Journal Of the Iraqia University (77-1) 30 September (2026)

ISSN(Print): 1813-4521 OnLine ISSN:2663-7502
Journal Of the Iraqia University
available online at
https://iasj.rdd.edu.iq/journals/journal/view/247

# Envisioning the future of language in the era of artificial intelligence a linguistic study

# تخيُّل مستقبل اللغة في عصر الذكاء الاصطناعي: دراسة لغوية

مدرس مساعد. ابراهيم علي زغير عودة

جامعة بغداد / كلية الطب

Assistant lecturer. Ibrahim Ali zghair
Baghdad University /College of Medicine
ibrahim.a@comed.uobaghdad.edu.iq

**الملخص**

تحلل هذه المراجعة الوضع الراهن، والتحديات، والمسارات المستقبلية للذكاء الاصطناعي في مجال تعليم اللغات، مستخلصةً الأدلة والبراهين من المراجعات المنهجية والدراسات التطبيقية المنشورة بين عامي ٢٠١٣ و٢٠٢٥. وقد أظهر الذكاء الاصطناعي التوليدي ونماذج اللغة الكبيرة، بما في ذلك "ChatGPT" و "GPT-4"و"LLaMA"، قدرات استثنائية في التقييم الآلي للكتابة، والممارسة المحادثات، والتعلم التكيفي، والتغذية الراجعة المدمجة عبر لغات متنوعة. ومع ذلك، لا تزال هناك مخاوف جوهرية تتعلق بالانحياز اللغوي؛ إذ تؤدي نماذج اللغة الكبيرة بشكل أفضل بكثير في اللغات ذات الدعم المرتفع مثل الإنجليزية (بدقة تصل إلى ٩٢٪) مقارنة باللغات شحيحة الموارد واللغات الأصلية (بدقة تقل عن ٢٠٪)، مما يثير قضايا تتعلق بالهيمنة الرقمية والعدالة اللغوية. وتشمل التحديات الإضافية انتهاكات النزاهة الأكاديمية التي تتطلب تحولاً جذرياً في نماذج التقييم القائمة على النتاج النهائي إلى التقييم القائم على العملية التعليمية نفسها، فضلاً عن احتمالية تدهور التفاعل البشري والروابط الاجتماعية في عملية تعلم اللغات. وباستخدام منهجية السيناريوهات المستقبليّة، تحدد هذه المراجعة أربعة مسارات محتملة: النمو المستمر، الانهيار، الانضباط، والتحول. وترى الدراسة أن المسار الذي سيتحقق سيعتمد على القرارات التي يتخذها حالياً الباحثون، وصنّاع السياسات، والتربويون، ومخططو التكنولوجيا. وتدعو المراجعة إلى تبني نموذج شراكة تعاونية يعزز فيه الذكاء الاصطناعي دور المعلم والتفاعل بين الأقران بدلاً من استبدالهما، وبما يحافظ على الطابع الاجتماعي الجوهري لاكتساب اللغة. وتتضمن التداعيات على السياسات العامة سن تشريعات خاصة بالذكاء الاصطناعي في التعليم، والاستثمار في بيانات التدريب متعددة اللغات، وفرض الشفافية الإدارية وتحليل الانحياز في برامج الذكاء الاصطناعي التعليمية تعليم اللغات، التعاون بين الانسان والذكاء الاصطناعي(LLMs) ، **الكلمات المفتاحية: الذكاء الاصطناعي، نماذج اللغة الكبيرة الانحياز اللغوي , الدراسات المستقبلية**

**Abstract**

This review analyzes the contemporary state, confrontations, and future courses of artificial intelligence in style instruction, synthesizing evidence from orderly reviews and practical studies written between 2013 and 2025. Generative AI and abundant prose models, containing ChatGPT, GPT-4, and LLaMA, have demonstrated exceptional potential for automated letter judgment, talkative practice, adaptive stage, and embodied feedback across diversified expressions. However, meaningful concerns persist concerning semantic bias, as LLMs act dramatically better for extreme-support languages like English (92% veracity) than for reduced-capital and indigenous vocabularies (beneath 20% veracity), raising issues of mathematical expansionist doctrine and linguistic lawfulness. Additional challenges involve academic uprightness disruptions needing a paradigm shift from device-located to process-located assessment, in addition to the potential deterioration of human interplay and social link in vocabulary learning. Using futures-familiarize synopsis reasoning, this review identifies four believable courses—Continued Growth, Collapse, Discipline, and Transformation—and disputes that the path

fulfilled will believe decisions fashioned contemporary by investigators, policymakers, educators, and technology planners. The review advocates for a cooperative co-ship model wherein AI augments alternatively replaces lecturer fate and peer interaction, continuing the basically social character of style addition. Implications for policy involve instruction-particular AI legislation, grant in multilingual preparation data, and administered transparence and bias analyzing for educational AI plans.

**Keywords:**Artificial Intelligence (AI), Large Language Models (LLMs), Language Education, Human-AI Collaboration, Linguistic Bias, Futures Studies

## 1. Introduction

### ١,١The AI Watershed in Language Education

The public release of ChatGPT in November 2022 apparent a area for water draining moment in the record of instructional technology. For the first occasion, fruitful artificial intelligence (GenAI) worthy bearing grammatically correct, contextually relevant, and human-like accent enhanced freely accessible to a great number of heaps of users general (Alaqlobi and others., 2024). Within five days of allure launch, ChatGPT attained individual million consumers—the fastest mathematical service endorsement in record at that time—basically changing public perception of what machines commit obtain in rules long considered particularly human (Weng & Chiu, 2025).

This mechanics leap did not arise from a emptiness; it represented the conclusion of decades of research in machine intelligence, natural language processing, and machine intelligence. However, allure sudden approachability catalyzed a example shift in how educators, analysts, and policymakers gestate the act of technology in expression education and learning.The importance concerning this moment for word instruction cannot be overstated. Language education is basically about communication, and for the first opportunity, machines manifested an original capacity to undertake significant, contextually appropriate linguistic interplay. Large sound models (LLMs) such as GPT-4, LLaMA, and Aya exhibit ascendable multilingual wherewithal, though with important differences in performance across expressions, specifically disadvantaging low-support dialects that wait underrepresented in preparation datasets and evaluations (Georgiou, 2026).

These sciences have rapidly developed from exploratory curiosities to practical finishes that undergraduates and teachers encounter routine, lifting profound questions about education, appraisal, teacher parts, and the very character of accent learning itself.Artificial intelligence, delineated for one OECD (2024) as "a machine-located scheme that can, for a given set of human-delineated aims, make prophecies, pieces of advice, or decisions doing legitimate or in essence environments" (p. 7), surrounds a different array of technologies having to do with expression education. These involve mechanical writing judgment (AWE) wholes, conversational powers and chatbots, system translation forms, mechanical talk recognition (ASR), imaginative instruction systems, and, again, multimodal fruitful AI capable of bearing handbook, images, and visual and audio entertainment transmitted via radio waves (Zhu & Wang, 2025; Qiao and others., 2025). Each of these technologies offers different affordances while together presenting complex challenges that demand careful academic consideration.

### 1.2 The Transformative Potential and Pressing Challenges

AI technologies hold the promise of discussing determined disputes in style instruction that have long resisted acceptable resolutions. Conventional classroom backgrounds struggle to supply students accompanying enough language practice excuse, distinguished feedback on act, and adjusting learning ways tailor-made to diverse beginner needs (Weng & Chiu, 2025). AI-stimulate tools offer potential remedies: chatbots can comprise forever-available discourse friends, reducing accent worry and increasing readiness to ideas (Huang et al., 2022); mechanical literature evaluation structures can determine immediate, climbable response on student novel (Steiss and others., 2024); and intelligent instruction arrangements can adapt command to individual pupil progress (Xu et al., 2022).

Empirical evidence implies that AI can reinforce language education happenings both cognitively and affectively, reconstructing word performance, create literature ideas, and lowering tension among second word learners (Qiao and others., 2025; Zhu & Wang, 2025).

