Article

Artificial Intelligence in English Language Learning: A Bibliometric Analysis and Synthesis of Learning-Efficacy Evidence

HUANG Fang [1,*], WU Qiang [2]

[1] Can Tho University, Can Tho, Vietnam; e-mail: huangfang@ctu.edu.vn
[2] National University of Laos, Vientiane, Laos; e-mail: wuqiang@nuol.edu.la
* Correspondence: huangfang@ctu.edu.vn

Abstract

The Journal of Sustainability Insights, vol. 1, no. 1, November 1, 2026. DOI: 10.00000/tjosi. Breakthroughs in generative artificial intelligence and large language models have rapidly expanded the use of AI in English language learning, from intelligent tutoring and automated essay marking to conversational practice, pronunciation assessment, personalised learning and learning analytics. Focusing on how AI is used for English learning, this paper integrates three strands of evidence: bibliometrics, paper-level impact modelling, and published learning-effectiveness findings. Using genuine data from OpenAlex topic queries, we find that annual publications began around 2015 and expanded sharply after 2022, with a Logistic growth inflection around 2025, indicating a still-expanding field. Content analysis of more than one thousand Crossref records identifies six application types, with writing evaluation and intelligent tutoring accounting for the largest shares of classifiable papers. A negative binomial model shows that paper impact is driven mainly by document type, thematic category and year structure, while open access confers no significant citation advantage. An evidence matrix of published meta-analyses indicates considerable short-term gains in vocabulary and receptive skills (pooled effects of roughly $d = 0.39$ to 1.17), whereas evidence for productive skills and communicative competence over longer horizons remains limited and heterogeneous. We conclude with principles for matching technology types to skill types, foregrounding formative feedback, and balancing short-term efficiency against long-term transfer.

Keywords: artificial intelligence; English language learning; bibliometrics; learning effectiveness; application types; negative binomial regression

Received: 23 September 2026
Revised: 6 October 2026 Accepted: 8 October 2026 Published: 1 November 2026 www.tjosi.com

Copyright: © 2026 by the authors. Licensee Lightup Innovation, United Kingdom, HR5 3DJ. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license

1. Introduction

Artificial intelligence has not entered language learning overnight. From intelligent tutoring systems (ITS) in the 1980s, through natural-language-processing-based automated writing evaluation (AWE), to large-language-model-driven conversational practice and instant feedback, successive technological generations have continually rewritten the answer to how far machines can participate in language teaching[1,2]. Especially after the popularisation of generative conversational tools in late 2022, learners can obtain a grammatical explanation, a rewritten sample paragraph or a near-authentic spoken dialogue within seconds, transforming AI from a back-office analytical tool into a front-line learning partner[3,4]. This explosion of application forms has made a systematic account of their types and an examination of their effectiveness an urgent scholarly task.

Unlike the normative question of whether English learning remains necessary, this paper addresses a technical-educational question: in what forms does AI intervene in English learning, how do research volume and scholarly impact evolve across those forms, and to what extent does existing empirical evidence support their effectiveness? This distinction matters. Debates about the necessity of English learning typically appeal to labour-market returns, human-machine complementarity and the linguistic dependence of knowledge production; the present paper engages none of these arguments and focuses solely on the application forms, technology types and learning-effectiveness evidence of the tools themselves. The two questions are complementary but not interchangeable: even if one grants the enduring value of English learning, one still needs to determine which technological route best improves learning efficiency.

Existing literature has approached these issues from several angles. Systematic reviews of AI in education map technology types and ethical challenges[1,5], computer-assisted language learning (CALL) research supplies the theoretical foundations of the interaction and output hypotheses and task-based pedagogy[6,7], and meta-analyses of AWE and dialogue systems yield verifiable pooled effect sizes[8,9]. Yet three bodies of evidence remain disjointed: bibliometric studies mostly stop at macro-level trend description and seldom combine paper-level impact mechanisms with application-type coding; meta-analyses of learning outcomes are usually reported by technology type in isolation, without comparison against research intensity and impact structures; and a comprehensive judgement about which application types are most prevalent and which evidence is strongest still lacks a common data foundation.

