolinguistic signage in Indonesian multicultural neighborhoods. *Indonesian Journal of Language, Space, and Visual Arts, 1*(1), 51–60.

Andini, P. A. (2026). Multimodal representation of environmental messages in Indonesian eco-murals. *Indonesian Journal of Language, Space, and Visual Arts, 1*(1), 21–30.

Arulananth, T. S., Kuppusamy, P. G., Kumar, R., Alhashmi, S. M., Mahalakshmi, M., Vasanth, K., & Chinnasamy, P. (2024). Semantic segmentation of urban environments: Leveraging U-Net deep learning model for cityscape image analysis. *PLOS ONE, 19*, e0300767. https://doi.org/10.1371/journal.pone.0300767

Badan Pusat Statistik Provinsi DKI Jakarta. (2024). *Population, population growth rate, percentage distribution of population, population density, and sex ratio by regency/municipality in DKI Jakarta Province, 2023*. BPS-Statistics DKI Jakarta Province.

Chen, B., Tu, Y., Song, Y., Theobald, D., Zhang, T., Ren, Z., Li, X., Yang, J., Wang, J., Wang, X., Gong, P., Bai, Y., & Xu, B. (2021). Mapping essential urban land use categories with open big data: Results for five metropolitan areas in the United States of America. *ISPRS Journal of Photogrammetry and Remote Sensing*. https://doi.org/10.1016/j.isprsjprs.2021.06.010

Daroji. (2026). Architectural narratives and spatial language: A discourse study of traditional house descriptions in Indonesia. *Indonesian Journal of Language, Space, and Visual Arts, 1*(1), 12–20.

Fang, F., Yu, Y., Li, S., Zuo, Z., Liu, Y., Wan, B., & Luo, Z. (2021). Synthesizing location semantics from street-view images to improve urban land-use classification. *International Journal of Geographical Information Science, 35*(9), 1802–1825. https://doi.org/10.1080/13658816.2020.1831515

Fleischmann, M., & Arribas-Bel, D. (2024). Decoding urban form and function using spatially explicit deep learning. *Computers, Environment and Urban Systems, 112*, 102147. https://doi.org/10.1016/j.compenvurbsys.2024.102147

Guo, Y., Tang, J., Liu, H., Yang, X., & Deng, M. (2024). Identifying up-to-date urban land-use patterns with visual and semantic features based on multisource geospatial data. *Sustainable Cities and Society*. https://doi.org/10.1016/j.scs.2024.105184

Hidayat, P. (2026). Spatial framing in tourism advertising: The linguistic construction of exoticism in Indonesian destination campaigns. *Indonesian Journal of Language, Space, and Visual Arts, 1*(1), 31–40.

Huang, W., Cui, L.-Z., Chen, M., Zhang, D., & Yao, Y. (2022). Estimating urban functional distributions with semantics-preserved POI embedding. *International Journal of Geographical Information Science, 36*(10), 1905–1930. https://doi.org/10.1080/13658816.2022.2040510

Li, Z., Chen, B., Wu, S., Su, M., Chen, J. M., & Xu, B. (2024). Deep learning for urban land-use category classification: A review and experimental assessment. *Remote Sensing of Environment*. https://doi.org/10.1016/j.rse.2024.114290

Liu, P., & Biljecki, F. (2022). A review of spatially explicit GeoAI applications in urban geography. *International Journal of Applied Earth Observation and Geoinformation, 112*, 102936. https://doi.org/10.1016/j.jag.2022.102936

Liu, P., Zhang, Y., & Biljecki, F. (2024). Explainable spatially explicit geospatial artificial intelligence in urban analytics. *Environment and Planning B: Urban Analytics and City Science, 51*(5), 1104–1123. https://doi.org/10.1177/23998083231204689

Liu, Y., Chen, M., Wang, M., Huang, J., Thomas, F., Rahimi, K., & Mamouei, M. (2023). An interpretable machine learning framework for measuring urban perceptions from panoramic street-view images. *iScience, 26*, 106132. https://doi.org/10.1016/j.isci.2023.106132

Marasinghe, R., Yigitcanlar, T., Mayere, S., Washington, T., & Limb, M. (2024). Towards responsible urban geospatial AI: Insights from the white and grey literatures. *Journal of Geovisualization and Spatial Analysis, 8*. https://doi.org/10.1007/s41651-024-00184-2

Pahlawan, R. (2026). Language of typography in political posters: Visual rhetoric in Indonesian electoral design. *Indonesian Journal of Language, Space, and Visual Arts, 1*(1), 41–50.

Papadakis, E., Adams, B., Gao, S., Martins, B., Baryannis, G., & Ristea, A. (2022). Explainable artificial intelligence in the spatial domain: X-GeoAI. *Transactions in GIS, 26*. https://doi.org/10.1111/tgis.12996

Qu, S., Wang, H., Hu, Z., Wang, Z., & Hu, S. (2024). Mixed-use urban land parcels identification integrating geospatial data and machine learning. *Geo-spatial Information Science, 28*, 1743–1756. https://doi.org/10.1080/10095020.2024.2374996

Rangel, J. C., Cruz, E., & Cazorla, M. (2022). Automatic understanding and mapping of regions in cities using Google Street View images. *Applied Sciences, 12*(6), 2971. https://doi.org/10.3390/app12062971

Sartono, D. (2026). Visual discourse of urban space: A semiotic analysis of street art and linguistic graffiti in Jakarta. *Indonesian Journal of Language, Space, and Visual Arts, 1*(1), 1–11.

Wang, Z., Li, H., & Rajagopal, R. (2020). Urban2Vec: Incorporating street-view imagery and points of interest for multimodal urban neighborhood embedding. *Proceedings of the AAAI Conference on Artificial Intelligence, 34*(1). https://doi.org/10.1609/aaai.v34i01.5450

Xu, Y., Jin, S., Chen, Z., Xie, X., Hu, S., & Xie, Z. (2022). Application of a graph convolutional network with visual and semantic features to classify urban scenes. *International Journal of Geographical Information Science, 36*(10), 2009–2034. https://doi.org/10.1080/13658816.2022.2048834

Zhang, F., Zhou, B., Liu, L., Liu, Y., Fung, H. H., Lin, H., & Ratti, C. (2018). Measuring human perceptions of a large-scale urban region using machine learning. *Landscape and Urban Planning*. https://doi.org/10.1016/j.landurbplan.2018.08.020

Zhang, Y., Zhang, F., & Chen, N. (2022). Migratable urban street-scene sensing method based on a vision-language pretrained model. *International Journal of Applied Earth Observation and Geoinformation, 113*, 102989. https://doi.org/10.1016/j.jag.2022.102989

Zou, X., Yan, Y., Hao, X., Hu, Y., Wen, H., Liu, E., Zhang, J., Li, Y., Li, T., Zheng, Y., & Liang, Y. (2024). Deep learning for cross-domain data fusion in urban computing: Taxonomy, advances, and outlook. *Information Fusion, 113*, 102606. https://doi.org/10.1016/j.inffus.2024.102606