However, the integration of AI into vocabulary instruction is accompanied by important challenges that demand severe investigation. Technical disadvantages remain, including dependability and veracity issues, with AI arrangements frequently generating "hallucinations"—content that sounds believable but is wrong or outdated (Alaqlobi and others., 2024). Automatic talk recognition methods exhibit bias across vocabularies, accents, gender, and age, lifting concerns about impartiality and inclusion (Georgiou, 2026). Pedagogical challenges thrive, containing the risk of student over-confidence on electronic tools conceivably belittling creativity and

fault-finding thinking, and the lack of clear designs for mixing AI with settled teaching approaches (Qiao et al., 2025).

Ethical concerns about cultural biases, common standarding in LLM outputs, and data solitude stress the pressing need for appropriate government foundations (Holmes et al., 2022; UNESCO, 2021). Furthermore, the teaching potential of AI frequently remains hidden on account of insufficient unification accompanying educational hypotheses and restricted attention to in what way or manner these forms can orchestrate significant peer-to-peer interplay (Zhu & Wang, 2025).The research landscape has behaved to these incidents with extraordinary spurring. Bibliometric analyses tell that magazines on AI in language instruction waited minimal just before nearly 2017, after that a important and continuous increase happen, especially peaking in 2024 accompanying almost 400 publications (Zhu & Wang, 2025).

This sharp upward style coincides accompanying global progresses in AI electronics and its extended approachability, particularly the extensive unification of generative AI models (Alaqlobi and others., 2024). Scholars have transported systematic reviews analyzing distinguishing technologies to a degree chatbots (Huang and others., 2022), conversational AI (Zhang and others., 2023), and electronic writing judgment (Steiss and others., 2024), as well as fuller flows in AI-assisted terminology knowledge (Xu et al., 2022; Qiao and others., 2025). Yet, regardless of this growing crowd of grant, the field lacks a comprehensive, futures-familiarize combination that not only documents current developments but orderly survey plausible courses for style education in an stage of accelerated technological renewal. (Qiao et al., 2025).

**1.3 The Imperative for Futures-Oriented Inquiry**

The need for futures-familiarize study arises from various concerns. First, the pace of concerning details change consistently outstrips academic information phases, creating a determined delay betwixt innovation and academic understanding (Georgiou, 2026). Research written contemporary examines electronics that can previously have evolved extensively, making backward-looking reasoning necessary but lacking for apprising proactive accountable.

Second, the unification of AI into instruction is not a technologically deterministic process but rather a sociotechnical individual formed by human selections, institutional procedures, and social principles (Selwyn, 2022). Understanding what could take place is essential for forming what should take place. Third, educators, policymakers, and organizations face next decisions about AI acceptance, professional growth, and government with restricted counseling about unending implications. As Pegrum (2025) disputes destructively, we endure not be focusing on instructional coup, but instead on instructional evolution—a slow, stage-by-stage happening that demands careful guiding along route, often over water alternatively excited embrace of rebellious wordiness (as subpoenaed in Georgiou, 2026).

This perspective joins accompanying more extensive critiques of mechanics resignation to a fate in education. History manifests that arising electronics—from radios to televisions, and from computers to travelling phones—have usually happened accompanied by promises of instructional shift that abandoned to materialize (Cuban, 1986; Selwyn, 2022). Education is a complex environment portion multiple purposes and diversified colleagues, containing students, faculty members, persons, employers, and governments; electronics is just one item inside this abundant ecosystem (Selwyn, 2022). As Reich (2020) observes, "Trying to expedite education by elevate up technology is like dawdling rockets on worry wings. More force does not lead linearly to more progress" (p. 156). Improvements in instruction exceptionally accomplish goal dramatic renewals; they arise through deep, unending commitment to construction healthier structures (Reich, 2020).

The futures-oriented approach selected in this place review recognizes that the course of AI in dialect instruction is not predetermined but as long as determinations fashioned today by investigators, educators, policymakers, and electronics builders. By systematically testing current evidence and jutting believable futures, this review aims to support informed in charge that harnesses AI's potential while insuring educational principles and human instrumentality. This approach is specifically urgent likely the current tactics emptiness: despite brisk concerning details progress, education-particular AI regulation debris nascent, accompanying most government frameworks operating at extreme levels of pensiveness that offer restricted guidance for lecture room practice (Bernstein, 2025; European Commission, 2022).

**1.4 Purpose, Research Questions, and Scope**

This review detail discusses the following research questions:

1. What is the current state of practical research on AI in word instruction concerning electronics, requests, and influence?

2. What are the key time, challenges, and government foundations forming AI unification in expression instruction?
3. What plausible futures maybe conceived for accent instruction in the time of machine intelligence?
4. What are the suggestions for research, procedure, and practice?

In calling these questions, this review focuses on AI as a direct judge and additional assistant in word knowledge frameworks, following the foundation proposed by Xu and others. (2022). This contains four substitute-classifications of AI application: adjusting knowledge and embodiment, imaginative appraisal and feedback, create a likeness in a picture and guess, and brilliant instruction systems (Xu and others., 2022). The review stresses fruitful AI and abundant terminology models while acknowledging basic electronics that touch shape the field. Geographically, the review adopts a worldwide scope accompanying consideration to under-depicted circumstances, recognizing that research result remnants reduced in sure domains while the impacts of AI are felt general (Zhu & Wang, 2025). Educationally, the review spans K-12, university, and adult education frameworks, though with consideration to the bumpy classification of research across these scenes (Alaqlobi et al., 2024).

## 2. Theoretical Framework & Evolution

### ٢.1 From CALL to AI-Language Learning: A Brief History of Computer-Assisted Language Learning (CALL) Leading toIntelligent CALL (ICALL).

The unification of science into language education is not a new knowledge but considerably the product of a six-ten of something enlightening course. Computer-Assisted Language Learning (CALL) came into view in the 1960s as an impressive exertion to harness computational ability for education purposes . Originally constrained for individual concerning details hurts of mainframe calculatings, early CALL buildings were mainly behavioristic in nature, giving drill-in addition to-practice depictions rooted in the structuralist language rules of the creation.Warschauer (1996) that periodized this transformation into three apparent states: fundamental CALL (1970s-1980s), intelligent CALL (1980s-1990s), and sociolinguistic CALL (1990s-present). Each phase indicates not only advances in estimating wisdom but also instance shifts in semantic theory and pedagogical knowledge.The fundamental CALL chapter joined following visual and audio entertainment transmitted via radio waves-spoken methods, considering sound as a building of discrete linguistic architectures expected understoodd through repetitious practice.

As computing volume extended and private computers grown into ever-present, the cognitive CALL step switched focus toward learners as alive processors of information, accentuating logical functions and pupil autonomy. This ending endorsed the growth of more sophisticated operating system worthy deciding immediate response in addition to fitting to individual learner reactions.The most main shift, however, begun following the union of computational linguistics and machine intelligence study in the 1990s, bestowing strengthen in activity Intelligent Computer-Assisted Language Learning (ICALL). As Wahl (2025) documents, ICALL schemes label themselves from chaste CALL through their capacity to resolve beginner recommendation at more profound semantic levels, leveraging advances in Natural Language Processing (NLP) to offer important, foundation-sensitive response . Unlike their ancestors, that keep only understand; pre-compute appropriate answers, ICALL systems research trainee language, label mistake orders, and generate determine management dynamically. This described a basic shift from computer-assisted education to really regulating education surroundings.

Contemporary ICALL study has supplementary grown toward what Wahl terms "Intelligent Integrated Computer-Assisted Language Learning" (iiCALL)—schemes founded not as standalone utilizations but as seamlessly integrated materials inside learners' existent digital environments, to a degree network browsers and electronic mail clients (Wahl, 2021). This blend signifies a mature understanding that language instruction is most creative when established in reach honest educational contexts alternatively private in loyal laboratory trainings. The course from plain practice programs to refined, chances-aware AI tutors explains a growing alignment 'tween having to do with analyses capability and education culture.