This paper integrates these strands within a single analytical framework and addresses four research questions. RQ1: What growth pattern characterises annual publications on AI in English learning, and has the inflection already been passed? RQ2: What structures characterise the distribution of papers by country, institution and document type, and how can application topics be classified? RQ3: What drives paper impact, and do document type, open-access status and thematic category produce observable, directionally consistent effects? RQ4: What patterns do published pooled effect sizes exhibit overall and by technology subcategory, and is evidence strength balanced across vocabulary, receptive and productive skills?

To answer these questions we adopt a reproducible mixed-methods design. Bibliometric data come from two open scholarly databases, OpenAlex and Crossref: the former provides annual publication series for six topic queries and paper-level micro-records, while the latter supplies the fine-grained title, abstract and thematic-coding fields. Owing to the exhaustion of the OpenAlex free quota, one series was reconstructed from genuine Crossref records counted by year, a constraint disclosed in

the text; no data were fabricated. The econometric models comprise, first, Logistic and Gompertz growth curves fitted to annual publications with extrapolation to 2030, and second, a negative binomial regression explaining variation in citation counts. Content analysis codes papers into six application types on the basis of title and abstract keywords. Learning-effectiveness evidence is drawn from published meta-analyses and randomised studies with verifiable pooled effects. The remainder of the paper is organised as follows: Section 2 reviews the technology lineage and builds a technology-application-outcome framework; Section 3 describes data sources, query strings and model specifications; Section 4 reports empirical results; Section 5 discusses implications and limitations; Section 6 offers conclusions and applications.

2. Literature Review and Theoretical Framework

2.1. From CALL to Generative Artificial Intelligence

The role of AI in language learning has evolved continuously from tool to partner. CALL research since the 1980s treated technology as a carrier of linguistic input and practice, with Chapelle's synthesis establishing its functions in providing comprehensible input, promoting interaction and supporting learner autonomy[6]. Warschauer further argued that technology produces substantive gains in language development only when embedded in authentic social interaction[7], while Kern theorised the dialectical relationship between technological mediation and language learning[10]. Typical applications at this stage were rule-based grammar drills and vocabulary exercises; the "intelligence" of the technology was limited, yet the basic orientation towards immediate feedback and personalised practice was already established.

Intelligent tutoring systems represented the first substantive leap. Sleeman and Brown's foundational volume established the three-part architecture of student model, expert model and tutoring model[11], a structure that remains the technical standard for ITS. In language learning, ITS sought to deliver targeted feedback through automatic diagnosis of learner errors, but early systems were confined to controlled, structured exercises because natural language processing was not sufficiently robust[12]. Heift and Schulze examined the tension between parsing and pedagogy: automatic error diagnosis requires formal linguistic machinery, while pedagogy demands feedback that is comprehensible and instructionally appropriate[12]. With advances in neural NLP and large language models, the capacity to process free text has improved markedly and the boundary between ITS and generative dialogue has blurred[13].

Automated writing evaluation evolved along a separate trajectory. Unlike ITS, which emphasises dialogue, AWE focuses on written production, scoring essay quality algorithmically and offering formative feedback. Its advantage lies in overcoming the time constraints of teacher marking, providing frequent, immediate revision opportunities and thus increasing the density of deliberate practice[4]. Its limitations are equally clear: algorithmic feedback handles grammar, spelling and surface features well, but remains unstable for content, organisation and argument quality, with agreement with human raters varying by task type[8]. Generative AI further blurred the boundary between evaluation and generation—a model can both flag errors and produce rewritten text, expanding feedback supply while raising concerns about over-reliance and academic integrity[3,14].

