Inaccuracy of Malay Language Usage in Artificial Intelligence (AI) Applications: Challenges To Standard Language Quality in Malaysia: A Conceptual Paper

Nur 'Izzatty Muhiddin*., Nurul Aisyah Kamrozzaman., Noraini Saro

Faculty of Education and Humanities, UNITAR International University, Kelana Jaya, Selangor, Malaysia

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.101000276

Received: 09 October 2026; Accepted: 14 October 2026; Published: 01 November 2026

ABSTRACT

Artificial intelligence (AI) applications, including generative chatbots, machine translation systems, and automated writing assistants, are increasingly used to produce Malay-language text. These tools offer efficiency and access to language support, but their output may vary in accuracy across orthography, morphology, syntax, lexis, and pragmatic register. This conceptual paper examines the implications of such variation for standard Malaysian Malay (bahasa Melayu baku) as codified by Dewan Bahasa dan Pustaka (DBP). Drawing on a narrative, integrative synthesis of literature in language planning, sociolinguistics, computational linguistics, and educational technology, the paper proposes a framework linking data and resource conditions to possible mechanisms and manifestations of AI-generated language variation. It theorises potential consequences for users' acceptance of non-standard forms, the perceived authority of codified norms, language learning, and the functional use of Malay, while identifying AI literacy, human mediation, institutional oversight, and locally adapted language technologies as possible moderators. Six research propositions are offered for empirical testing. The paper does not present a corpus audit or establish that AI-generated Malay is inaccurate in every context; rather, it identifies a conceptual problem and specifies how its claims could be evaluated. It also considers the potential for improved multilingual AI systems to support standard Malay when they are tested against authoritative linguistic resources and used with appropriate human oversight.

Keywords: artificial intelligence, Bahasa Melayu, standard language, language planning, large language models

INTRODUCTION

Artificial intelligence has moved from the periphery to the centre of everyday textual production. Large language models (LLMs) built on the transformer architecture (Vaswani et al., 2017) and popularised through systems such as GPT-3 and its successors (Brown et al., 2020; Floridi & Chiriatti, 2020) now draft emails, translate documents, answer examination-style questions, and generate essays for millions of users, including students, teachers, journalists, and civil servants in Malaysia. The public release of conversational agents has been described as a watershed moment for education and knowledge work alike, prompting an unprecedented volume of scholarly commentary on both opportunities and risks (Dwivedi et al., 2023; Kasneci et al., 2023; van Dis et al., 2023). In the Malaysian context, these technologies increasingly operate in and on Bahasa Melayu, the national language enshrined in Article 152 of the Federal Constitution and elaborated through six decades of deliberate corpus and status planning by Dewan Bahasa dan Pustaka (DBP) (Asmah Haji Omar, 1992; Coluzzi, 2017; Gill, 2005).

Nevertheless, an important asymmetry may shape this technological uptake. Large language models are trained on corpora whose language coverage and quality are uneven, and Malay has received less NLP research and benchmarking attention than English and several other high-resource languages (Joshi et al., 2020; Poh et al., 2024). Research on multilingual systems shows that fluent output can still contain errors in language use or factual grounding, although performance varies by model, task, prompt, and language (Bang et al., 2023;

Guerreiro et al., 2023). For Malay specifically, MalayMMLU provides evidence of performance gaps on a benchmark of Malaysian school-curriculum knowledge, but this benchmark is not itself a direct measure of spelling, grammar, or adherence to DBP norms (Poh et al., 2024). It is therefore reasonable to investigate whether AI-generated Malay contains non-standard spelling, affixation, syntax, lexical choices, or register; the frequency and severity of these patterns require direct language-quality evaluation.

This issue may be understood as a language-planning question as well as a technical one. Standard languages are shaped by selection, codification, elaboration, and acceptance processes sustained by institutions over time (Haugen, 1959, 1966; Milroy, 2001). If AI systems repeatedly present forms that diverge from codified Malaysian Malay, they could influence users' perceptions of what is acceptable. However, the pathway from exposure to acceptance is a theoretical possibility, not an effect established by the present conceptual review. It depends on factors such as the quality of the output, users' existing proficiency, the task, the frequency of use, and whether the output is checked against reliable sources. The paper therefore treats normalisation and changes in language practice as propositions for empirical testing rather than demonstrated consequences.

The question is particularly relevant to education. AI tools can provide learners with accessible explanations, drafting support, translation, and feedback, but their output may also be accepted without sufficient evaluation. Research on automated feedback and AI-assisted writing identifies both opportunities and risks, including the possibility that learners may use suggestions they do not fully understand (Ranalli, 2018; Stevenson & Phakiti, 2014; Warschauer et al., 2023). Whether such patterns occur in Malaysian Malay classrooms, and whether they lead to lasting changes in learners' independent writing, requires targeted classroom and longitudinal studies. This paper consequently considers learner exposure and internalisation as plausible pathways to investigate, while recognising that guided AI use may also strengthen language learning.

Despite the growing literature, the relevant discussions are often separated. Computational studies evaluate model performance, language-planning scholarship explains how standards are maintained, and educational technology research examines adoption and learning; these strands are not always brought together to consider the quality of standard Malaysian Malay. This paper addresses that conceptual gap. It does not report new empirical data or claim to establish causal effects. Instead, it synthesises multidisciplinary literature to develop a typology, an integrative framework, and propositions that can be tested through corpus analysis, controlled experiments, and classroom research.

The objectives of this paper are fourfold: (1) to synthesise the interdisciplinary literature on AI language technologies, low-resource natural language processing, and Malay language planning; (2) to develop a typology of Malay-language inaccuracies observable in AI applications; (3) to propose an integrative conceptual framework linking the antecedents, mechanisms, manifestations, and consequences of AI-induced inaccuracy for standard language quality, together with its moderating conditions; and (4) to derive research propositions and practical implications for policymakers, educators, and technology developers in Malaysia. The remainder of the paper reviews the relevant literature and theoretical foundations, describes the conceptual methodology, presents the typology, framework, and propositions as its results, and then discusses the challenges, implications, and limitations before concluding.