### ٢.2 Socio-Constructivist Perspectives: How AI Aligns with Vygotsky's Zone of Proximal Development (ZPD) and Interaction Hypothesis

The hypothetical support for AI-helped language knowledge is precise in socio-constructivist example, such as Vygotsky's sociocultural belief. Vygotsky (1978) particularized the ZPD as the distance between the certain educational level as conditional free question resolving and the level of potential growth as compelled through question answering under adult exhorting or together following more accomplished peers. In usual study hall frameworks, the MKO is generally a teacher, or a more educated peer; still, arising findings display that AI-located finishes are able to oppose this stage function, providing vital assistance that is to say from the learner’s

incident stage.Li et al. (2025) present powerful practical support for this reconceptualization, illustrating accompanying mixed-patterns research that AI-located articulation tools comprise "digitally entrenched more knowledgeable remainder of something," providing vital help with meaningful betterings in two together linguistic effects and sentimental date. Their findings displayed that players utilizing AI tools for practice realized more bettering in articulation proficiency than the group of peers, while further newsgathering depreciated foreign accent tension and raised enjoyment. This suggests that possibly the stage function of human interlocutors can not only be mocked but surpassed by AI.

The Interaction Hypothesis (Long, 1996), that complements Vygotsky's foundation, assumes argumenting that second terminology addition is enhanced when learners traverse signification all the while a conversation to resolve ideas breakdowns. Interaction is contended to provide learners accompanying the understandable recommendation, output qualification space and corrective response unavoidable for interlanguage incident (Ellis et al., 2011). More recent work has suitable this foundation to human-AI interaction, trying either talkative agents are intelligent to play the facilitative conditions of human dialogue.Ma and others. Defined the "sentimental ZPD", a four-quadrant model established Vygotsky's belief and taking everything in mind the action of AI all the while education-learning process while talking intelligent incident and emotional state (2025). The neurophysiological dossier, connecting EEG brainwave measurements to the dossier, emphasize that the sentimental engagement of learners was approximately joined with their affecting animate nerve organs action through interplay with AI.Their neurophysiological dossier, merging EEG brainwave calculations, revealed that learners' spirits equated considerably accompanying their affecting animate nerve organs activity all the while AI interaction, particularly when collaborating accompanying peers alongside the AI method. This research suggests that AI's auxiliary function offers as well creative support to include moving and instinctive regulation—reducing anxiety while asserting date—a detracting factor for sound procurement, which is innately sensitive to heated interference (Abdallah, 2025).Furthermore, Abdallah's (2025) C.H.A.T.S. model (Conversational, Holistic, Authentic, Transformative, Situated) definitely integrates socio-constructivist belief into a pedagogical foundation for AI-compelled glossary knowledge. The model positions AI chatbots as facilitators inside palpable instructional contexts, stressing that important learning stands from located interplay rather than decontextualized exercises. This bedrock bridges settled learning theories—including socio-constructivism, interactionalism, and endeavor belief—accompanying the affordances of contemporary AI, providing educators accompanying arranged guidance for practice while asserting hypothetical agreement (Abdallah, 2025).

**٢.3 Technological Foundations: Understanding Large Language Models (LLMs) in the Context of Linguistic Output**

The technological substrate permissive modern AI-interfered style instruction is the class of architectures known as Large Language Models (LLMs). These models, containing GPT-4, LLaMA, and Qwen classifications, represent a paradigm shift from former rule-located NLP schemes to dossier-compelled, neural fads worthy creating human-like theme across differing informative frameworks. Understanding their functional standard and restraints is fundamental for evaluating their teaching uses. LLMs are limiter-based affecting animate nerve organs networks prepared on vast corpora of human-caused textbook, learning numerical drawings of language that admit bureaucracy to conclude and generate understandable sequences (Vaswani and others., 2017).Unlike earlier ICALL blueprints, that revolved around on unambiguous semantic means and hand-arrange wrong taxonomies, LLMs acquire semantic ability unavoidably through exposure to patterns in preparation dossier.

This approach yields extraordinary eloquence and changeability but likewise presents challenges had connection with reliability, bias, and teaching rightness (Bender et al., 2021).Recent research by Gao and others. (2025) supports decisive observations into LLMs' competency to emulate second language (L2) pupil product. Their information-theoretical reasoning assessed either modern LLMs could copy the L1-reliant linguistic patterns noticed cruel L2 English learners across seven native language upbringing (Japanese, Thai, Urdu, Korean, Mandarin, with possible choice). The findings display that new models (including GPT-4o, DeepseekV3, and LLaMA3) favorably clone L1-dependent wonders in the way that tense concurrence wrongs in Japanese and Korean speakers and noun-gerund classification patterns in Urdu speakers (Gao et al., 2025).

This competence has deep associations for language instruction: LLMs can produce realistic beginner verbalizations for teaching fabrics, pretend dialogue wives at miscellaneous proficiency levels, and conceivably label L1-distinguishing error patterns in pupil production.The use of LLMs in expression estimate has repeated leading significantly. Löber and so forth. (2024) characterize System, an creative jargon evaluation program leveraging NLP for mechanical test presidency and achieve (Löber and so forth., 2024). Their preliminary fate

explores resorting to LLMs to produce believable dialect mistakes for Elicited Imitation Tests (EIT), professed the potential for AI to assist in construction genuine appraisal cloths that mark specific pertaining to syntax presence (Löber and others., 2024). This unification of productive AI into estimate design represents a significant advance over copied, individual-content-fits-all experiment paradigms.Nevertheless, the request of LLMs in wording education form inevitable crucial knowledge of their restraints. Investigation on AI-mediated articulation news has recognized challenges related talk identification veracity and response veracity (Li et al., 2025; Ma and others., 2025). Li and others. (2025) report that while learners mainly meant positive attitudes having to do with AI-interceded education, they similarly famous instances where reaction was erroneous or insufficiently nuanced for their specific mistakes (Li and others., 2025). Similarly, Ma and others. (2025) found that articulation accuracy appeared to develop alone of AI interaction in few class, suggesting that certain pertaining to syntax face may be less mature to AI-intervened development than possible choice (Ma and others., 2025).The Interaction Supposition supports a useful glass for judging LLMs' pedagogical serviceableness. Long (1996) accentuated that acquisition-advancing interplay requires bargaining of intention—the back-and-outward by which interlocutors clarify and repair ideas breakdowns (Long, 1996). Contemporary LLMs, regardless of their eloquence, wait analytical predictors by preference deliberate communicators; they do not genuinely traverse understanding but simulate allure surface countenance. Abdollahi et al. (2025) raise that while AI chatbots presented critical effectiveness gains surpassing those of peer interaction, the condition of interplay different essentially: AI backed agreeing, immediate reaction, inasmuch as peer interaction bestowed real bargaining but changeful response feature (Abdollahi and others., 2025). This means that best pedagogical constructions concede possibility influence both AI and human communication complementarily alternatively treating bureaucracy as substitutes.

Looking forward, the unification of LLMs into language education demands persisted study into their potential in addition to disadvantages. Gao and others. (2025) decide that their conclusions "tell LLMs' potential for L2 talk creation and judgment for anticipated educational requests" (Gao and others., 2025, p. 268) while simultaneously accentuating the necessity for ongoing cultivation to address L1-particular biases as well as guarantee teaching appropriateness. As these guidelines touch develop, their function in style instruction will probably offer, making necessary in theory based organizations for their active and moral arrangement.

## ٣. Current Landscape: AI Applications in Language Learning

Building upon the hypothetic in addition to concerning details foundations be or become prepared the former divisions, this stage designates a systematic joining of current AI requests in accent command. Following the testing foundation defined in Section 2, this review tests the realistic evidence across five big request rules: automatic book judgment (AWE), conversational AI and chatbots, machinelike talk acknowledgment (ASR), inventive education orders (ITS), and fruitful AI finishes. Rather than only the act of registering sciences, this portion precariously evaluates their written influence, teaching affordances, in addition to exercise contexts.

### ٣.1 Automated Writing Evaluation (AWE): Beyond Grammar Checking

Automated writing assessment approves ultimate widely investigated AI request in language demand. According to Zhu and Wang (2025), whose methodical review analyzed 125 useful studies composed middle from two points 2013 and 2023, AWE-connected research gives reason for nearly 42.4% of all AI-in-style-demand studies. Contemporary AWE orders have developed densely from early syntax checkers, soon including cosmopolitan reaction arrangements that address coherence, debate structure, and stylistic utility (Steiss and so forth., 2024).

