Journal of Supply Chain and Entrepeneurship
2025, VOL. 1, NO. 2, 45-52
https://doi.org/10.64268/josce.v1i2.35

Harnessing NLP and Big Data to Solve Linguistic Challenges in Indonesian Humanoid Robots: Pathways to Innovation and Entrepreneurship

Syaifullah, Wenny Noorahim
Universitas Lambung Mangkurat

***Corresponding author: syaifullah_fkip@ulm.ac.id**

ABSTRACT

Aim: Indonesian, as a national language, contains intricate linguistic features such as agglutinative morphology, idioms, and numerous dialectal variations. These characteristics present significant challenges in developing humanoid robots capable of natural interaction through Natural Language Processing (NLP). This study aims to address these linguistic complexities while exploring the entrepreneurial potential of localized NLP applications in Indonesia.
Methods: The research employs a qualitative literature review method, focusing on existing studies related to Indonesian NLP datasets, transformer-based language models, and speech technologies. Key sources include IndoNLI for inference, IndoSentiment for sentiment analysis, and case studies of humanoid robots like Lumen. The analysis also includes approaches utilizing Big Data, multi-pass decoders, and contextual language modeling to optimize performance in Indonesian linguistic settings.
Findings: Findings indicate that the successful development of Indonesian-speaking humanoid robots relies on context-aware NLP models trained on representative, culturally relevant datasets. Integrating multimodal systems and Big Data enables enhanced comprehension of idiomatic, regional, and informal expressions. The research also reveals that NLP-based innovations can be commercialized through AI-powered assistants, educational bots, and digital customer service, opening new opportunities for tech-driven entrepreneurship.
Significance: This study contributes to both technological advancement and business innovation by linking linguistic AI research with entrepreneurial applications. It underscores the importance of building a robust local data ecosystem and designing language models that reflect Indonesia's linguistic diversity. These insights are vital not only for improving human-robot interaction but also for fostering sustainable digital entrepreneurship within emerging markets like Indonesia.

ARTICLE HISTORY
Submitted: May 05, 2025
Accepted: June 07, 2025
Published: July 05, 2025

KEYWORD
Digital Entrepreneurship
Innovation Ecosystem
Language Technology
Commercialization
NLP-Based Startups
Supply Chain of AI Products

INTRODUCTION

Indonesian, as the national and unifying language for over 270 million people, presents a unique linguistic structure that differs significantly from widely adopted global languages. Its agglutinative nature, extensive use of affixes, and flexible syntactic structures create complexities in computational processing (Ernawati et al., 2023). Unlike English, where sentence elements often follow a fixed structure, Indonesian allows variations in subject-predicate-object positioning. This flexibility introduces challenges for Natural Language Processing (NLP) systems in recognizing intent, meaning, and context. Moreover, idiomatic expressions and colloquialism in daily conversation further complicate parsing and machine understanding. These linguistic intricacies are particularly problematic in the development of humanoid robots intended to interact naturally with Indonesian speakers. Robots must not only interpret literal meanings but also navigate figurative language and cultural norms. This need underlines the importance of localizing AI systems in language-sensitive regions like Indonesia.

Another dimension of complexity arises from Indonesia's rich dialectal diversity, where language usage varies by region, ethnicity, and social context. A single word may carry different meanings across provinces, leading to ambiguity in speech recognition and semantic modeling (Diao & Hu, 2021 and Zhao et al. 2021). For humanoid robots to function effectively in Indonesia, they must be equipped with NLP capabilities that are sensitive to these variations. Unfortunately, most NLP advancements globally remain

centered on English and Mandarin, leaving Indonesian underrepresented in model development and data resources (Cahyawijaya et al. 2022). As a result, there is a growing gap between linguistic diversity and available technological solutions in AI. Recent initiatives such as IndoNLI and IndoSentiment demonstrate how local datasets can bridge this gap by tailoring training materials to Indonesian linguistic characteristics (Mahendra et al. 2021 and Budiharto. 2020). These datasets include logical inference, idiomatic expression recognition, and sentiment analysis, all aligned with real-world conversational patterns. However, these resources remain limited in scale and require continual updates to keep pace with evolving language use.