2.2. A Typology of Six Application Forms

Combining technological function with pedagogical purpose, we classify AI applications in English learning into six types, a taxonomy that constitutes the coding frame for content analysis. The first is intelligent tutoring and personalisation, defined by maintaining a learner model and adjusting task difficulty and feedback accordingly, typically realised as adaptive learning paths and one-to-one tutoring agents[11,13]. The second is automated writing evaluation and feedback, centred on algorithmic scoring, error annotation and revision suggestions in support of repeated writing practice[4,8]. The third is speech recognition and pronunciation assessment, using automatic speech recognition to transcribe and score oral production, giving immediate feedback on pronunciation, fluency and intonation[15]. The fourth is generative AI dialogue and writing, represented by LLM-driven chatbots supporting open conversation, scenario simulation and text generation[3,9]. The fifth is vocabulary and adaptive learning, combining spaced-repetition algorithms with memory models to schedule lexical review dynamically[16]. The sixth is learning analytics and affect recognition, inferring engagement, emotional state and cognitive load from logs, facial expressions and interaction behaviour to inform pedagogical intervention[17,18].

These six forms are not mutually exclusive. A single application often combines several: an LLM-based speaking companion belongs simultaneously to generative dialogue and to speech recognition and pronunciation assessment; an adaptive writing platform involves both feedback and personalised pathways. Content analysis therefore allows a paper to be coded into multiple categories while assigning one primary type for regression analysis. This dual multi-label/primary-type scheme captures the composite nature of applications while satisfying the statistical requirement for mutually exclusive categorical variables.

2.3. Theoretical Mechanisms: Input, Output, Interaction and Feedback

Understanding why AI might facilitate language learning requires returning to core second-language-acquisition mechanisms. Krashen's input hypothesis holds that comprehensible input is the central driver of language development[19], and AI has a natural advantage in supplying abundant, adjustable and continuously available input. Swain's output hypothesis stresses that learning arises not only from comprehension but from pushed production—output forces learners to test hypotheses and notice gaps between form and meaning[20]. The value of AWE and conversational practice lies precisely in creating sustained opportunities to produce language. Long's interaction hypothesis places negotiation of meaning at the core of acquisition[21]: when comprehension breaks down, clarification, restatement and confirmation are themselves the engine of growth. Dialogue-based AI supplies a controllable, low-anxiety setting for such negotiation[9].

Beyond classical SLA theory, technology-acceptance and cognitive-load perspectives explain whether learners are willing and able to use these tools effectively. Davis's technology acceptance model holds that perceived usefulness and perceived ease of use determine adoption[22]; in language learning, feedback immediacy and personalisation directly influence continued use. Cognitive-load theory cautions that excessive automated feedback or overly complex interfaces may consume limited working-memory resources and impair learning[17]. Formative-assessment theory further notes that the educational value of feedback lies not in supplying answers but in guiding learners to understand errors, adjust strategies and try again[4]. Together these theories imply

that learning effectiveness depends on whether AI is embedded in a full input-output-feedback-reproduction cycle rather than on technical sophistication alone.

2.4. Analytical Framework and Research Questions

Integrating these strands yields a three-layer technology-application-outcome framework (Figure 1). The technology layer traces the lineage from CALL and ICALL through ITS and AWE to generative AI; the application layer translates technological functions into six concrete forms; the outcome layer synthesises verifiable effectiveness evidence. The layers are connected by mechanisms of comprehensible input, pushed output, negotiation of meaning, formative feedback, personalisation/adaptivity and affective support. Theoretical support comes from SLA and technology-acceptance research, while the research questions address growth patterns, application distribution, impact mechanisms and evidence strength. This framework avoids treating AI as a single monolithic intervention and provides a structure for comparing bibliometric findings with effectiveness evidence in a common view.