LITERATURE REVIEW

Standard Malay and the institutional architecture of language planning

Bahasa Melayu occupies a constitutionally privileged position in Malaysia, and its modern standard variety is the outcome of one of Southeast Asia's most sustained language-planning enterprises. Following independence, corpus planning (the codification of spelling, grammar, and terminology) was entrusted to Dewan Bahasa dan Pustaka, whose reference grammar Tatabahasa Dewan remains the authoritative statement of standard morphology and syntax (Nik Safiah Karim et al., 2008), and whose terminological work has generated hundreds of thousands of standardised technical terms (Asmah Haji Omar, 1992). Status planning elevated Malay as the medium of national education and administration, while acquisition planning embedded the standard in the school curriculum and teacher education (Awang Sariyan, 2004; Cooper, 1989; Gill, 2005). Scholars differ in their appraisal of the enterprise: Coluzzi (2017) reads the trajectory of Malay planning as a qualified success

shadowed by the gravitational pull of globally hegemonic English (Crystal, 2003), while Albury (2017) highlights the contested folk understandings of what the national language is and should do in a multilingual polity. What is not disputed is that standard Malay is institutionally produced and maintained. This point is of central importance for the present paper, because whatever threatens the maintenance mechanisms threatens the standard itself.

Sociolinguistic theory clarifies why maintenance is perpetual work. Haugen's (1959, 1966) classic model analyses standardisation into four processes (selection of a norm, codification of form, elaboration of function, and acceptance by the community) and stresses that acceptance is never final; it must be continually re-secured against competing varieties. Milroy (2001) adds that standard languages survive on an ideology of correctness sustained by gatekeeping institutions such as schools, publishers, and language academies. Ferguson's (1959) account of diglossia further reminds us that formal, codified 'high' varieties coexist with vigorous colloquial 'low' varieties. In Malaysia, colloquial Malay and the code-mixed style popularly labelled bahasa rojak flourish in informal domains, while the formal domains of education, administration, and serious publishing have historically anchored the standard (Asmah Haji Omar, 1992; Spolsky, 2004). The entry of AI text generators into precisely these formal domains (drafting official letters, essays, and reports) is therefore of acute sociolinguistic significance: the technology intervenes at the anchor points of the standard.

Large language models and the low-resource language problem

Contemporary AI language applications are built on transformer-based neural networks (Vaswani et al., 2017) pre-trained on internet-scale corpora, whether as encoder models such as BERT and its multilingual variants (Conneau et al., 2020; Devlin et al., 2019) or as generative decoder models in the GPT family (Brown et al., 2020). The linguistic composition of the pre-training corpus decisively shapes model competence. Joshi et al. (2020) mapped the state and fate of the world's languages in NLP and placed the vast majority, including regional varieties of Malay, in resource-poor classes that receive negligible research attention and data. The consequences are systematic: multilingual models allocate limited representational capacity to under-represented languages, tokenise their morphology inefficiently, and default to patterns transferred from dominant languages (Bender et al., 2021; Conneau et al., 2020).

Empirical evaluation confirms the gap for Malay and its close relatives. Poh et al. (2024) constructed MalayMMLU, a 24,213-item benchmark drawn from the Malaysian school curriculum, and found that leading LLMs answered substantially less accurately in Malay than in English, with weaker models exhibiting pronounced deficits in language-specific subjects. Multitask evaluations of ChatGPT similarly document degraded reasoning and generation quality in non-English languages, alongside a persistent tendency to hallucinate, that is, to produce confident but false or unfounded content (Bang et al., 2023; Ji et al., 2023). In machine translation, Guerreiro et al. (2023) show that hallucination is concentrated precisely in low-resource translation directions, while regional studies of the Indonesian-Malay language ecology emphasise that scarce, fragmented, and poorly standardised data compound the difficulty of building reliable systems for the Malay world (Aji et al., 2022). A further, distinctly Malaysian complication is corpus contamination across the Malay-Indonesian boundary: because Indonesian dwarfs Malaysian Malay on the web, models frequently interpolate Indonesian lexis, spelling, and idiom into ostensibly Malaysian output, producing forms such as bisa, mobil, and gimana that are non-standard in Malaysia even though they are unremarkable in Jakarta (Aji et al., 2022; Poh et al., 2024).

The critical AI literature interprets these behaviours structurally rather than incidentally. Bender et al. (2021) argue that scale itself produces a 'stochastic parrot' dynamic in which fluency outruns understanding, and that harms fall disproportionately on communities whose languages are marginal in the training data. Weidinger et al. (2022) taxonomise the risks of language models and identify representational and quality-of-service harms for speakers of under-served languages; Blodgett et al. (2020) and Hovy and Spruit (2016) locate such harms within a broader account of the social impact of NLP; and Bird (2020) calls for decolonising language technology by centring speech communities in system design. Foundation-model scholarship likewise warns that homogenisation (many downstream applications inheriting the biases of a few base models) amplifies whatever linguistic distortions those models encode (Bommasani et al., 2021). Taken together, this literature establishes

the technical genealogy of the inaccuracy problem: it is a predictable artefact of how, and on what, the systems are trained.

AI in language education: Promise and peril

Educational technology research paints a Janus-faced picture of AI in language learning. On the promissory side, AI tutors, chatbots, and writing assistants offer personalised, always-available practice and feedback that earlier generations of computer-assisted and mobile-assisted language learning could only approximate (Godwin-Jones, 2022; Kukulska-Hulme & Shield, 2008; Luckin et al., 2016). Systematic reviews document accelerating adoption across higher education (Zawacki-Richter et al., 2019), and policy bodies have moved to articulate guardrails, with UNESCO (2021, 2023) urging human-centred deployment and explicit attention to linguistic and cultural diversity. Within applied linguistics, Kohnke et al. (2023) catalogue productive classroom uses of ChatGPT while cautioning that outputs require expert vetting; Warschauer et al. (2023) theorise the affordances and contradictions of AI-generated text for second-language writers; and Barrot (2023) warns that overreliance may short-circuit the development of writing competence.

In the Malaysian and regional context, researchers have directly examined how emerging technologies reshape language learning. Kamrozzaman et al. (2020) developed a heutagogical model of mobile learning for sustainability education, demonstrating that Malaysian adult learners readily appropriate mobile technologies for self-determined learning, a disposition that today extends seamlessly to generative AI tools. Extending this line of inquiry into the AI era, Kamrozzaman (2025) used a fuzzy Delphi procedure to validate an expert framework converging heutagogy, m-learning, and AI for sustainable language education, finding consensus that AI can scaffold each phase of self-determined language learning provided that learners exercise critical oversight of machine output. Related work in the Asian Journal of Social Science Research reinforces the same duality from a professional-education perspective: Kajiita and Murote (2025) argue that AI integration in social work education demands deliberate curricular attention to the technology's limitations and ethical risks rather than passive adoption. These regional contributions matter for the present argument because they document both the enthusiasm with which AI is being woven into educational practice and the emerging consensus that unmediated machine output cannot be trusted as a linguistic model.