Big Data has emerged as a valuable tool to enhance NLP model training, especially in linguistically complex environments. By leveraging data from online news, social media, and audio transcripts, language models gain exposure to diverse language forms and usage contexts (Jiono, 2020 and Chen et al. 2024). Yet, this process introduces new problems such as data imbalance, noise, and ethical concerns surrounding user privacy. NLP research must therefore consider the quality, diversity, and integrity of the data being used. The Lumen robot project, which incorporated local NLP modules and cloud-based voice processing, demonstrated notable success in public interaction scenarios, reaching over 96% recognition accuracy under real-world conditions (Sya & Prihatmanto, 2015). However, challenges remain in handling informal language, short utterances, and regional dialects. Heinrich and Wermter (2011) proposed the use of multi-pass decoder systems to enhance speech recognition accuracy, particularly in noisy environments. These technological innovations illustrate the potential for combining data-driven models with algorithmic advancements to strengthen human-robot communication in Indonesian.

While much of the existing research focuses on technical optimization, relatively few studies explore the entrepreneurial potential of Indonesian NLP technologies. The commercial applications of NLP-driven humanoid robots include customer service automation, educational robots, and smart public service kiosks (Paramesha et al. 2024 and Younis et al. 2023). The development of localized language technologies opens new markets for AI startups, especially those focusing on regional services and user interfaces. For example, ventures like Bahaso and Prosa. ai are already creating tools tailored for Indonesian users, emphasizing the feasibility of NLP commercialization (Luckyardi, Karin, et al. 2024 and Luckyardi, Munawaroh, et al. 2024). Moreover, integrating speech and emotion recognition into social robots could support industries like tourism, health tech, and inclusive education (Ragno et al. 2023 and Schiavo et al. 2024). These innovations contribute not only to technological advancement but also to building innovation ecosystems in emerging economies. By translating linguistic research into marketable products, Indonesia has the opportunity to become a regional leader in AI entrepreneurship. However, this transition requires strategic investments in research, policy support, and scalable infrastructure.

Existing literature also reveals a lack of multimodal and context-aware systems tailored to Indonesian cultural communication. For instance, many current models fail to account for nonverbal cues, speech tone, or turn-taking conventions commonly practiced in Indonesian social interactions. A study by Gruetzemacher & Paradice. (2022) and Lin et al. (2025) found that culturally grounded NLP models significantly outperformed general-purpose ones in tasks involving politeness detection and intent classification. Additionally, contextual understanding is key to resolving ambiguity in utterances, especially in informal or emotionally charged interactions. Transformer-based models such as IndoBERT and ID-GPT have shown promise in adapting context-aware learning to local languages (Li, 2024). Yet, their deployment in embodied AI systems like humanoid robots remains limited and experimental. This gap highlights the importance of designing NLP systems that not only understand language semantically but also socially and pragmatically. Closing this gap will accelerate the development of socially intelligent robots, capable of fostering trust and engagement across diverse user groups.

In summary, this article addresses a dual challenge in the field of NLP and AI for the Indonesian context: the linguistic barriers that hinder NLP performance, and the untapped entrepreneurial pathways

for commercializing localized AI technologies. By conducting a comprehensive literature-based analysis, this study aims to map key problem areas and offer scalable solutions. It integrates perspectives from computational linguistics, robotics, and digital entrepreneurship to construct a holistic view of the Indonesian AI landscape. The novelty of this research lies in its interdisciplinary approach, blending language technology with business model innovation. Moreover, it emphasizes the importance of local knowledge, ethical data practices, and context-aware system design. By identifying technical and commercial enablers, this paper seeks to contribute both to academic discourse and practical roadmaps for AI development in Southeast Asia. In doing so, it lays a foundation for inclusive, adaptive, and economically viable NLP solutions embedded in humanoid robot technologies for Indonesian users.