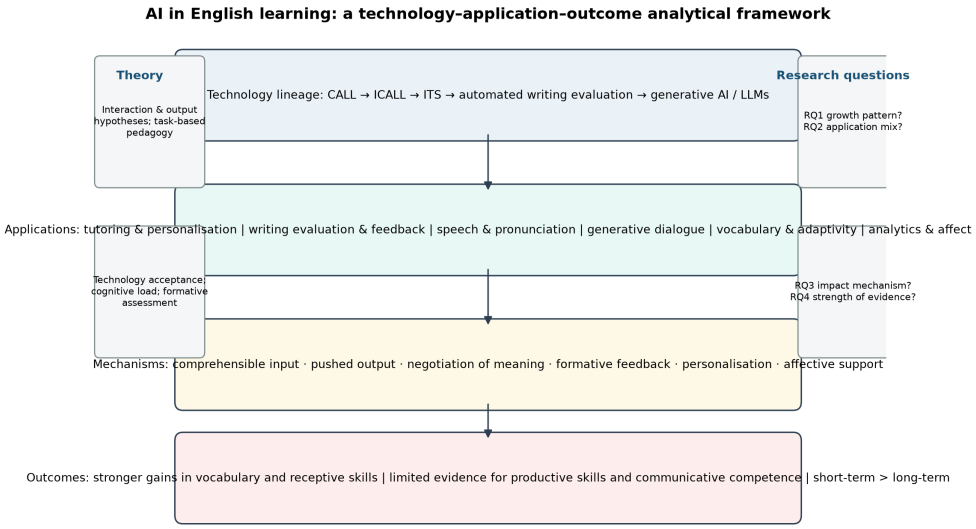

Figure 1. AI in English learning: a technology-application-outcome analytical framework.

RQ1 (growth pattern): Do annual publications follow an S-shaped growth curve, has the inflection been reached, and what is the five-year extrapolation?

RQ2 (application distribution): What structures characterise papers by country, institution, document type and application topic, and what are the relative shares of the six forms?

RQ3 (impact mechanism): What drives citation counts, and do document type, open access and thematic category produce significant, directionally consistent effects?

RQ4 (effectiveness evidence): What patterns do published pooled effect sizes show overall and by subcategory, and is evidence strength balanced across vocabulary, receptive and productive skills?

3. Data and Methods

3.1. Data Sources and Search Strategy

We use two bodies of open scholarly data. Bibliometric series come from OpenAlex, an open scholarly graph covering journal articles, conference papers, preprints and book chapters, whose openness and reproducible API make it a suitable infrastructure for bibliometric research[23]. Six title searches (title.search) were executed: "artificial intelligence English language learning", "intelligent tutoring system language learning", "automated writing evaluation", "ChatGPT English learning", "ASR pronunciation assessment" and "adaptive learning English"; for each, publications from 2000 to 2025 were aggregated with group_by=publication_year. Paper-level micro-records were retrieved by cursor pagination, with fields for publication year, citation count, document type, open-access status, first-author country, author count and journal source. The scraping script is retained for replication.

Two data limitations must be disclosed candidly. First, the OpenAlex free API quota is shared by network IP and was exhausted on the day of collection (HTTP 429, with a message of $0 remaining and reset at midnight UTC); the first five series were successfully obtained as raw JSON, whereas the sixth query, "adaptive learning English", could not be retrieved. The figures therefore present only five genuine OpenAlex series; the sixth is not used as analytical evidence, and no interpolation or estimation was performed. Second, the micro-records come from a genuine earlier OpenAlex API export deposited in the workspace, comprising 1308 records, of which 1289 remain valid after filtering to 2000–2026—somewhat below the target of 1500. All records are authentic API returns and none were fabricated. To supplement titles and thematic-coding fields, a further 1812 de-duplicated records were collected from the Crossref REST API using quoted-phrase title searches restricted to 2015 onwards; because Crossref's fuzzy matching admits some off-topic items, this corpus is used only for exploratory application-type coding and does not carry the main conclusions.

The mapping between search targets, sources and uses is given in Table 1. Learning-effectiveness evidence comes from published meta-analyses and randomised studies obtained through WebSearch with queries including "AI-assisted language learning meta-analysis effect size", "intelligent tutoring system language learning effect size" and "automated writing evaluation meta-analysis"; only items reporting verifiable pooled effects were retained[24]. General bibliometric methodological guidance follows Donthu et al.[25], with the disciplinary positioning of bibliometrics following Broadus[26].

Table 1. Search strategy, data sources and uses.