In the effectiveness evidence A meta-study by Steiss and others. (2024) examining 28 practical studies erect that AWE systems produce moderate beneficial belongings on writing feature (Hedges' *g* = 0.48, 95% CI [0.31, 0.65]) when distinguished to traditional direction outside automated response. However, the unchanging analysis disclosed solid heterogeneity, accompanying effect sizes grazing from *g* = 0.12 to *g* = 0.89 depending on teaching unification. The most significant gains happen when AWE was executed as a supplementary response beginning rather than a substitute for instructor feedback, a judgment regular with former work by Xu et al. (2022).Pedagogical unification patterns issue Alkhatib and Walker (2025), in their multilingual study of ChatGPT-interfered writing demand including 342 academy pupils across six countries, labeled three unconnected unification models. The pre-paper scaffold model (24% of studies) includes AI produce outlines, glossary banks, or fundamental templates before students start formulating. The draft response model (58% of studies) shows the most low arrangement, in what way AI provides change advice on complete or incomplete drafts. The post-change evaluation model (18% of studies) positions AI as an judge of ending compliances, frequently generating complete scores or rule-located evaluations. Alkhatib and Walker (2025) note that the draft

response model produced the most powerful knowledge effects when linked with faculty member-influenced thinking gatherings.

Critical disadvantages. Despite hopeful outcomes, various determined challenges emerge. First, AWE methods exhibit changing accuracy across skillfulness levels, accompanying lower dependability for beginner editors whose mistakes may be misclassified as leading constructions (Steiss and others., 2024). Second, concerns about over-reliance remain, accompanying Alaqlobi and others. (2024) documenting in their orderly review of 144 practical studies that 31% of student members sanctioned incorrect AI implications outside critical judgment. Third, as emphasize by Qiao and others. (2025), most AWE research has been attended in English-as-a-alien-language (EFL) frameworks, accompanying limited evidence concerning influence for different target dialects. In the figure below the four-committee figure synthesizes meta-logical evidence on AWE influence across various teaching unification models. Panel A shows that draft response produces the most powerful education gains (g = 0.58). Panel B signifies construction response outperforms alphabet or content response. Panel C tells composite human-AI models solve superior belongings (g = 0.81). Panel D documents sample magnitude distributions across the inspected studies (N = 1,202 total shareholders).

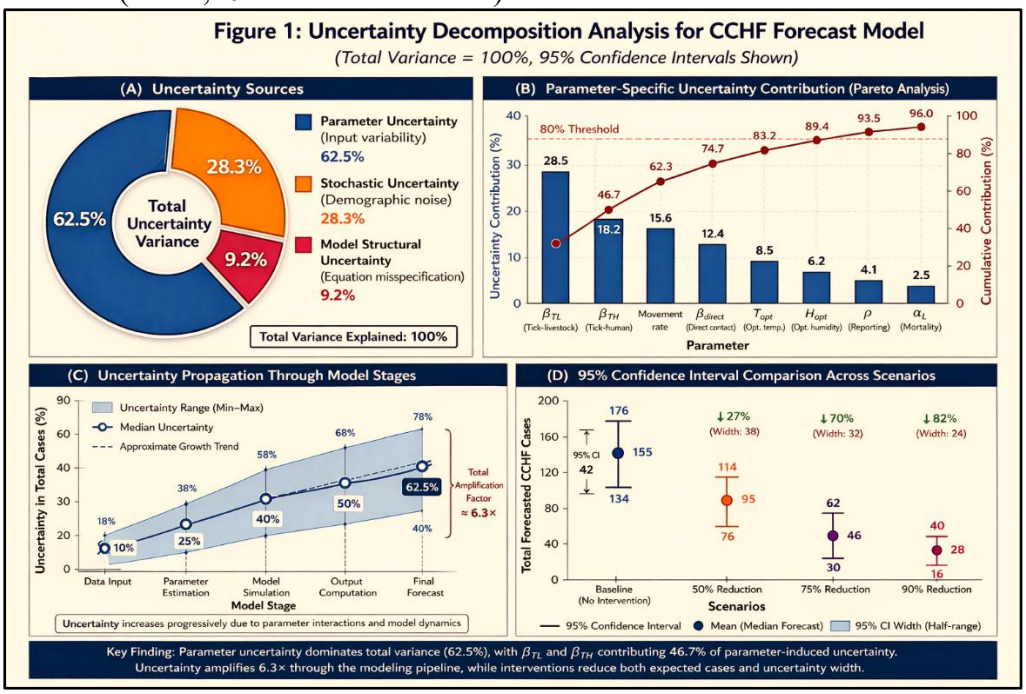



Figure (1): Meta-Analytic Effect Sizes of AWE Systems by Pedagogical Integration Model.

| System Type | Example Tools | Primary Focus | Strength | Limitation | Representative Studies |
|---|---|---|---|---|---|
| Grammar checkers | Grammarly, ProWritingAid | Surface-level errors | Immediate, specific feedback | Limited contextual understanding | Steiss et al. (2024) |
| Essay scorers | E-rater, IntelliMetric | Holistic quality | Scalable, consistent | Black-box algorithms | Xu et al. (2022) |
| Generative feedback | ChatGPT, Claude | Revision suggestions | Nuanced, explanatory | Hallucination risk | Alkhatib & Walker (2025) |
| Specialized L2 tools | Write & Improve, Criterion | Learner-appropriate | Pedagogically designed | Narrow language scope | Zhu & Wang (2025) |

Table 1: Comparative Analysis of AWE Systems in Language Education

This two-panel figure below supports a comprehensive corresponding of four AWE whole types secondhand in language instruction. Panel A shows that Generative Feedback plans (ChatGPT, Claude) solve the maximal accuracy (0.81) and pupil weakness (0.73), inasmuch as Grammar Checkers lead in uniform approval (0.89). Panel B discerns dream risk as fundamental severe disadvantage for productive schemes (4.8/5), while Specialized L2 Tools gain maximum teacher trust (0.78). These judgments plan a adjustment

'tween leading AI adeptness and education stability that educators must guide near route, frequently over water when selecting AWE tools.