METHOD

Research Design

This study applies a descriptive qualitative design using a library research approach that is both analytical and exploratory. The purpose of this design is to deeply understand the intersection of linguistic complexity, NLP, and entrepreneurial innovation in the development of Indonesian-speaking humanoid robots. This approach allows for the evaluation of existing theories, datasets, case studies, and models relevant to Natural Language Processing and Big Data technologies. Qualitative thematic analysis is employed to extract patterns and connections from scholarly works in computational linguistics and AI applications. According to Moleong (2018), qualitative methods are appropriate when the research seeks to explore deep meanings and systemic relationships rather than statistical generalizations. The study examines how linguistic phenomena—such as affixation, dialectal diversity, and semantic ambiguity—impact the performance of NLP systems in humanoid robots. Furthermore, it also explores how technological advancements in language processing may lead to innovation ecosystems and digital entrepreneurship opportunities. The combination of qualitative synthesis and strategic entrepreneurship framing makes this research suitable for understanding both technological and commercial implications.

Population and Sample

As a literature-based qualitative study, the research does not involve human or experimental populations. Instead, the "population" refers to published academic materials relevant to Indonesian NLP, Big Data, robotic applications, and AI-based entrepreneurship. These sources include journal articles indexed in Scopus, SINTA, and Google Scholar, as well as conference proceedings in computational linguistics and intelligent systems. The "sample" comprises selected studies that meet thematic relevance, recency, and scientific credibility criteria. Sampling was conducted purposively to ensure that the selected materials adequately reflect both technological complexity and the emerging entrepreneurial dimensions of AI integration. Special attention was given to works discussing localized datasets such as IndoNLI (Mahendra et al., 2021), IndoSentiment (Budiharto, 2020), and NusaCrowd (Cahyawijaya et al., 2022), which are essential for training language models in Indonesian contexts. The inclusion of entrepreneurial case studies—such as the commercial deployment of humanoid robots like Lumen (Sya & Prihatmanto, 2015)—further enriches the sample diversity. Thus, the sampled literature is carefully curated to provide both scientific depth and practical business relevance.

Data Collection

Data were collected through a systematic and comprehensive review of secondary sources using keyword-based search strategies. The search employed combinations such as "NLP Indonesian," "Big Data linguistics," "humanoid robots Indonesia," "digital entrepreneurship," and "AI-based startups." Sources were retrieved from academic journal databases, open-access repositories, and university libraries. The inclusion criteria focused on peer-reviewed articles, official documentation on dataset development, and empirical case studies related to NLP implementation in Indonesian-speaking robots. Key data included technological frameworks, accuracy evaluations, training datasets, and commercial

applications in AI-driven environments. The method of data extraction followed the "read-record-organize" technique to ensure fidelity and traceability. In addition to mainstream academic content, the study also considered innovation reports and industry white papers that discuss AI deployment in the Indonesian digital economy. Supporting references include work on multi-pass decoders (Heinrich & Wermter, 2011), informal language corpora, and data sources from social platforms (Jiono, 2020). Altogether, the data collection process ensured the inclusion of both academic rigor and applied innovation perspectives.

Measurement and Variables

In qualitative research, variables are interpreted as thematic constructs rather than measurable quantities. The key constructs examined in this study include linguistic complexity, NLP system performance, dataset representativeness, commercial viability, and innovation ecosystem potential. Linguistic complexity is operationalized through features such as morphological agglutination, idiomatic variation, and dialect diversity (Ernawati et al. 2023 and Fitri, 2024). NLP system performance is evaluated based on reported metrics in prior studies such as inference accuracy, sentiment analysis precision, and contextual understanding in humanoid interactions (Mahendra et al. 2021 and Budiharto, 2020). Dataset representativeness considers the coverage, diversity, and cultural relevance of Indonesian corpora like IndoNLI and IndoSentiment. Meanwhile, entrepreneurial variables are framed through indicators such as the presence of market-driven NLP applications, scalability potential of AI solutions, and the integration of local technology in supply chain models. These constructs are coded and interpreted to map the relationship between language technology development and its commercialization pathways. This multidimensional approach allows for a balanced understanding of both the scientific and entrepreneurial aspects of the research topic.