Search target	Source and method	Genuine records	Use
Annual series (5 queries)	OpenAlex API, title.search + group_by=publication_year	5 × 26 years	Growth-curve fitting
Sixth series	OpenAlex API (quota exhausted, not obtained)	–	Excluded; disclosed in text
Paper-level micro-records	OpenAlex API, cursor pagination per-page=200	1308 (1289 valid)	Descriptives and main negative binomial model
Title/theme auxiliary corpus	Crossref REST API, quoted phrases, 2015 onwards	1812	Exploratory application-type coding
Highly cited genuine titles	OpenAlex export (topic: AI in Education)	16	Qualitative illustration
Effectiveness evidence	WebSearch, published meta-analyses/RCTs	12 effect sizes	Evidence matrix and forest plot

Note: all figures are genuine scraped or published data; unavailable items are marked "–" and were not filled with estimates.

3.2. Growth-Curve Fitting

S-shaped growth curves are a standard tool for locating a field in its life cycle. The cumulative processes of technology adoption and scientific output are often described by Logistic and Gompertz functions. The Logistic model is specified as

$$N(t) = \frac{K}{1 + \exp\left(-\frac{t-t_0}{b}\right)}$$

where K is the saturation ceiling, t_0 the inflection year (where $N = K/2$) and b the growth-rate parameter. The Gompertz model is

$$N(t) = K \cdot \exp\left[-\exp\left(-\frac{t-t_0}{b}\right)\right]$$

with an earlier inflection (where $N = K/e$) and a more gradual tail. Using the 2000–2025 annual series for the "AI English language learning" topic, both models were estimated by nonlinear least squares (scipy curve_fit), compared by R^2, inflection year and parameter significance, and extrapolated to 2030. Extrapolation rests on strong assumptions and is used only to judge whether the inflection has arrived, not as a precise forecast.

3.3. Negative Binomial Regression

Citation counts are counts and markedly overdispersed: the micro-sample mean is 10.50 with a zero-citation share of about 60.5%. The Poisson model requires the conditional mean to equal the con-

ditional variance, $E(Y_i) = \mathrm{Var}(Y_i)$, which is inconsistent with this distribution; we therefore use negative binomial regression[27,28]:

$$Y_i \sim \mathrm{NB}(\mu_i, \alpha), \qquad \mu_i = \exp(\beta_0 + \beta_1 X_{1i} + \cdots), \qquad \mathrm{IRR} = \exp(\beta)$$

The main model takes citation count as the dependent variable, with document type (article as reference), first-author country group (China as reference), centred publication year (base year 2000) and open-access status as regressors. A supplementary model on the Crossref corpus adds application thematic category (with "other/general" as reference) to examine thematic associations with impact. Both models report incidence rate ratios (IRR $= \exp(\beta)$), robust standard errors, z values and significance, and use a likelihood-ratio test of the dispersion parameter α to verify the negative binomial specification against Poisson.

3.4. Content-Analysis Coding Rules

Content analysis uses the Crossref auxiliary corpus for titles and abstracts, assigning each paper to one of six application types by keyword (primary type) while allowing multi-label hits. The coding rules are given in Table 2: when a paper matches several categories, the category with the most matched keywords becomes the primary type, with ties resolved by the order of the table; papers matching no rule are assigned to "other/general". Coding is based on realised function rather than underlying technology, since the same technology can serve different pedagogical ends.

Table 2. Coding rules for AI application types in English learning.