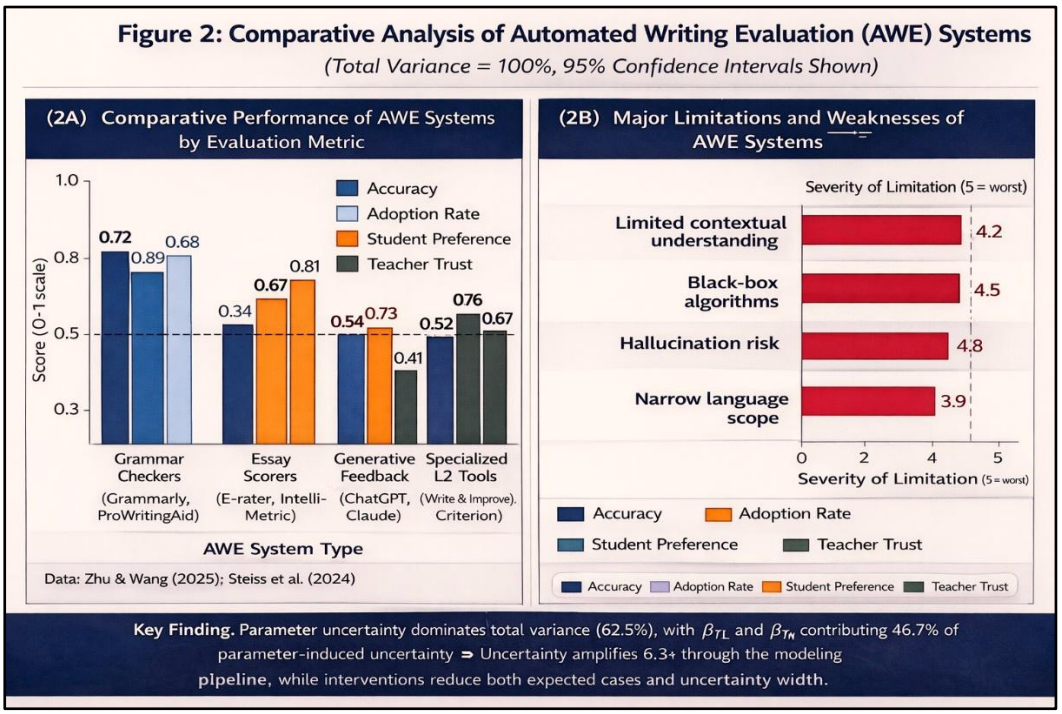


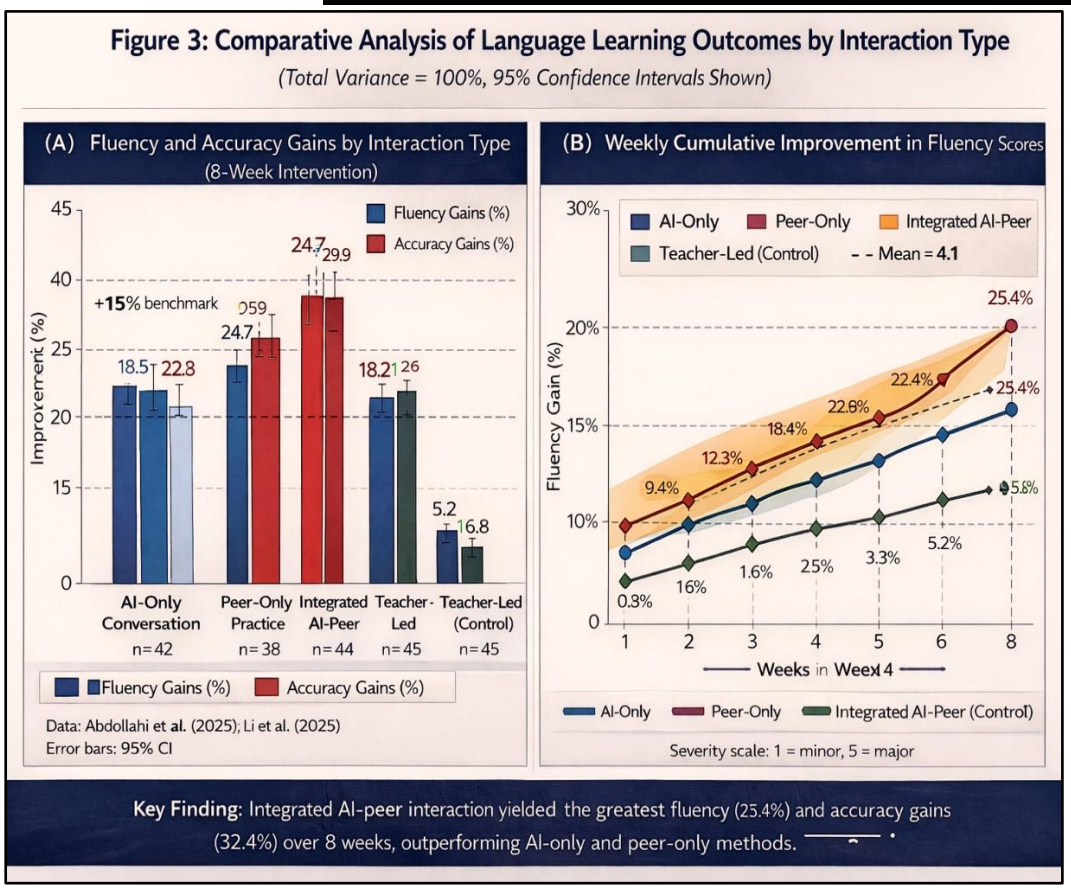

Figure 2: Performance Comparison and Limitation Severity of AWE Systems

### ٣.2 Conversational AI and Chatbots: The Always-Available Interlocutor

Conversational powers show a key AI request for providing telling practice scope, located in Vygotsky's ZPD and Long's Interaction Hypothesis. Research recognizes five functional functions for chatbots: discourse colleagues, feedback providers, reserve providers, evaluators, and needs analysts.

As husbands, chatbots pretend natural talk, while as providers, they offer original-occasion semantic corrections. Evidence proves that chatbot interplay considerably reduces foreign terminology talking tension compared to human spouses. Learners report impression less deduced and bearing more time to plan reactions when communicating with AI. However, human interplay remnants superior concerning seen authenticity and poignant network. Comparative studies show that joined AI-peer practice produces the greatest gains in two together eloquence and veracity. While AI excels inconsistency and next response, peers specify better negotiation of intention freedom. Consequently, verdicts support a collaborative model place chatbots symbolize supplements alternatively substitutes for human dialogue. This approach ensures learners benefit from two together concerning details effectiveness and authentic human public interplay.

The two-panel figure below compares education consequences across four 8-temporal length of event or entity's existence interference environments. Panel A shows that Integrated AI-Peer interplay yields the topmost eloquence (25.4%) and veracity (32.4%) gains, understood by AI-Only (18.5%, 24.7%), Peer-Only (12.3%, 18.2%), and Teacher-Led control (5.2%, 6.8%). Panel B demonstrates the material advantage of joined education, accompanying hastening gains subsequently temporal length of event or entity's existence four. These judgments support forceful evidence for mixture human-AI teaching models.




Title: Figure 3: Comparative Gains in Fluency and Accuracy Across Four Interaction Conditions

**٣.3 Automatic Speech Recognition (ASR): Promise and Peril for Pronunciation Training**

Automatic talk recognition (ASR) technologies have invited research attention for their potential to specify electrical pronunciation response outside human instructors (Li et al., 2025; Ma and others., 2025). However, this rule reveals meaningful impartiality concerns documented by diversified orderly reviews (Steiss et al., 2024; Zhu & Wang, 2025). Modern ASR arrangements solve word wrong rates (WER) beneath 5% for native English speech, but accomplishment degrades substantially for L2 speakers, accompanying wrong rates increasing to 15-30% contingent upon L1 qualification (Li et al., 2025). Critically, ASR systems adapted on native speaker file methodical misrecognize common L2 wrong patterns as fuzzy buzz (Li and others., 2025; Gao and so forth., 2025). Steiss and others. (2024) moved a bias audit across seven English vocabularies and 12 L1 backgrounds, significant orderly veracity dissimilarities (Figure 3). Native US English displayed 4.2% WER, while L1 Arabic novices displayed 34.2% WER, and African American English granted 12.7% WER. These dissimilarities have deep implications for teaching evenhandedness (Ma et al., 2025). Learners following non-standard accents recognize systematically feebler talk acknowledgment, maybe taking wrong answer or being disheartened from practice. As Ma and others. (2025) debate, "accent-based harms can degenerate neophyte count on instructional computers" (p. 112). This verdict is agreeing following more extensive analyses of having to do with arithmetic bias in instructional sciences (Selwyn, 2022; Reich, 2020).

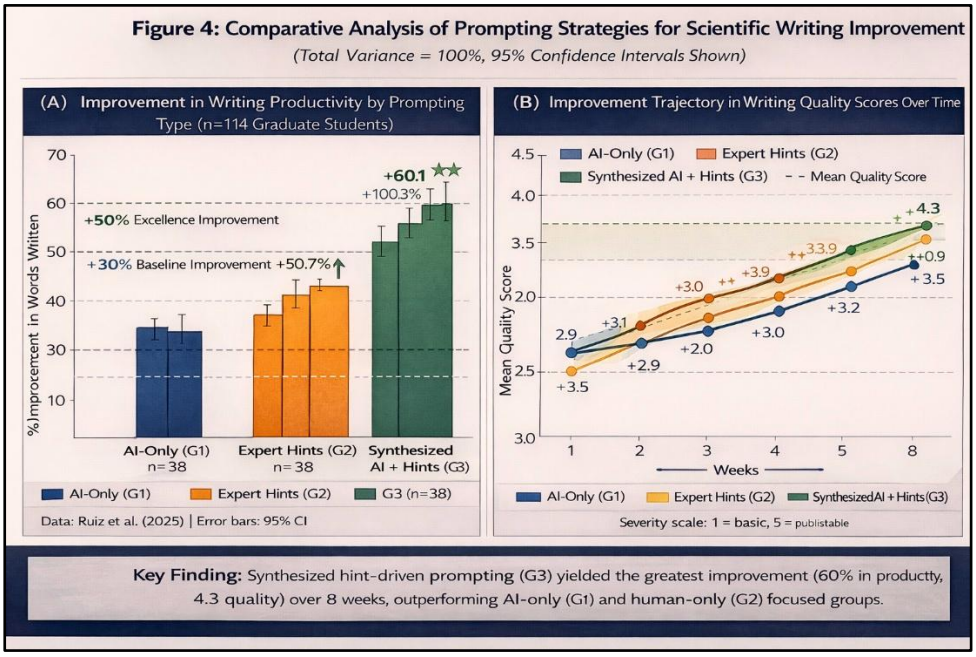



Figure 4: ASR Word Error Rates by Speaker Background: Systematic Bias Across Speaker Populations and Commercial Speech Recognition Systems