On the perilous side, the feedback literature raises a specific alarm for language quality. Studies of automated writing evaluation consistently find that learners struggle to discriminate valid from invalid machine feedback and often incorporate erroneous suggestions into their texts (Grimes & Warschauer, 2010; Ranalli, 2018; Stevenson & Phakiti, 2014). Generative systems intensify the risk because they do not merely evaluate learner text; they produce model text that learners imitate wholesale. Where that model text is non-standard, as the preceding subsection shows it frequently is for Malay, imitation transmits error. AI-literacy research therefore assumes central importance: users' capacity to understand, evaluate, and critically use AI is the key individual-level defence (Long & Magerko, 2020; Ng et al., 2021), yet such literacy remains unevenly distributed across Malaysian classrooms (Kohnke et al., 2023; Selwyn, 2019).

Theoretical foundations

Two complementary theoretical traditions anchor the proposed framework. The first is Haugen's (1959, 1966) process model of language standardisation: selection of a norm, codification in grammars and dictionaries, elaboration of the norm's functional range, and acceptance by the speech community. In Malaysia, DBP's work on grammar and terminology, together with education and public administration, has helped sustain standard Malay (Asmah Haji Omar, 1992; Cooper, 1989; Nik Safiah Karim et al., 2008). Haugen's model is useful here because standardisation is not treated as a completed historical event; its norms continue to be taught, used, and renegotiated. AI systems could intersect with these processes by reproducing codified forms, introducing alternatives, or making some forms more visible than others. The direction and scale of any resulting change, however, cannot be inferred from theory alone and must be examined empirically.

The second tradition is the critical account of standard-language ideology. Milroy (2001) argues that standards are sustained partly through shared beliefs about correctness and the authority of institutions such as schools, publishers, and language academies. AI assistants may become additional sources of language advice because

they are accessible and produce fluent responses, but fluency does not guarantee conformity to an official standard (Bender et al., 2021; Ji et al., 2023). It is therefore possible that some users will treat AI output as authoritative, while others will use it as a draft and verify it against dictionaries, grammar references, or expert feedback. This framework directs attention to how users assess authority, rather than presuming that AI necessarily displaces existing language institutions. Critical NLP scholarship further highlights the importance of data representation and accountability in the design of language technologies (Bird, 2020; Blodgett et al., 2020; Bommasani et al., 2021).

Synthesis and research gap

Table 1 summarises representative works across the four literatures and their contribution to this conceptual synthesis. Three observations guide the paper. First, computational research provides evidence of performance differences across languages and tasks, but such evidence does not by itself establish the rate of deviation from DBP norms in generated Malay. Second, language-planning scholarship explains the institutional processes through which standards are maintained, but much of it predates the widespread use of generative AI. Third, educational research identifies both the affordances and risks of AI-supported learning, while the specific effects on standard Malaysian Malay remain insufficiently tested. The gap is therefore integrative and empirical: a framework is needed to connect technical conditions with possible sociolinguistic and educational outcomes and to specify how those outcomes could be evaluated.

Table 1. Representative literature informing the conceptual synthesis

Domain and source(s)	Focus	Key insight	Relevance to this paper
Language planning (Haugen, 1959, 1966; Cooper, 1989)	Processes of standardisation	Standards require selection, codification, elaboration, and continual re-acceptance	Supplies the theoretical lens: AI destabilises the acceptance process
Standard language ideology (Milroy, 2001; Ferguson, 1959)	Ideology and diglossia	Standards rest on institutionally sustained beliefs about correctness	AI emerges as a rival, non-institutional authority on 'correct' Malay
Malay language policy (Asmah Haji Omar, 1992; Gill, 2005; Coluzzi, 2017; Albury, 2017)	Status, corpus and acquisition planning for Malay	Standard Malay is an institutional achievement under pressure from English	Establishes the stakes and the institutional actors (DBP, MOE)
Low-resource NLP (Joshi et al., 2020; Conneau et al., 2020; Aji et al., 2022)	Linguistic inequality in NLP	Under-represented languages receive degraded model capacity and data	Explains why Malay output is error-prone by design, not accident
Malay benchmarking (Poh et al., 2024)	MalayMMLU evaluation of 18 LLMs	LLMs perform markedly worse in Malay than in English	Provides empirical anchoring for the inaccuracy premise
Hallucination and risk (Ji et al., 2023; Guerreiro et al., 2023; Bender et al., 2021; Weidinger et al., 2022)	Failure modes of generative models	Fluent output can be unfaithful; harms concentrate on low-resource languages	Identifies the generative mechanisms in the framework

AI in (language) education (Kasneci et al., 2023; Kohnke et al., 2023; Barrot, 2023; Warschauer et al., 2023)	Adoption, affordances, risks	Learners adopt AI text as a model but vet it poorly	Specifies the transmission channel from machine error to human usage
Regional AI-education scholarship (Kamrozzaman et al., 2020; Kamrozzaman, 2025; Kajiita & Murote, 2025)	Heutagogy, m-learning and AI; AI in professional education	Self-determined learners embrace AI, but expert consensus demands critical oversight of machine output	Grounds the moderating role of AI literacy and pedagogy in regional evidence

Source: Author's synthesis of the reviewed literature.

METHODOLOGY

Research design

This study adopts a conceptual-paper design intended to develop theory and generate testable propositions rather than to estimate effect sizes or test causal relationships. It uses a narrative, integrative review because the topic spans computational linguistics, sociolinguistics and language planning, educational technology, and AI ethics. These literatures use different methods and outcomes, making statistical meta-analysis inappropriate for the present purpose. The review should not be interpreted as a systematic review: the aim is conceptual integration, and the synthesis is qualitative.

Data sources and search strategy

Literature was identified through searches of Scopus, Web of Science, the ACL Anthology, ERIC, and Google Scholar using combinations of the terms 'artificial intelligence', 'large language models', 'ChatGPT', 'machine translation', 'low-resource languages', 'Malay', 'Bahasa Melayu', 'language planning', 'standard language', 'language quality', and 'language education'. The review prioritised peer-reviewed journal articles and major computational-linguistics proceedings published from 2016 to 2025; foundational works on language planning and standardisation regardless of publication date; and Malaysian or regional scholarship on Malay language policy, pedagogy, and technology. Reference lists of relevant works were also examined to identify additional sources. More than fifty works across four areas—computational linguistics, sociolinguistics and language planning, educational technology, and AI ethics—informed the synthesis. Sources were retained when they contributed to at least one of these areas and helped explain the linguistic, institutional, or educational dimensions of the research question. Works focused on unrelated AI applications or lacking a clear connection to language, language learning, or language policy were outside the conceptual scope. The manuscript does not currently report a database-by-database screening log; consequently, exact retrieval totals, duplicate counts, and exclusion counts are not available for reporting in this revision. This limitation is stated explicitly; the review is intended as a transparent narrative synthesis, not a fully reproducible systematic review.