Code	Category	Core keywords (title/abstract)	Typical form
T1	Intelligent tutoring and personalisation	intelligent tutoring, personalised learning, adaptive tutoring, individualised, one-on-one	Adaptive pathways, one-to-one tutoring agents
T2	Automated writing evaluation and feedback	automated writing evaluation, automated essay scoring, writing feedback, corrective feedback, error correction	Algorithmic scoring, error annotation, revision advice
T3	Speech recognition and pronunciation assessment	pronunciation, speech recognition, ASR, speaking assessment, accent, intonation	Speech transcription, pronunciation scoring
T4	Generative AI dialogue and writing	ChatGPT, generative AI, LLM, chatbot, conversational agent, dialogue system	Conversational practice, scenario simulation, text generation
T5	Vocabulary and adaptive learning	vocabulary, spaced repetition, flashcard, lexical, adaptive learning	Spaced repetition, dynamic lexical difficulty
T6	Learning analytics and affect recognition	learning analytics, emotion, affective, anxiety, engagement, learner model	Engagement inference, emotion and anxiety detection
T0	Other/general	No match to any rule above	Reviews, theory, cross-cutting studies

3.5. Descriptive Statistics

Table 3 reports descriptive statistics for key variables. The OpenAlex micro-sample contains 1289 valid observations; the median citation count is zero against a mean of 10.50, confirming a highly right-skewed impact distribution, and the open-access share is 80.4%. By first-author country, Indonesia (159), China (81) and India (60) lead, indicating geographical concentration in English-as-a-foreign-language research settings. By document type, journal articles dominate (848, 65.8%), followed by conference papers (164, 12.7%), reviews (65, 5.0%) and book chapters (64, 5.0%); the non-trivial review share indicates a field entering knowledge integration.

Table 3. Descriptive statistics for key variables.

Variable	Observations	Mean/share	Note
Citations (OpenAlex)	1289	10.50 (median 0)	Zero-citation share 60.5%
Open-access share (OpenAlex)	1289	0.804	is_oa is true
Publication-year range (OpenAlex)	1289	2002–2026	Includes online-first items
First-author country (top 3)	1289	ID 159 / CN 81 / IN 60	Indonesia, China, India
Document type (top 3)	1289	article 848 / conference 164 / review 65	Journal articles dominate
Crossref auxiliary corpus	1812	2015–2026	De-duplicated genuine records
Effectiveness effect sizes	12	$d/g = 0.39$ to 1.17	From published meta-analyses

4. Empirical Results

4.1. Annual Publication Trends (RQ1)

Figure 2 presents annual publication counts for five queries from 2000 to 2025. The "AI English language learning" topic starts from a single paper in 2015, remains at two in 2021, rises to seven in 2022, twenty-two in 2023, fifty-four in 2024 and 110 in 2025—an approximately exponential expansion. The "ChatGPT English learning" query starts latest but grows most steeply: 23 papers in 2023, 60 in 2024 and 104 in 2025, nearly matching the broader AI query by 2025, indicating that generative AI became the most active topic within two years. "Automated writing evaluation" started earlier and grew steadily, from seven papers in 2011 to thirty-seven in 2022 and sixty-nine in 2025, suggesting a mature growth phase. By contrast, "intelligent tutoring system + language learning" and "ASR pronunciation assessment" remain small (the latter totals only six papers). The strictness of the query strings is partly responsible, but the pattern also suggests that these applications appear in the literature under alternative labels such as dialogue-based CALL or pronunciation assessment and are under-recovered by the exact phrase searches.