Data Analysis

The analysis was conducted through three key stages: data reduction, thematic display, and conclusion verification. In the reduction phase, irrelevant or outdated literature was filtered out, retaining only content aligned with the dual focus on NLP technologies and entrepreneurial integration. Selected sources were then organized into thematic clusters such as Indonesian language structure, NLP model architecture, Big Data processing, human-robot interaction, and innovation strategies. These themes were visually mapped to identify intersections between technical challenges and business opportunities (Ferasso et al. 2023 and Hussain. 2023). In the display phase, the data were categorized under focus areas like transformer-based NLP models, multimodal systems, commercial robot deployment, and AI product supply chains. Triangulation was used to verify conclusions by comparing claims across multiple sources, ensuring internal consistency and conceptual alignment. Validation was strengthened by checking whether insights from NLP-focused studies could be corroborated with examples from real-world commercialization efforts. The integration of literature from both academia and industry enhances the trustworthiness of findings and aligns them with international research standards. Ultimately, the data analysis supports the conclusion that linguistic AI solutions, when localized and strategically deployed, hold substantial potential for innovation and entrepreneurship in Indonesia.

RESULTS AND DISCUSSION

Results

The findings from this study highlight five major linguistic and technical challenges encountered in the development of Indonesian-speaking humanoid robots, especially when incorporating NLP and Big Data technologies. These challenges include morphological and idiomatic complexity, regional dialect variation, acoustic limitations in voice interaction, contextual ambiguity in meaning, and data imbalance between formal and informal communication. The literature confirms that Indonesian's agglutinative structure and idiomatic richness create tokenization and semantic processing issues (Ernawati et al., 2023 and Nasution & Onan. 2024). Regional dialects introduce further variability that current NLP

models struggle to interpret accurately (Fitri, 2024). Moreover, real-world interactions are complicated by noisy environments and short or informal speech, which degrade the performance of speech recognition modules. Ambiguous words that rely on contextual cues are often misinterpreted by systems that lack deep contextual training. Lastly, the dominance of formal language corpora in current NLP models results in insufficient exposure to everyday Indonesian communication styles. These interrelated findings emphasize the urgent need for context-sensitive NLP strategies and localized dataset enhancement.

Table 1. Challenges and Solutions for Indonesian NLP in Humanoid Robots

Aspect	Challenges	Solutions
Morphology & Idioms	Agglutinative structure, idioms, syntactic variations	Context-based NLP using BERT or GPT models
Local Dialects	Regional accents and terminology	Local datasets combined with model fine-tuning
Voice Interaction	Noisy environments, short or fragmented sentences	Multi-pass decoder for enhanced speech recognition
Meaning Variation	Ambiguous words and context-dependent expressions	Contextual interpretation via Big Data-based dialogue
Data Imbalance	Over-reliance on formal language data; lack of informal expressions	Integration of informal sources like social media & podcasts

Table 1 outlines five core challenges faced in implementing Natural Language Processing (NLP) for Indonesian-speaking humanoid robots and provides corresponding solutions based on the literature review. The first challenge relates to the agglutinative nature of Indonesian, where words change significantly with affixes and idioms. This requires advanced NLP models such as BERT or GPT, which can understand context more holistically. The second challenge is the presence of numerous local dialects, which necessitates the collection and use of localized datasets along with fine-tuning of existing models. The third challenge addresses the technical issue of recognizing voice in noisy public environments or in short utterances, which can be mitigated by applying a multi-pass decoder approach. The fourth challenge is ambiguity in meaning depending on the context, where the solution lies in training systems with context-rich, conversational Big Data. Lastly, the imbalance in language data—dominated by formal structures—can be addressed by enriching training datasets with informal communication forms such as those from social media and podcasts. Together, these strategies provide a comprehensive roadmap for developing NLP systems that are linguistically, culturally, and commercially viable in the Indonesian context.

Discussion

Addressing linguistic complexity requires advanced NLP models that are capable of handling Indonesian morphology, affixes, and syntactic variations. Transformer-based architectures like BERT and GPT offer promising solutions by learning contextual relationships across sentences. When trained on Indonesian-specific corpora such as IndoNLI, these models improve understanding of complex sentence structures and idioms (Mahendra et al. 2021 and Cucchiarini et al., 2022). The integration of such models within humanoid robots provides more natural interaction patterns and enhances semantic accuracy. However, they still require ongoing refinement through fine-tuning with culturally relevant datasets that reflect local expressions and informal tones.