Data analysis

The synthesis proceeded in three analytical steps. First, the literature was organised around three questions: how language models perform across lower-resource languages; how standard Malaysian Malay is codified and maintained; and how AI is being adopted in language-related educational and professional settings. Second, concepts and findings relevant to the research question were compared and grouped into an antecedents–mechanisms–manifestations–potential consequences structure. Third, this structure was examined against Haugen's (1959, 1966) model of language standardisation and Milroy's (2001) account of standard-language ideology. The five-part typology and the framework were developed as analytical proposals from this synthesis. Claims about model-error frequency, user acceptance, learner uptake, and long-term language change are not

treated as findings of this review unless directly supported by cited empirical research; in this paper, these pathways are framed as propositions awaiting direct testing.

RESULTS

A typology of Malay-language inaccuracy in AI applications

Drawing together research on multilingual language technologies and language learning (Aji et al., 2022; Bang et al., 2023; Guerreiro et al., 2023; Poh et al., 2024) with reference works on standard Malay (Nik Safiah Karim et al., 2008), this paper proposes a five-dimensional typology for analysing possible inaccuracies or contextually unsuitable choices in AI-generated Malaysian Malay: orthographic, morphological, syntactic, lexical-terminological, and pragmatic-register. Table 2 gives illustrative examples to support conceptual discussion. These examples are not outputs collected from a defined set of AI systems and are not evidence of frequency or prevalence. In addition, not every form that differs from formal Malaysian Malay is inherently incorrect: some are valid in another national variety, informal context, or communicative relationship. The table therefore distinguishes clear orthographic or usage issues from forms that require contextual judgement. Any empirical application should verify each item against current DBP references and a documented annotation protocol.

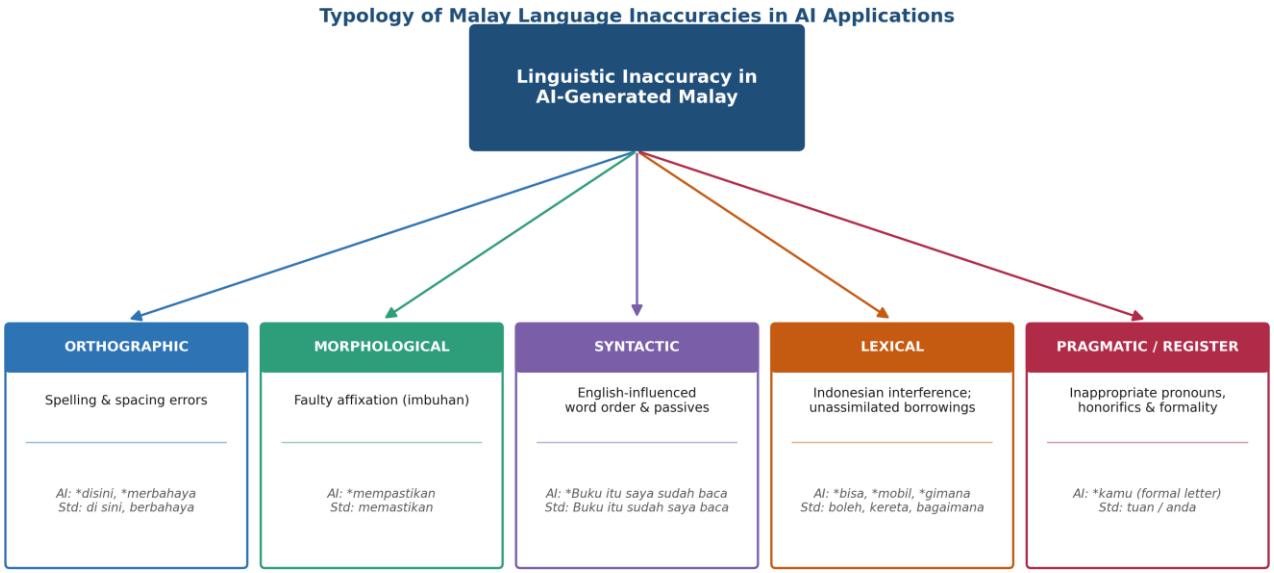

Figure 1. Typology of Malay language inaccuracies in AI applications (Source: Author)

Table 2. Illustrative examples for evaluating possible linguistic inaccuracies or contextually unsuitable forms in AI-generated Malaysian Malay

Dimension	Description	Illustrative form or issue*	Editorial interpretation / reference form	Possible mechanism to investigate
Orthographic	Spelling and word-spacing choices that may depart from standard Malaysian Malay	*disini; *dikalangan	di sini; dalam kalangan. Verify each item against current DBP spelling and usage references.	Informal spelling patterns in training or prompt data

Morphological	Affixation or word formation that may not match the intended standard form	*mempastikan	memastikan. The status of any other affixed form should be checked in context before it is labelled incorrect.	Representation of Malay morphology and tokenisation
Syntactic	Construction or word choice that may be redundant or unsuitable for the intended sentence	Repeated copular construction, e.g., *adalah merupakan	Choose a construction appropriate to the sentence; avoid treating all uses of adalah or merupakan as interchangeable without grammatical analysis.	Translation-like patterns or overgeneralisation
Lexical-terminological	Lexical choice that may be unsuitable for Malaysian standard Malay or may bypass a codified term	bisa; mobil; gimana; download-kan	These are not universally incorrect: some are associated with Indonesian usage or informal code-mixing. Assess against the intended Malaysian audience, register, and current DBP terminology.	Cross-variety transfer, code-mixing, or terminology selection
Pragmatic-register	Address forms or pronouns that may not suit the relationship, audience, or genre	kamu / engkau in a formal letter	Suitability depends on context. Select forms such as tuan, puan, or anda according to the recipient and communicative situation; do not label a pronoun wrong in every context.	Register modelling and insufficient audience context

Note. Asterisks mark forms presented for evaluation, not a claim that every item is universally incorrect. Examples are illustrative composites and have not been sampled from AI-system outputs. PRPM guidance supports “memastikan” rather than “mempastikan” and recommends “dalam kalangan” as the more grammatically precise form; the full table should still be reviewed against current DBP references and the intended audience, register, and sentence context before each item is used as an error label (Dewan Bahasa dan Pustaka, n.d.-a, n.d.-b).

The integrative conceptual framework

Figure 2 assembles the argument into an integrative framework organised along an antecedents–mechanisms–manifestations–potential consequences pathway, with four possible moderating conditions. The framework is intentionally parsimonious. Each arrow represents a hypothesised relationship to be investigated, not a causal effect established by this paper. The framework also allows for positive outcomes: where model output is accurate, well aligned with standard Malaysian Malay, and appropriately mediated, AI may support language learning, editing, and the use of Malay in professional and educational domains.