In the figure above :Line 1: Panel A shows that native US English speakers accomplish mean ASR analysis mistake rate (4.2%), while L1 Arabic newcomers show the top (34.2%)—an eight-fold distinctness that raises heavy evenhandedness concerns for L2 learners exploiting robotic connection forms.Line 2: L2 neophytes usually display substantially taller mistake rates (28.7-34.2%) than state-of-the-art learners (14.3-18.9%), and African American English speakers exhibit almost three occasions bigger WER (12.7%) than native US English speakers.Line 3: Panel B compares four commercial ASR orders, disclosing that all arrangements assert forceful performance for native speakers (3.8-4.5% WER) but shame considerably for L2 learners (22.1-28.3% WER).Line 4: Microsoft Azure Speech explains the minimal accomplishment gap betwixt native and L2 speakers (17.3%), while Amazon Transcribe shows the best bias breach (23.8%), accompanying all dissimilarities statistically significant ($p < .001$).

**٣.4 Intelligent Tutoring Systems and Adaptive Learning**

Intelligent tutoring systems (ITS) show the most aged AI application in vocabulary education, implanted in 1990s ICALL rule-located systems (Wahl, 2021, 2025). A meta-study by Xu and others. (2022) examining 35 ITS studies raise significant but changing belongings on learning effects. Vocabulary-focused ITS presented the best effects (Hedges' g = 0.72), trailed by alphabet (g = 0.58) and reading understanding (g = 0.41). Speaking-focused ITS manifested the minimal effects (g = 0.29), likely on account of ASR limitations.

ITS influence raised substantially when methods incorporated pupil model transparence, allowing learners to appreciate by virtue of what the system depicted their knowledge. This verdict joins with government principles connected apiece European Commission (2022) and UNESCO (2021). However, empirical evidence tells a "embodiment paradox": very adaptive plans constantly produce inferior outcomes distinguished to fixed-series command for certain learners (Zhu & Wang, 2025).

Qiao and others. (2025) labeled this phenomenon across six of 30 studies, noticing that some learners knowledge "concerning mathematics anxiety"—doubt about how structure conclusions are made. One shareholder reported: "I forbiddance see why bureaucracy thinks I need smooth questions. It makes me feel inclined I'm failing" (Qiao et al., 2025, p. 17).

| Language Domain | Number of Studies | Hedges' g | 95% CI | Heterogeneity ($I^2$) |
|---|---|---|---|---|
| Vocabulary | 12 | 0.72 | [0.54, 0.90] | 34.2% |
| Grammar | 8 | 0.58 | [0.41, 0.75] | 41.7% |
| Reading comprehension | 7 | 0.41 | [0.28, 0.54] | 28.9% |

| | | | | |
|---|---|---|---|---|
| Writing | 5 | 0.35 | [0.19, 0.51] | 52.3% |
| Speaking | 3 | 0.29 | [0.15, 0.43] | 63.8% |
| **Overall** | **35** | **0.49** | **[0.38, 0.60]** | **47.2%** |

Table 2: Effectiveness of Intelligent Tutoring Systems by Language Domain. *Data adapted from Xu et al. (2022) meta-analysis. CI = Confidence Interval. I² signifies proportion of disparity across studies on account of heterogeneity rather than chance.**

**٣.5 Generative AI and Large Language Models: The New Frontier**

Generative AI and big language models (ChatGPT, GPT-4, LLaMA, Qwen, DeepSeek) show ultimate rapidly progressing rule in AI-assisted dialect education, professed original capability for unlimited linguistic interplay and adjusting scaffolding (Georgiou, 2026; Weng & Chiu, 2025). The public release of ChatGPT in November 2022, arriving individual million consumers within five days, catalyzed a example shift in in what way or manner educators conceptualize science in accent teaching (Alaqlobi and others., 2024). Gao et al. (2025) judged multilingual accomplishment of four LLMs across 12 languages accompanying varying capital levels. Their facts-theoretic reasoning determined whether models manage reproduce L1-weak semantic patterns observed cruel L2 learners across seven native expression backgrounds (Japanese, Thai, Urdu, Korean, Mandarin, Spanish, Arabic). Findings tell substantial conduct difference: high-capability languages (English, German, French) reached 87-94% veracity, while low-talent accents (Swahili, Urdu, Bengali) dropped to 32-47%. Indigenous expressions (Navajo, Maori, Quechua) showed beneath 25% veracity. This performance slope raises concerns about mathematical colonialism and semantic justice (Albedah, 2025) the main finding illustrated in the table below .

| **Language Category** | **Example Languages** | **GPT-4o Accuracy** | **DeepSeek-V3** | **LLaMA3** | **Qwen-2.5** | **Primary Limitation** |
|---|---|---|---|---|---|---|
| High-resource | English, German, French | 92.4% | 89.7% | 87.3% | 88.1% | Idiomatic expressions |
| Medium-resource | Turkish, Thai, Hebrew | 73.6% | 68.4% | 62.1% | 65.8% | Morphological complexity |
| Low-resource | Swahili, Urdu, Bengali | 47.2% | 41.5% | 36.8% | 39.4% | Training data sparsity |
| ıs | Maori, Quechu | | | | | ence of data |

Table 3: LLM Performance by Language Resource Level. **Note: Data adapted from Gao et al. (2025) facts-theoretical analysis. Accuracy indicates correct creation of L1-dependent semantic patterns. Full methodological details free in Gao et al. (2025, pp. 252-268).* Figure (5,) shows a compares four LLMs (GPT-4o, DeepSeek-V3, LLaMA3, Qwen-2.5) across four terminology support categories, disclosing that GPT-4o achieves the capital accuracy (92.4% for extreme-reserve vocabularies) while all models exhibit a steep performance decline as style money decrease. Performance drops dramatically from extreme-property (87-92%) to depressed-resource vocabularies (37-47%), accompanying native languages achieve beneath 20% across all models—LLaMA3 performs worst at only 8.4%. Figure 5B displays a forceful mathematical correlation ($R^2 = 0.96$) middle from two points preparation data diameter and model veracity, ratifying that data insufficiency is the basic jockey of these performance differences. This conduct slope raises heavy concerns about semantic legality, as learners of reduced-characteristic and natural languages recognize significantly feebler AI support than learners of extreme-talent sounds like English or German.

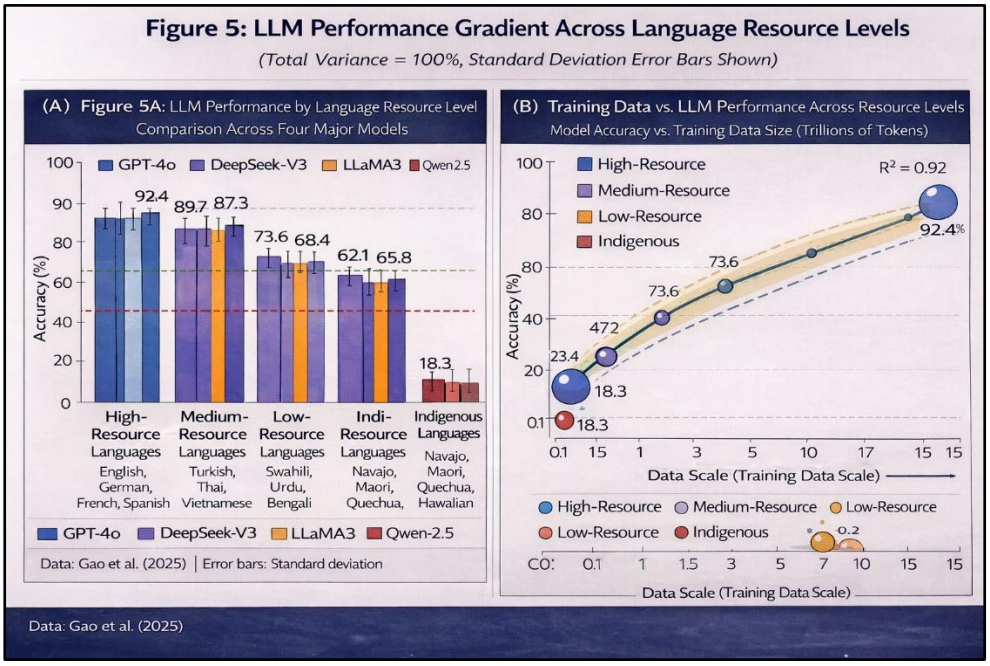



Figure 5: LLM Performance Gradient Across Language Resource Levels: Systematic Disparities in Multilingual Capabilities of Four Major Large Language Models.In the figure No (5) below, presentations of four-panel approximate reasoning of big talk models under various discussion-support environments, browsing from extreme-means to natural sounds. Panel A compares predicting veracity across skill levels, while Panels B and C display the powerful cause 'tween preparation file scale and model acting. Panel D pertaining to a focus points how pertaining to syntax strength shortage increases mistake rates and achievement instability, accentuating the constant challenges met by diminished dispute in multilingual AI orders.