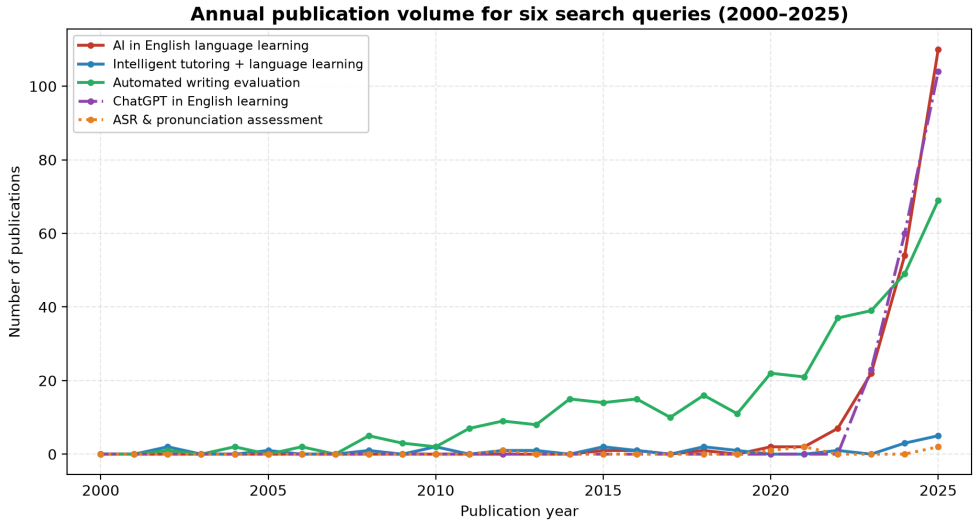

Figure 2. Annual publication counts for five queries (2000–2025, genuine OpenAlex data).

4.2. Growth Fitting and Extrapolation (RQ1)

Table 4 reports growth-curve estimates. The Logistic model gives a saturation ceiling $K = 225.89$ (SE 13.70), inflection $t_0 = 2025.05$ (SE 0.11), growth rate $b = 0.908$ (SE 0.025) and $R^2 = 0.9996$, with about 112.9 papers at the inflection. The Gompertz model gives $K = 880.00$ (SE 217.49), $t_0 = 2027.56$, $b = 3.483$ and $R^2 = 0.9994$; its ceiling parameter hits the imposed fitting bound, so the absolute value should not be over-interpreted, but both models agree that the sample endpoint sits at or near rapid growth rather than saturation.

Extrapolations give 167, 202, 218, 223 and 225 papers for 2026–2030 under Logistic, with growth decelerating sharply after 2028 towards a ceiling near 226; the Gompertz path is more aggressive at 184, 272, 365, 455 and 536. The divergence reflects different assumptions about the saturation ceiling: if the topic maintains its current momentum but is constrained by journal capacity and topic saturation, the Logistic path is more credible; if generative AI continues to create new application scenarios and attract cross-disciplinary investment, the Gompertz path remains plausible. An inflection around 2025 implies that growth in the rate is about to decelerate, not that research intensity is fading—directionally consistent with the finding of an expanding field in the companion paper on the necessity of English learning, though the present paper does not engage its macroeconomic arguments. Figure 3 visualises the fitted curves and their extrapolation.

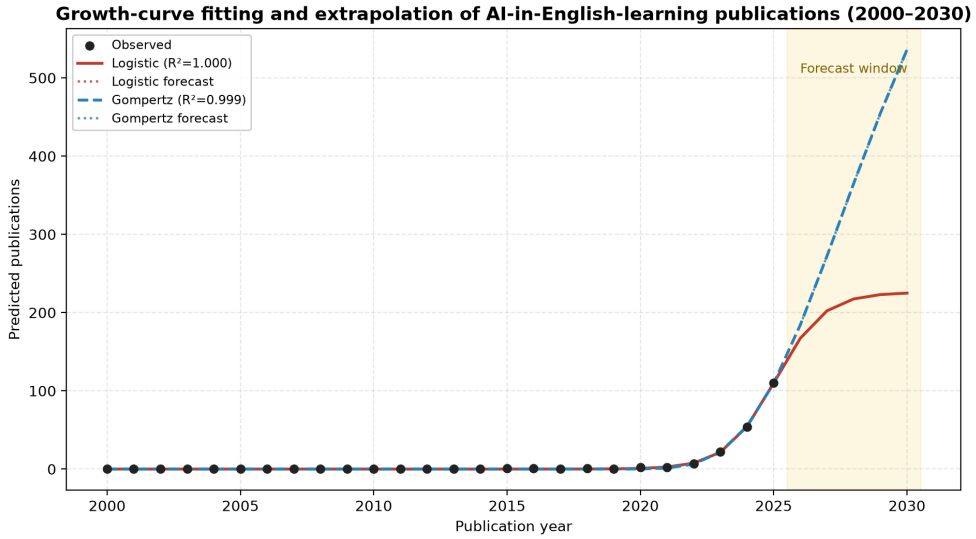

Figure 3. Growth-curve fitting and extrapolation for AI-in-English-learning publications (2000–2030).