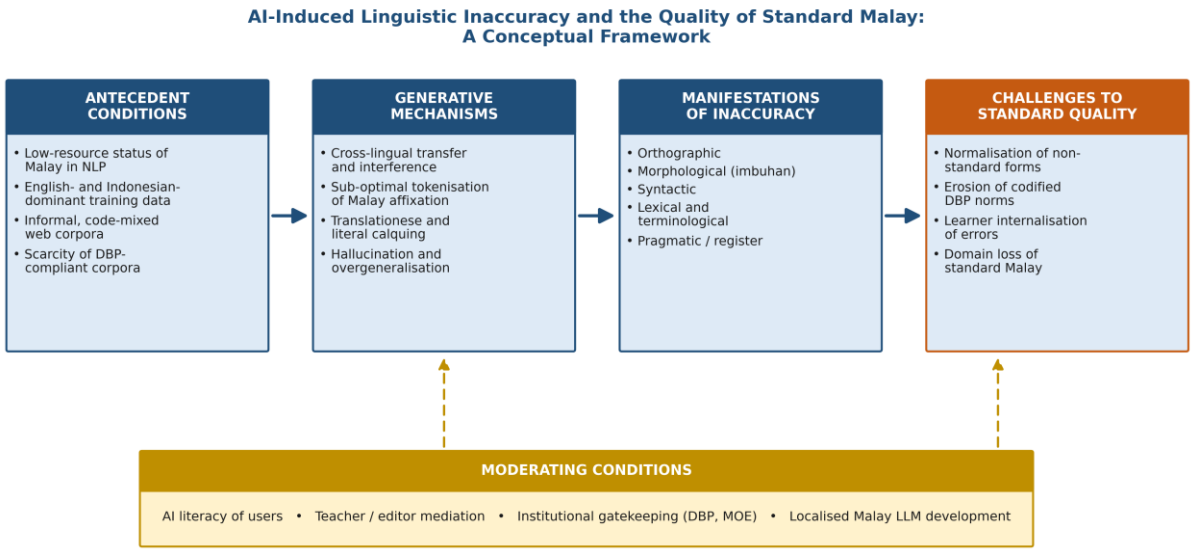

Figure 2. Conceptual framework of AI-induced linguistic inaccuracy and its challenges to standard Malay language quality (Source: Author)

Antecedent conditions. The framework identifies four conditions that may shape the quality of AI-generated Malay. These include uneven NLP resources for Malay (Joshi et al., 2020; Poh et al., 2024); the influence of English and Indonesian within multilingual training environments (Aji et al., 2022; Conneau et al., 2020); the presence of informal and code-mixed material in publicly available text; and limited availability of curated corpora explicitly aligned with DBP standards. The relative contribution of each condition is likely to vary across systems, model versions, tasks, and data sources; it should therefore be measured rather than assumed to be identical for all AI applications.

Generative mechanisms. Four mechanisms are proposed as possible explanations for non-standard or contextually unsuitable output: cross-lingual transfer, sub-optimal tokenisation of Malay morphology, translation-like syntax, and hallucination or overgeneralisation (Bang et al., 2023; Conneau et al., 2020; Guerreiro et al., 2023; Ji et al., 2023; Joshi et al., 2020; Poh et al., 2024). These mechanisms are not directly measured in the present paper. They serve as explanatory constructs that future studies can operationalise through error annotation, model comparisons, prompt variation, and analysis of outputs across tasks.

Manifestations. The proposed typology groups output issues into five dimensions: orthography, morphology, syntax, lexical terminology, and pragmatic register. These categories are intended to guide systematic evaluation. The paper does not establish that errors cluster most frequently in any one category; that claim requires a defined dataset, consistent coding rules, and inter-rater agreement assessment.

Potential consequences for standard-language quality. The framework proposes four possible downstream outcomes: greater familiarity with non-standard forms; a widening perceived gap between codified norms and everyday practice; learner uptake of AI-generated forms in subsequent writing; and reduced use of Malay in domains where users perceive English-language systems to perform better. These outcomes are theoretically plausible in light of language-planning and AI-education scholarship (Gill, 2005; Haugen, 1966; Milroy, 2001; Ranalli, 2018; Stevenson & Phakiti, 2014), but they have not been demonstrated for AI-generated Malaysian Malay by this conceptual study. They should be treated as hypotheses for experimental, longitudinal, and classroom investigation.

Moderating conditions. Four conditions may reduce the risk of inaccurate or unsuitable output being accepted without review: (a) AI literacy, including the ability to check language output against reliable sources (Long & Magerko, 2020; Ng et al., 2021); (b) teacher, editor, and professional mediation; (c) institutional guidance, auditing, and quality assurance; and (d) development or adaptation of models using curated Malaysian Malay

data and explicit standard-language evaluation. These conditions are proposed intervention points at the individual, professional, institutional, and technical levels. Their effectiveness and possible interactions require evaluation rather than assumption.

Research propositions

To make the framework empirically tractable, six propositions are advanced in Table 3. P1–P3 concern the form and distribution of linguistic deviations; P4–P5 concern possible user and learner outcomes; and P6 concerns moderating conditions. The propositions are not findings. They are hypotheses derived from the conceptual synthesis and should be refined against operational definitions, sampling decisions, and validated annotation procedures before testing.

Table 3. Research propositions derived from the conceptual framework

No.	Proposition (to be tested)	Suggested operationalisation / basis
P1	In a matched evaluation dataset, AI-generated Malaysian Malay will contain a higher rate of deviations from predefined DBP-aligned criteria than professionally edited Malaysian Malay.	Compare matched texts using a validated rubric; report deviations per 1,000 words and by category. Prior NLP benchmarking motivates evaluation but does not establish this outcome (Joshi et al., 2020; Poh et al., 2024).
P2	The distribution of deviations will differ by linguistic feature, with affixation, formal register, and codified terminology emerging as candidate high-risk categories.	Code errors by category and compare rates; do not assume clustering before observing the dataset. Cross-lingual transfer and tokenisation provide candidate explanations (Conneau et al., 2020).
P3	Indonesian-associated lexical or orthographic forms may constitute a substantial category of lexical mismatch in AI-generated Malaysian Malay.	Define the category with Malaysian Malay experts and compare its count with other lexical categories. Whether it is the largest category remains an open empirical question (Aji et al., 2022).
P4	Repeated exposure to selected non-standard AI-generated forms may increase their acceptability ratings relative to a comparison condition, with possible variation by age and proficiency.	Use a controlled pre/post exposure design with standard-form and control conditions; measure acceptability and report subgroup analyses. This is a proposed test of the acceptance mechanism (Haugen, 1966; Milroy, 2001).
P5	Learners using AI without structured linguistic mediation may reproduce more AI-typical deviations in subsequent independent writing than learners receiving explicit mediation or writing without AI.	Compare independent writing across clearly defined conditions, controlling for proficiency and task difficulty. Automated-feedback research motivates the question but does not establish this specific Malay-language effect (Ranalli, 2018; Stevenson & Phakiti, 2014; Warschauer et al., 2023).
P6	AI literacy, teacher/editor mediation, institutional quality assurance, and locally adapted models may moderate relationships between AI exposure, acceptability, and subsequent writing outcomes.	Measure each moderator separately and test interaction effects. An additive effect should be hypothesised only where the design explicitly tests combined effects (Long & Magerko, 2020; Ng et al., 2021; UNESCO, 2023).