The lack of representation for regional dialects poses a significant limitation to the scalability of NLP-powered humanoid robots across Indonesia. To overcome this, researchers have emphasized the importance of collecting geographically distributed datasets, including diverse speech styles and pronunciation variations. IndoSentiment has contributed to this goal by analyzing emotion and sentiment in informal language, yielding improved performance in interpreting user intent (Budiharto, 2020). Speech recognition modules must also be supported by robust decoding systems capable of processing local accents in dynamic, non-standard environments. Without such localization, the reach of commercial AI products will remain confined to formal or urban language contexts.

The technological solutions discussed in this study, such as multi-pass decoders and Big Data-based contextual learning, address real-time interaction constraints. Multi-pass decoders, as introduced by Heinrich and Wermter (2011), enable humanoid robots to reduce speech recognition errors in noisy public spaces such as exhibitions or schools. Meanwhile, Big Data aggregation from social media, podcasts, and online discussions allows language models to better interpret slang, humor, and emotional tones. These advancements can significantly enhance the effectiveness of humanoid robots deployed in customer service, education, and public engagement roles.

From an entrepreneurial standpoint, these technologies open new markets and business models in the digital economy. The development of localized NLP systems encourages the emergence of AI startups focused on Indonesian voice assistants, educational bots, and automated service platforms. By commercializing language datasets and interaction modules, entrepreneurs contribute to building a domestic AI supply chain that fosters innovation. Moreover, these business opportunities help preserve linguistic diversity while increasing inclusivity in AI-driven products tailored for local users.

Implications

The implications of this research span both academic and practical domains. On the academic side, it emphasizes the necessity of culturally grounded NLP systems to overcome linguistic barriers in emerging economies. By using Indonesian as a case study, this work contributes to the broader discourse on AI localization and the ethical deployment of language technologies. Practically, the results highlight the commercial viability of NLP innovations that align with local language behaviors, social norms, and digital consumption patterns. Policymakers and entrepreneurs alike can leverage these findings to promote technology startups, especially in underserved regions where linguistic diversity is both an asset and a barrier. Educational institutions may also adopt humanoid robots as language-learning tools, thereby integrating technology and pedagogy. Furthermore, this research supports the development of inclusive AI strategies within national digital transformation agendas. Overall, the alignment of NLP solutions with entrepreneurship fosters both technological sovereignty and socio-economic impact in the Indonesian context.

Limitations and Future Research

Despite the insights presented, this study is limited by its reliance on secondary data and theoretical modeling. It does not include empirical field validation of NLP systems in real-world robot-human interactions, which could offer a more nuanced understanding of implementation challenges. Additionally, many referenced datasets such as IndoNLI and IndoSentiment, while valuable, are still evolving and may not fully capture the breadth of Indonesia's linguistic diversity. Future research should focus on empirical testing of NLP modules in multilingual field trials involving users from various socio-demographic groups. Longitudinal studies could also examine how AI-driven language tools influence user behavior, perception, and adoption over time. Another critical area for future work is the monetization model for language datasets and AI-powered products within Indonesia's entrepreneurial ecosystem. Studies integrating human-computer interaction (HCI), digital business frameworks, and local cultural analysis are needed to build sustainable, inclusive, and commercially scalable language technology solutions.

CONCLUSION

This study highlights that the development of Indonesian-speaking humanoid robots requires not only advanced linguistic technology but also a strategic understanding of Indonesia's diverse language landscape and entrepreneurial potential. By examining the linguistic complexity of Indonesian—ranging from agglutinative morphology and idioms to regional dialects and informal variations—this research demonstrates the critical need for localized NLP models supported by representative datasets such as IndoNLI and IndoSentiment. The integration of Big Data and transformer-based architectures like BERT and GPT enables systems to recognize meaning beyond surface-level grammar, while technologies such as multi-pass decoders improve interaction quality in real-world conditions.