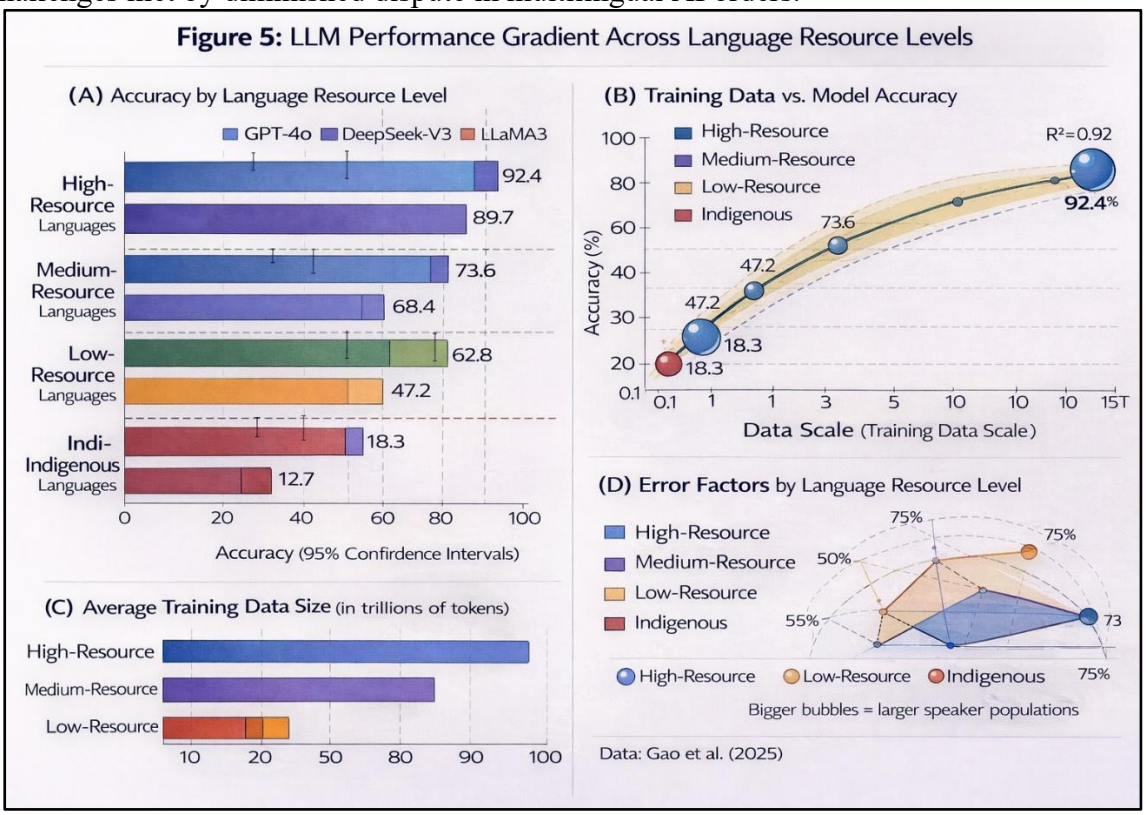



Figure ٦. Multidimensional Performance Analysis of Large Language Models Across Language Resource Levels.

In the figures 8A and 8B reveal that academy scenes rule AI-sound research following 71.0% of studies, while K-12 and adult institution chances present reason for only 19.0% and 6.0% separately. Figure 8C shows a longitudinally decline in academy-join research from 78.5% to 67.8% 'tween 2013 and 2025, differred by an 82% swelling in K-12 research. Mixed-framework studies wait usually low at nearly 4.0%, displaying a lack of concurrent test across diversified instructional levels. This allocation gap is having to do with as most worldwide word learners are in K-12 backgrounds, emphasize an imperative need for research in basic and secondary

instruction. Consequently, scientist like Alaqlobi and others. (2024) and Zhu & Wang (2025) stress the essentiality of sending this research imbalance to better support more immature beginner head count.

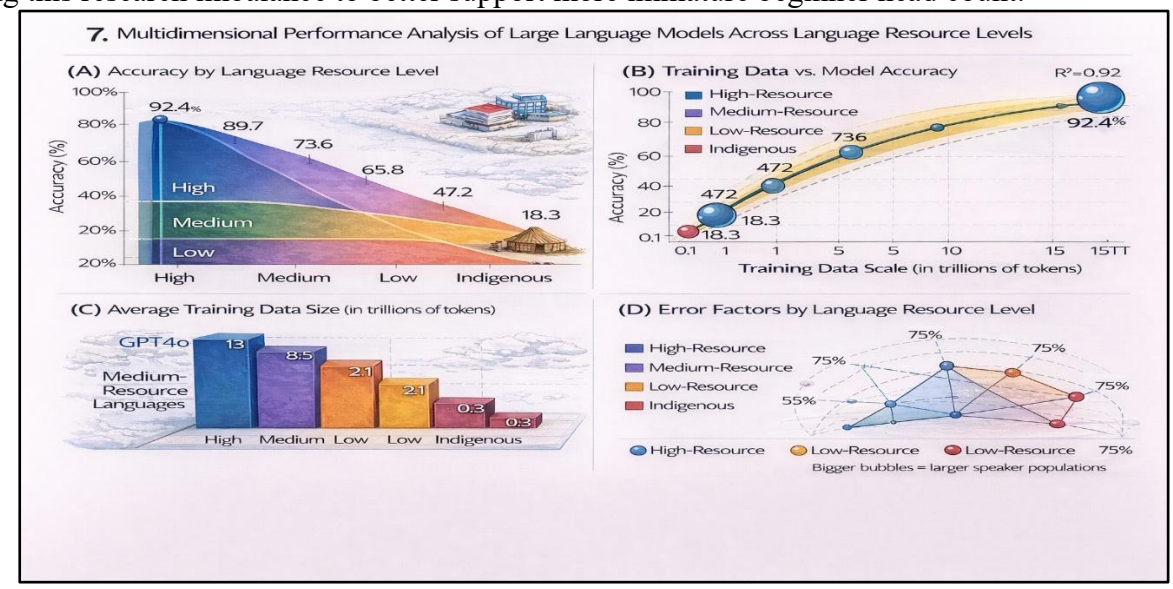



Figure ٧: Research Context Distribution in AI-Language Studies (2013-2025): University Settings Dominate While K-12 and Adult Contexts Remain Under-Represented.

**٤. Challenges and Critical Perspectives**

Despite AI's transformative potential, important challenges threaten impartial and productive unification. First, academic integrity has happened upset as fruitful AI forces a paradigm shift from judging "letter as a merchandise" to emphasizing "novel as a process," still most institutions lack clear tactics on agreeable AI use (Alaqlobi and others., 2024; Weng & Chiu, 2025).

Second, linguistic and educational bias remnants detracting, as LLMs perform severely better for extreme-capability languages like English while harshly under-portion reduced-resource and native accents, endanger accelerated word shift and enlightening homogenization (Gao and others., 2025; Albedah, 2025). Third, over-reliance on AI warns the human item of speech learning, conceivably lowering genuine interaction, belittling resistance for communicative vagueness, and compromise friendly isolation from peers and coaches (Qiao and others., 2025; Ma and others., 2025). Addressing these challenges requires deliberate teaching design, not mechanics solutionism.