Source: Author's formulation.

DISCUSSION

The inversion of the acceptance function

The first proposed challenge concerns the relationship between exposure and acceptance. Historically, textbooks, broadcasting, schools, and official publications have helped circulate codified forms (Cooper, 1989; Haugen, 1966). AI systems add another source of text that may be encountered frequently and interactively. If an AI system repeatedly produces a form that diverges from standard Malaysian Malay, repeated exposure could increase familiarity with that form. However, familiarity does not necessarily lead to acceptance, and acceptance does not necessarily lead to use. These links may be affected by prior proficiency, explicit instruction, source credibility, task type, and access to authoritative references. The proposed 'inversion' of the acceptance function is therefore a theoretical mechanism to test, not a conclusion that AI has already weakened standard Malay.

Algorithmic authority and the displacement of gatekeepers

A second issue concerns perceived authority. Standard-language ideology is partly maintained through recognition of institutions that define and teach codified norms (Milroy, 2001). Because AI assistants are accessible and often fluent, some users may consult them for language advice. Yet users may also treat them as brainstorming tools or verify responses against DBP resources. The relevant empirical question is not whether AI inevitably displaces language gatekeepers, but when, for whom, and in what tasks AI output is trusted over established references. The possibility of a divergence between official codification and the forms generated by AI merits investigation, especially where systems do not explicitly evaluate output against Malaysian standard-language norms (Bender et al., 2021; Bird, 2020; Ji et al., 2023).

Educational transmission of error

Education provides a particularly important setting in which to investigate these questions. Studies of automated feedback and AI-assisted writing suggest that learners may accept some machine suggestions without sufficient scrutiny, but uptake depends on task design, learner proficiency, and instructional mediation (Barrot, 2023; Grimes & Warschauer, 2010; Ranalli, 2018; Stevenson & Phakiti, 2014; Warschauer et al., 2023). For Malaysian Malay, it remains an empirical question whether learners reproduce AI-typical deviations in later independent writing and whether guided use improves their ability to evaluate language. Classroom studies should compare unassisted writing, AI-assisted writing with explicit teacher mediation, and AI-assisted writing without such mediation. Assessment design and professional development can help teachers make these distinctions visible while preserving the potential benefits of AI-supported drafting and feedback (Holmes & Tuomi, 2022; UNESCO, 2023).

Terminological bypass and domain loss

A further question concerns terminology and the functional use of Malay. AI tools may provide useful explanations and translations, but they may also select English terms where a standard Malay equivalent is available or generate plausible terms that require verification. The example 'download-kan' should be treated as an illustrative code-mixed form, not as evidence that systems systematically bypass DBP terminology. Likewise, users' choice of English or Malay may depend on the task, audience, available terminology, and perceived model performance. Research should therefore assess terminology against current DBP resources and investigate whether AI support expands or narrows the use of standard Malay in specific professional and educational domains (Asmah Haji Omar, 1992; Coluzzi, 2017; Gill, 2005; Poh et al., 2024).

The evolving capabilities of multilingual AI and opportunities for standard Malay

The risk-focused framework should be balanced against the possibility that multilingual AI systems can improve and provide useful support for Malay. AI tools may help users brainstorm, translate, explain unfamiliar vocabulary, practise writing, and access language assistance outside formal learning settings. Their value depends on the system, task, prompt, and the quality criteria used to judge the output; results from one benchmark or model version should not be generalised to all applications. Language-specific evaluation, retrieval from

authoritative sources, explicit prompting for Malaysian standard Malay, and human post-editing may improve usefulness, but these approaches should be tested rather than assumed to guarantee correctness. Future benchmarking should compare model versions and task types, assess output against current DBP references, and report both error rates and successful language-support outcomes. This balanced approach treats AI as a technology whose effects are shaped by design and use, rather than as an inherently harmful or inherently reliable language authority.

Operationalisation and empirical evaluation of the propositions

The six propositions require explicit measures before they can be tested. For P1, researchers could compare a matched sample of AI-generated Malay and professionally edited Malaysian Malay using a predefined rubric aligned with DBP spelling, morphology, syntax, terminology, and register guidance. For P2, annotators could code each deviation by linguistic category and report rates per word or sentence, with inter-rater agreement. For P3, researchers should define in advance what counts as Indonesian-associated lexical or orthographic interference, consult Malaysian Malay experts, and test whether this category is the largest among all coded lexical deviations rather than assuming it is. For P4, a controlled exposure study could measure acceptability ratings before and after repeated exposure to carefully selected standard and non-standard forms, with appropriate consent and debriefing. For P5, learners' subsequent independent writing could be compared across unassisted, AI-assisted with mediation, and AI-assisted without mediation conditions, controlling for proficiency and task difficulty. For P6, AI literacy, human mediation, institutional review, and model localisation should each have observable indicators; their separate and combined effects can then be tested. Claims that their effects are 'additive' should be retained only if the study design and analysis explicitly test that assumption. Across all studies, the model name and version, date, prompts, settings, sampling procedure, reference standard, coding guide, and reliability statistics should be reported.

Implications for policy, pedagogy, and technology development

The framework's proposed moderators suggest a shared responsibility among Malaysian stakeholders (Table 4). DBP could consider making authoritative orthographic, grammatical, and terminological resources easier to use in machine-readable formats, alongside a public benchmark for evaluating standard-Malay output. The Ministry of Education and higher education institutions could incorporate linguistic AI literacy into teaching, including explicit practice in checking affixation, terminology, spelling, and register against authoritative sources. Assessment could require students to document how they checked and revised AI-assisted text, rather than treating AI use itself as evidence of poor learning. These recommendations align with broader guidance on human-centred and responsible AI in education (UNESCO, 2021, 2023), while their design should be adapted to local needs and evaluated for effectiveness.