Beyond its technological focus, this study identifies significant opportunities for commercializing NLP solutions through the creation of AI-powered services tailored for Indonesian users. These include educational bots, digital assistants, and customer service platforms that can foster innovation and enhance the inclusiveness of AI deployment in the country. The alignment of language technologies with entrepreneurial ecosystems not only addresses linguistic barriers but also stimulates the domestic supply chain of AI products, contributing to the broader goals of digital transformation.

In sum, the convergence of NLP, Big Data, and entrepreneurship presents a promising pathway for Indonesia to build context-aware humanoid robots that are both socially relevant and economically viable. Future advancements must continue to integrate linguistic precision with scalable business models to ensure that technological innovation benefits diverse user groups and supports sustainable growth within the national and regional AI economy.

AUTHOR CONTRIBUTIONS STATEMENT

Syaifullah was responsible for the conceptualization of the research, development of the research framework, and the drafting of the initial manuscript. He also led the analysis of the linguistic and technological dimensions of NLP and Big Data in humanoid robots.

Wenny Noorahim contributed to the refinement of the research methodology, literature synthesis, and the integration of entrepreneurial perspectives into the study. She also played a key role in reviewing and editing the manuscript for academic coherence and publication readiness.

Both authors jointly validated the findings, approved the final version of the manuscript, and take full responsibility for the content presented in this study.

REFERENCES

Budiharto, W. (2020). Deep Learning-Based Question Answering System For Intelligent Humanoid Robot. *Journal Of Big Data, 7*(1). Https://Doi.Org/10.1186/S40537-020-00341-6

Cahyawijaya, S., Aji, A. F., Lovenia, H., Winata, G. I., Wilie, B., Mahendra, R., Koto, F., Moeljadi, D. Vincentio, K., Romadhony, A., & Purwarianti, A. (2022). Nusacrowd: A Call For Open And Reproducible NLP Research In Indonesian Languages. *Arxiv*. Https://Doi.Org/10.48550/Arxiv.2205.15960

Ernawati, I. A., Brawijaya, K. S., Aini, F., & Nurhayati, E. (2023). Perkembangan Ragam Bahasa Dalam Komunikasi Mahasiswa Di Lingkungan Kampus UPN "Veteran" Jawa Timur. *Jurnal Pengabdian West Science, 2*(6), 406–420. Https://Doi.Org/10.58812/Jpws.V2i6.388

Heinrich, S., & Wermter, S. (2011). Towards Robust Speech Recognition For Human-Robot Interaction. In *Proceedings Of The IROS Workshop On Cognitive Neuroscience Robotics (CNR)* (Pp. 29–34).

Jiono, M. (2020). Self Localization Based On Neighborhood Probability Mapping For Humanoid Robot. In *4th International Conference On Vocational Education And Training (ICOVET 2020)* (Pp. 355–359). Https://Doi.Org/10.1109/ICOVET50258.2020.9230237

Mahendra, R., Aji, A. F., Louvan, S., Rahman, F., & Vania, C. (2021). Indonli: A Natural Language Inference Dataset For Indonesian. In *Proceedings Of The 2021 Conference On Empirical Methods In Natural Language Processing* (Pp. 10511–10527). Https://Doi.Org/10.18653/V1/2021.Emnlp-Main.821

Moleong, L. J. (2018). *Metodologi Penelitian Kualitatif* (Ed. Revisi). Remaja Rosdakarya.

Sya, S. S., & Prihatmanto, A. S. (2015). Design And Implementation Of Image Processing System For Lumen Social Robot-Humanoid As An Exhibition Guide For Electrical Engineering Days. *Proceedings Of Electrical Engineering Days*. Https://Ieeexplore.Ieee.Org/Document/7738307

Sugiyono. (2019). *Metode Penelitian Kualitatif, Kuantitatif, Dan R&D*. Alfabeta.