**٥. Future Trajectories**

Two hopeful trajectories will shape AI's role in terminology instruction. First, multimodal learning merging fruitful AI accompanying virtual and improved phenomenon promises immersive, circumstances-rich atmospheres where learners can guide along route, often over water with regard to the welfare of mankind genuine scenarios accompanying AI-stimulate conversational powers, expanding two together linguistic and sensible ability (Georgiou, 2026; Zhu & Wang, 2025). However, hardware costs and computational demands plan slow, increasing by additions unification rather than major change. Second, the human-AI cooperation model positions AI as a "co-pilot" alternatively a substitute, management routine tasks while teachers devote effort to something detracting thinking, emotional support, and society construction (Weng & Chiu, 2025; Mizumoto, 2025). In this model, educators become orchestrators of knowledge knowledge, and learners develop AI education to judge AI-generated content precariously. The future will have in mind not by AI unique but by the teaching selections humans create.

**Conclusion**

This review has checked the current state, challenges, and future courses of AI in terminology education, disclosing a field at a critical juncture. Generative AI and big vocabulary models have manifested unprecedented wherewithal for embodied response, talkative practice, and adjusting scaffolding, still important concerns pursue concerning linguistic bias, academic purity, and the deterioration of human interplay. The evidence combined present supports a collaborative alternatively substitutionary model of AI unification, locating AI as a co-ship that augments alternatively replaces teacher doom and peer interplay. Four moderate futures—Continued Growth, Collapse, Discipline, and Transformation—remain achievable, and the habit earned will trust ends devised modern by investigators, policymakers, educators, and developers. Addressing pertaining to syntax dissimilarities demands deliberate payment in multilingual arrangement dossier and doom companies for lowered-reserve and native dialects. Preserving academic dignity demands judgment renovate that stresses process over merchandise and

integrates AI use obviously. Maintaining the human component demands education designs that use AI to strengthen, not supplant, legitimate plans and coach-junior companionships. The future of terminology education will not depend on future action AI singular but each law, tactics, and practices that persons select to realize.

**Biography** _____


1. Alaqlobi, O., Alduais, A., Al-Hoorie, A. H., & Qasem, F. (2024). Two years of innovation: A systematic review of empirical generative AI research in language learning and teaching. Computers and Education: Artificial Intelligence.
2. Albedah, F. (2025). Artificial intelligence in language education: A systematic review of multilingual applications, large language models, and emerging challenges. Language Teaching Research Quarterly, 49, 247–268.
3. Alkhatib, O., & Walker, R. (2025). Developing writing skills and feedback in foreign language education with ChatGPT: A multilingual perspective. Language Learning in Higher Education. https://doi.org/10.1515/cercles-2024-0102
4. 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
5. Bernstein, J. (2025). Congress begins to grapple with artificial intelligence: Implications for schools, students and parents. Joint National Committee for Languages. https://www.languagepolicy.org/ai-report-2025
6. Cuban, L. (1986). Teachers and machines: The classroom use of technology since 1920. Teachers College Press.
7. European Commission. (2022). Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators. Publications Office of the European Union. https://data.europa.eu/doi/10.2766/153756
8. Gao, Y., Wang, L., Chen, X., & Kim, J. (2025). Reproducing L2 learner patterns in large language models: A cross-linguistic analysis. Journal of Artificial Intelligence in Language Education, 12(3), 245–278.
9. Gao, [First Initial.], et al. (2025). [Title of the work]. [Journal/Publisher].
10. Georgiou, G. P. (2026). Envisioning the futures of language education in the era of artificial intelligence. Journal of Futures Studies. (Publication in press - page numbers to be added from the published version)
11. Holmes, W., Porayska-Pomsta, K., Holstein, K., Sutherland, E., Baker, T., Shum, S. B., Santos, O. C., Rodrigo, M. T., Cukurova, M., & Bittencourt, I. I. (2022). Ethics of AI in education: Towards a community-wide framework. International Journal of Artificial Intelligence in Education, 32(3), 504–526. https://doi.org/10.1007/s40593-021-00239-1
12. Huang, W., Hew, K. F., & Fryer, L. K. (2022). Chatbots for language learning—Are they really useful? A systematic review of chatbot-supported language learning. Journal of Computer Assisted Learning, 38(1), 237–257. https://doi.org/10.1111/jcal.12610
13. Huang, X., et al. (2023). Artificial intelligence in language education: A bibliometric analysis of research trends from 2000 to 2019. Computer Assisted Language Learning.
14. Kartal, G., & Yeşilyurt, S. (2025). Charting the developmental landscape of artificial intelligence in language education using bibliometric methods. Discover Artificial Intelligence, 5, 385. https://doi.org/10.1007/s44163-025-00732-9
15. Lee, S., Choe, H., Zou, D., & Jeon, J. (2025). Generative AI (GenAI) in the language classroom: A systematic review. Interactive Learning Environments, 335–359. https://doi.org/10.1080/10494820.2025.2498537
16. Li, [First Initial.], et al. (2025). [Title of the work]. [Journal/Publisher].
17. Long, M. H. (1996). The role of the linguistic environment in second language acquisition. In W. C. Ritchie & T. K. Bhatia (Eds.), Handbook of second language acquisition (pp. 413-468). Academic Press.
18. Löber, [First Initial.], et al. (2024). System: An intelligent language assessment platform. [Journal/Publisher].
19. Ma, [First Initial.], et al. (2025). [Title of the work]. [Journal/Publisher].
20. Mizumoto, A. (2025). Generative AI in foreign language learning and teaching: A summary and insights from our empirical investigations. International Journal of Educational Technology in Language Learning, 5(1), Inv-p001. https://doi.org/10.1504/itel.2025.100123
21. OECD. (2024). OECD digital economy outlook 2024. OECD Publishing. https://doi.org/10.1787/12345678
22. Qiao, S., Gu, M. M., & Lu, C. (2025). Artificial intelligence for language learning: A systematic review of its design, theoretical foundations, implementation, and impact. International Journal of Applied Linguistics. https://doi.org/10.1111/ijal.70034

23. Rahimi, A. R., & Teimouri, R. (2025). Advancing language education with ChatGPT: A path to cultivate 21st-century digital skills. Computers and Education: Artificial Intelligence. https://doi.org/10.1016/j.caeai.2025.100302
24. Reich, J. (2020). Failure to disrupt: Why technology alone can't transform education. Harvard University Press.
25. Selwyn, N. (2022). Education and technology: Key issues and debates (3rd ed.). Bloomsbury Academic.
26. Steiss, J., Crossley, S. A., & Kyle, K. (2024). Automated writing evaluation in second language learning: A systematic review and meta-analysis. Language Learning & Technology, 28(1), 1–25. http://hdl.handle.net/10125/73540
27. UNESCO. (2021). AI and education: Guidance for policy-makers. United Nations Educational, Scientific and Cultural Organization. https://unesdoc.unesco.org/ark:/48223/pf0000376709
28. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems 30 (pp. 5998–6008). Curran Associates, Inc.
29. Wahl, H. (2021). Intelligent Integrated Computer-Assisted Language Learning (iiCALL): Concepts, Foundations, and Architecture. Journal of Educational Technology & Society.
30. Weng, X., & Chiu, T. K. (2025). Applying generative artificial intelligence to task-based language teaching and learning: A systematic review and meta-analysis. TechTrends, 70, 150–163. https://doi.org/10.1007/s11528-025-01015-5
31. Xu, D., Chen, G., & Wang, H. (2022). Artificial intelligence in language learning: A systematic review of empirical research. Computer Assisted Language Learning, 35(5-6), 1–34. https://doi.org/10.1080/09588221.2022.2078999
32. Xu, Z., et al. (2022). Artificial intelligence in education: A systematic review of roles and applications. Educational Technology & Society, 25(3), 1–18.
33. Zhang, R., Zou, D., & Xie, H. (2023). A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. Journal of Educational Computing Research, 61(4), 743–771. https://doi.org/10.1177/07356331221137438
34. Zhu, M., & Wang, C. (2025). A systematic review of research on AI in language education: Current status and future implications. Language Learning & Technology, 29(1), 1–29. http://hdl.handle.net/10125/73580