Technology developers and research funders could support curated Malaysian Malay datasets, transparent language-quality evaluation, and model adaptation that explicitly measures compliance with standard Malaysian Malay. Public-sector procurement could include language-quality criteria where AI tools are used for official communication. Teachers, editors, and professional communicators could use a verification workflow in which AI output is treated as a draft requiring review, with the level of review proportionate to the importance of the text. These actions are recommendations derived from the framework; their costs, feasibility, and effectiveness should be evaluated through pilot projects rather than assumed in advance (Bird, 2020; Holmes & Tuomi, 2022; Kohnke et al., 2023; Poh et al., 2024; Selwyn, 2019).

Table 4. Stakeholder implications derived from the moderating conditions

Stakeholder	Framework challenge addressed	Recommended strategic response
Dewan Bahasa dan Pustaka	Algorithmic displacement of codification authority; terminological bypass	Machine-readable standards; official standard-Malay compliance benchmark; periodic public audits of major AI systems

Ministry of Education Malaysia	Educational transmission of error	Linguistic AI-literacy curriculum; assessment redesign requiring human revision of AI-assisted text; teacher professional development
Higher education institutions	Learner internalisation; research gap	Institutional AI-use guidelines for Malay-medium writing; empirical research programmes testing P1 to P6
Technology developers and funders	Antecedent data conditions; low-resource mechanisms	Curated standard-compliant Malay corpora; locally fine-tuned LLMs; procurement standards for government AI
Teachers, editors, media practitioners	Normalisation of non-standard forms	Editorial verification workflows; professional upskilling as critical quality-assurers of AI-mediated Malay

Source: Author's formulation based on the conceptual framework.

Limitations and future research

Several limitations should guide interpretation. First, the paper is a conceptual synthesis and does not test the framework against a newly collected corpus or participant data. The examples in Table 2 are illustrative composites, not verified instances sampled from specific AI systems; they should not be interpreted as evidence of prevalence. Before empirical use, each example should be checked against current DBP references and reviewed by qualified Malay-language experts. Second, the manuscript does not report a complete database-by-database screening log, so exact search yields, duplicates, and exclusion counts are not available in the current version. Third, multilingual AI capabilities are changing, and the quality of Malay output may differ substantially across model versions, prompts, and tasks. Evaluation should therefore be repeated over time and should compare both limitations and improvements. Fourth, this paper focuses on standard Malaysian Malay and does not generalise its claims to Indonesian, Brunei Malay, or other languages and varieties. Future work should include (i) corpus-based benchmarking against DBP-aligned criteria; (ii) experiments on exposure, acceptability, and later production; (iii) classroom studies comparing mediated and unmediated AI use; and (iv) design-based evaluation of systems intended to support standard Malay.

CONCLUSION

This conceptual paper has proposed a five-dimensional typology and an integrative framework for examining how AI-generated Malay may diverge from standard Malaysian Malay and what consequences could follow. Its contribution is to connect research on multilingual NLP, language planning, and AI-supported education in a set of testable propositions, while identifying potential interventions at individual, professional, institutional, and technical levels. The paper does not establish the prevalence of linguistic inaccuracies in current AI systems or demonstrate that exposure has already weakened standard Malay. These are empirical questions requiring carefully sampled outputs, DBP-informed annotation, and user and classroom studies. At the same time, AI should not be regarded only as a source of risk: appropriately evaluated multilingual systems may support Malay-language drafting, translation, terminology access, and learning. The central practical challenge is to establish reliable evaluation, transparent human review, and collaboration between language institutions, educators, researchers, and developers so that AI can support rather than undermine the quality and continued use of standard Malaysian Malay.

ACKNOWLEDGEMENTS

The authors would like to thank UNITAR International University for supporting this research.

REFERENCES

1. Aji, A. F., Winata, G. I., Koto, F., Cahyawijaya, S., Romadhony, A., Mahendra, R., Kurniawan, K., Moeljadi, D., Prasojo, R. E., Fung, P., Baldwin, T., Lau, J. H., Sennrich, R., & Ruder, S. (2022). One country, 700+ languages: NLP challenges for underrepresented languages and dialects in Indonesia. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 7226-7249). Association for Computational Linguistics. https://doi.org/10.18653/v1/2022.acl-long.500
2. Albury, N. J. (2017). Mother tongues and languaging in Malaysia: Critical linguistics under critical examination. Language in Society, 46(4), 567-589.
3. Asmah Haji Omar. (1992). The linguistic scenery in Malaysia. Dewan Bahasa dan Pustaka.
4. Awang Sariyan. (2004). Teras pendidikan bahasa Melayu: Asas pegangan guru. PTS Publications & Distributors.
5. Bang, Y., Cahyawijaya, S., Lee, N., Dai, W., Su, D., Wilie, B., Lovenia, H., Ji, Z., Yu, T., Chung, W., Do, Q. V., Xu, Y., & Fung, P. (2023). A multitask, multilingual, multimodal evaluation of ChatGPT on reasoning, hallucination, and interactivity. In Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (pp. 675-718). Association for Computational Linguistics.
6. Barrot, J. S. (2023). Using ChatGPT for second language writing: Pitfalls and potentials. Assessing Writing, 57, Article 100745. https://doi.org/10.1016/j.asw.2023.100745
7. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (pp. 610-623). Association for Computing Machinery. https://doi.org/10.1145/3442188.3445922
8. Bird, S. (2020). Decolonising speech and language technology. In Proceedings of the 28th International Conference on Computational Linguistics (pp. 3504-3519). International Committee on Computational Linguistics.
9. Blodgett, S. L., Barocas, S., Daumé, H., III, & Wallach, H. (2020). Language (technology) is power: A critical survey of 'bias' in NLP. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 5454-5476). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.485
10. Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv. https://arxiv.org/abs/2108.07258
11. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., & Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901.
12. Coluzzi, P. (2017). Language planning for Malay in Malaysia: A case of failure or success? International Journal of the Sociology of Language, 2017(244), 17-38.
13. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2020). Unsupervised cross-lingual representation learning at scale. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 8440-8451). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.747
14. Cooper, R. L. (1989). Language planning and social change. Cambridge University Press.
15. Crystal, D. (2003). English as a global language (2nd ed.). Cambridge University Press.
16. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long and Short Papers) (pp. 4171-4186). Association for Computational Linguistics. https://doi.org/10.18653/v1/N19-1423
17. Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., & Wright, R. (2023). 'So what if ChatGPT wrote it?' Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational

AI for research, practice and policy. International Journal of Information Management, 71, Article 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