Chen, J., Liu, Z., Huang, X., Wu, C., Liu, Q., Jiang, G., Pu, Y., Lei, Y., Chen, X., Wang, X., Zheng, K., Lian, D., & Chen, E. (2024). When large language models meet personalization: Perspectives of challenges and opportunities. *World Wide Web*, *27*(4), 42. https://doi.org/10.1007/s11280-024-01276-1

Cucchiarini, C., Hubers, F., & Strik, H. (2022). Learning L2 idioms in a CALL environment: The role of practice intensity, modality, and idiom properties. *Computer Assisted Language Learning*, *35*(4), 863–891. https://doi.org/10.1080/09588221.2020.1752734

Diao, L., & Hu, P. (2021). Deep learning and multimodal target recognition of complex and ambiguous words in automated English learning system. *Journal of Intelligent & Fuzzy Systems*, *40*(4), 7147–7158. https://doi.org/10.3233/JIFS-189543

Ferasso, M., Tortato, U., & Ikram, M. (2023). Mapping the Circular Economy in the Small and Medium-sized Enterprises field: An exploratory network analysis. *Cleaner and Responsible Consumption*, *11*, 100149. https://doi.org/10.1016/j.clrc.2023.100149

Gruetzemacher, R., & Paradice, D. (2022). Deep Transfer Learning & Beyond: Transformer Language Models in Information Systems Research. *ACM Comput. Surv.*, *54*(10s), 204:1-204:35. https://doi.org/10.1145/3505245

Hussain, M. (2023). When, Where, and Which?: Navigating the Intersection of Computer Vision and Generative AI for Strategic Business Integration. *IEEE Access*, *11*, 127202–127215. https://doi.org/10.1109/ACCESS.2023.3332468

Lin, J., Dai, X., Xi, Y., Liu, W., Chen, B., Zhang, H., Liu, Y., Wu, C., Li, X., Zhu, C., Guo, H., Yu, Y., Tang, R., & Zhang, W. (2025). How Can Recommender Systems Benefit from Large Language Models: A Survey. *ACM Trans. Inf. Syst.*, *43*(2), 28:1-28:47. https://doi.org/10.1145/3678004

Luckyardi, S., Karin, J., Rosmaladewi, R., Hufad, A., & Haristiani, N. (2024). Chatbots as Digital Language Tutors: Revolutionizing Education Through AI. *Indonesian Journal of Science and Technology*, *9*(3), Article 3. https://doi.org/10.17509/ijost.v9i3.79514

Luckyardi, S., Munawaroh, S., Abduh, A., Rosmaladewi, R., Hufad, A., & Haristiani, N. (2024). Advancing Language Education in Indonesia: Integrating Technology and Innovations. *ASEAN Journal of Science and Engineering*, *4*(3), Article 3. https://doi.org/10.17509/ajse.v4i3.79471

Nasution, A. H., & Onan, A. (2024). ChatGPT Label: Comparing the Quality of Human-Generated and LLM-Generated Annotations in Low-Resource Language NLP Tasks. *IEEE Access*, *12*, 71876–71900. https://doi.org/10.1109/ACCESS.2024.3402809

Paramesha, M., Rane, N., & Rane, J. (2024). *Big Data Analytics, Artificial Intelligence, Machine Learning, Internet of Things, and Blockchain for Enhanced Business Intelligence* (SSRN Scholarly Paper 4855856). Social Science Research Network. https://doi.org/10.2139/ssrn.4855856

Ragno, L., Borboni, A., Vannetti, F., Amici, C., & Cusano, N. (2023). Application of Social Robots in Healthcare: Review on Characteristics, Requirements, Technical Solutions. *Sensors*, *23*(15), Article 15. https://doi.org/10.3390/s23156820

Schiavo, F., Campitiello, L., Todino, M. D., & Di Tore, P. A. (2024). Educational Robots, Emotion Recognition and ASD: New Horizon in Special Education. *Education Sciences*, *14*(3), Article 3. https://doi.org/10.3390/educsci14030258

Younis, H. A., Ruhaiyem, N. I. R., Ghaban, W., Gazem, N. A., & Nasser, M. (2023). A Systematic Literature Review on the Applications of Robots and Natural Language Processing in Education. *Electronics*, *12*(13), Article 13. https://doi.org/10.3390/electronics12132864

Zhao, S., Wu, Y., Tsang, Y.-K., Sui, X., & Zhu, Z. (2021). Morpho-semantic analysis of ambiguous morphemes in Chinese compound word recognition: An fMRI study. *Neuropsychologia*, *157*, 107862. https://doi.org/10.1016/j.neuropsychologia.2021.107862