18. Ferguson, C. A. (1959). Diglossia. Word, 15(2), 325-340. https://doi.org/10.1080/00437956.1959.11659702
19. Floridi, L., & Chiriatti, M. (2020). GPT-3: Its nature, scope, limits, and consequences. Minds and Machines, 30(4), 681-694. https://doi.org/10.1007/s11023-020-09548-1
20. Gill, S. K. (2005). Language policy in Malaysia: Reversing direction of language shift in tertiary education. Language Policy, 4(3), 241-260.
21. Godwin-Jones, R. (2022). Partnering with AI: Intelligent writing assistance and instructed language learning. Language Learning & Technology, 26(2), 5-24.
22. Grimes, D., & Warschauer, M. (2010). Utility in a fallible tool: A multi-site case study of automated writing evaluation. Journal of Technology, Learning, and Assessment, 8(6), 1-43.
23. Guerreiro, N. M., Alves, D. M., Waldendorf, J., Haddow, B., Birch, A., Colombo, P., & Martins, A. F. T. (2023). Hallucinations in large multilingual translation models. Transactions of the Association for Computational Linguistics, 11, 1500-1517.
24. Haugen, E. (1959). Planning for a standard language in modern Norway. Anthropological Linguistics, 1(3), 8-21.
25. Haugen, E. (1966). Dialect, language, nation. American Anthropologist, 68(4), 922-935.
26. Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. European Journal of Education, 57(4), 542-570. https://doi.org/10.1111/ejed.12533
27. Hovy, D., & Spruit, S. L. (2016). The social impact of natural language processing. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers) (pp. 591-598). Association for Computational Linguistics. https://doi.org/10.18653/v1/P16-2096
28. Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., & Fung, P. (2023). Survey of hallucination in natural language generation. ACM Computing Surveys, 55(12), 1-38. https://doi.org/10.1145/3571730
29. Joshi, P., Santy, S., Budhiraja, A., Bali, K., & Choudhury, M. (2020). The state and fate of linguistic diversity and inclusion in the NLP world. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (pp. 6282-6293). Association for Computational Linguistics. https://doi.org/10.18653/v1/2020.acl-main.560
30. Kajiita, R. M., & Murote, S. (2025). Social work and artificial intelligence: Implications for South African social work practice and education. Asian Journal of Social Science Research, 7(1), 1-19. https://doi.org/10.5281/zenodo.15803208
31. Kamrozzaman, N. A. (2025). Converging heutagogy, M-learning, and AI for sustainable language education: A fuzzy Delphi study. Edelweiss Applied Science and Technology, 9(9), 1600-1608. https://doi.org/10.55214/2576-8484.v9i9.10167
32. Kamrozzaman, N. A., Badusah, J., & Wan Mohammad, W. M. R. (2020). Development of heutagogy approach in M-learning for sustainability education. Education and Information Technologies, 25(4), 3035-3047.
33. Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 103, Article 102274. https://doi.org/10.1016/j.lindif.2023.102274
34. Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. RELC Journal, 54(2), 537-550. https://doi.org/10.1177/00336882231162868
35. Kukulska-Hulme, A., & Shield, L. (2008). An overview of mobile assisted language learning: From content delivery to supported collaboration and interaction. ReCALL, 20(3), 271-289. https://doi.org/10.1017/S0958344008000335
36. Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727
37. Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). Intelligence unleashed: An argument for AI in education. Pearson Education.

38. Milroy, J. (2001). Language ideologies and the consequences of standardization. Journal of Sociolinguistics, 5(4), 530-555. https://doi.org/10.1111/1467-9481.00163
39. Ministry of Education Malaysia. (2013). Malaysia Education Blueprint 2013-2025 (Preschool to post-secondary education). Kementerian Pendidikan Malaysia.
40. Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. Computers and Education: Artificial Intelligence, 2, Article 100041. https://doi.org/10.1016/j.caeai.2021.100041
41. Nik Safiah Karim, Farid M. Onn, Hashim Musa, & Abdul Hamid Mahmood. (2008). Tatabahasa Dewan (3rd ed.). Dewan Bahasa dan Pustaka.
42. Poh, S. C., Yang, S. J., Tan, J. M. L., Chieng, L. L. T. Y., Tan, J. X., Yu, Z., Foong, C. M., & Chan, C. S. (2024). MalayMMLU: A multitask benchmark for the low-resource Malay language. In Findings of the Association for Computational Linguistics: EMNLP 2024 (pp. 650-669). Association for Computational Linguistics. https://doi.org/10.18653/v1/2024.findings-emnlp.36
43. Ranalli, J. (2018). Automated written corrective feedback: How well can students make use of it? Computer Assisted Language Learning, 31(7), 653-674. https://doi.org/10.1080/09588221.2018.1428994
44. Selwyn, N. (2019). Should robots replace teachers? AI and the future of education. Polity Press.
45. Spolsky, B. (2004). Language policy. Cambridge University Press.
46. Stevenson, M., & Phakiti, A. (2014). The effects of computer-generated feedback on the quality of writing. Assessing Writing, 19, 51-65. https://doi.org/10.1016/j.asw.2013.11.007
47. UNESCO. (2021). AI and education: Guidance for policy-makers. UNESCO Publishing.
48. UNESCO. (2023). Guidance for generative AI in education and research. UNESCO Publishing.
49. van Dis, E. A. M., Bollen, J., Zuidema, W., van Rooij, R., & Bockting, C. L. (2023). ChatGPT: Five priorities for research. Nature, 614(7947), 224-226. https://doi.org/10.1038/d41586-023-00288-7
50. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998-6008.
51. Warschauer, M., Tseng, W., Yim, S., Webster, T., Jacob, S., Du, Q., & Tate, T. (2023). The affordances and contradictions of AI-generated text for writers of English as a second language. Journal of Second Language Writing, 62, Article 101071. https://doi.org/10.1016/j.jslw.2023.101071
52. Weidinger, L., Uesato, J., Rauh, M., Griffin, C., Huang, P.-S., Mellor, J., Glaese, A., Cheng, M., Balle, B., Kasirzadeh, A., & Gabriel, I. (2022). Taxonomy of risks posed by language models. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency (pp. 214-229). Association for Computing Machinery. https://doi.org/10.1145/3531146.3533088
53. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education: Where are the educators? International Journal of Educational Technology in Higher Education, 16(1), Article 39. https://doi.org/10.1186/s41239-019-0171-0
54. Dewan Bahasa dan Pustaka. (n.d.-a). Mempastikan. Pusat Rujukan Persuratan Melayu (PRPM). https://prpm.dbp.gov.my/cari1?d=175768&keyword=mempastikan
55. Dewan Bahasa dan Pustaka. (n.d.-b). Dikalangan. Pusat Rujukan Persuratan Melayu (PRPM). https://prpm.dbp.gov.my/Cari1?d=175768&keyword=dikalangan