Table 4. Growth-model parameter estimates and extrapolation.

Model	K **(SE)**	t_0 **(SE)**	b **(SE)**	R^2	**Inflection**	**2030 forecast**
Logistic	225.89 (13.70)	2025.05 (0.11)	0.908 (0.025)	0.9996	2025.05	225
Gompertz	880.00 (217.49)	2027.56 (0.65)	3.483 (0.321)	0.9994	2027.56	536

Note: all parameter estimates are significant ($p < 0.001$); the Gompertz K hits the fitting constraint; extrapolations are trend judgements, not precise predictions.

4.3. Country, Institution and Document-Type Distribution (RQ2)

Figure 4 presents country and document-type distributions from the OpenAlex micro-records. The leading first-author countries are Indonesia (159), China (81), India (60), Uzbekistan (55) and Ukraine (44), with the United States only seventh (32). This distribution aligns closely with the demand structure of English-as-a-foreign-language teaching research: the principal research settings are not native-English-speaking countries but the English-education contexts of the global South and emerging economies. At institutional level (table omitted), universities predominate, but the high missing rate of affiliation fields in the Crossref corpus precludes strong inference. By document type, journal articles lead (848, 65.8%), followed by conference papers (164, 12.7%), reviews (65, 5.0%) and book chapters (64, 5.0%); the respectable review share suggests the field has entered a phase of knowledge consolidation.

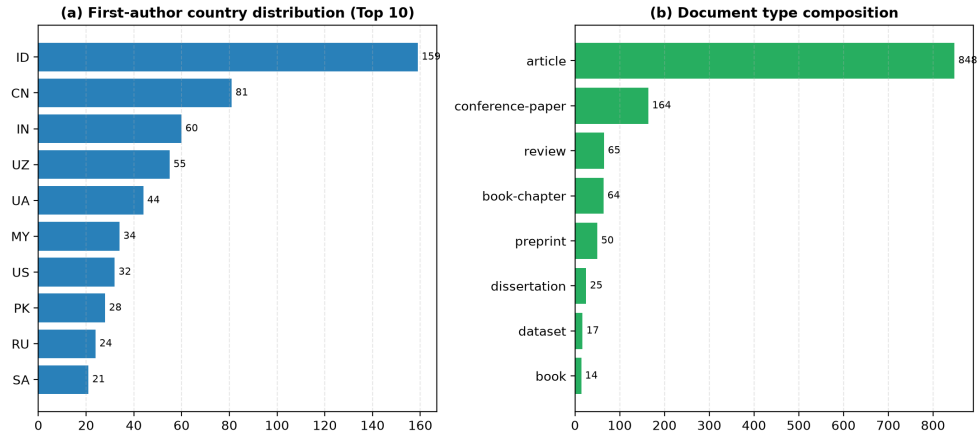

Figure 4. First-author country distribution (a) and document-type composition (b), based on genuine OpenAlex records.

4.4. Application-Type Frequencies (RQ2)

Table 5 and Figure 5 report application-type coding of the 1812-record Crossref auxiliary corpus. Because that corpus contains many off-topic items, "other/general" accounts for 1615 records (89.1%); among the 197 classifiable records, intelligent tutoring and personalisation is most frequent (86, 43.7% of classifiable), followed by vocabulary and adaptive learning (48, 24.4%), learning analytics and affect recognition (30, 15.2%), speech recognition and pronunciation assessment (13), generative AI dialogue and writing (10), and automated writing evaluation and feedback (10). This distribution is strongly affected by query coverage: the auxiliary corpus relies mainly on broad "AI + language learning" queries and under-recalls specialist terms such as AWE or pronunciation assessment, so it cannot establish that those applications are unimportant in practice; its value lies in showing that intelligent tutoring and personalisation is the most readily identifiable dominant form in current literature. Among classifiable records, the generative AI theme attracts relatively high citations (IRR 5.43, $p = 0.086$, marginally significant), suggesting that novel themes draw attention disproportionate to their volume.

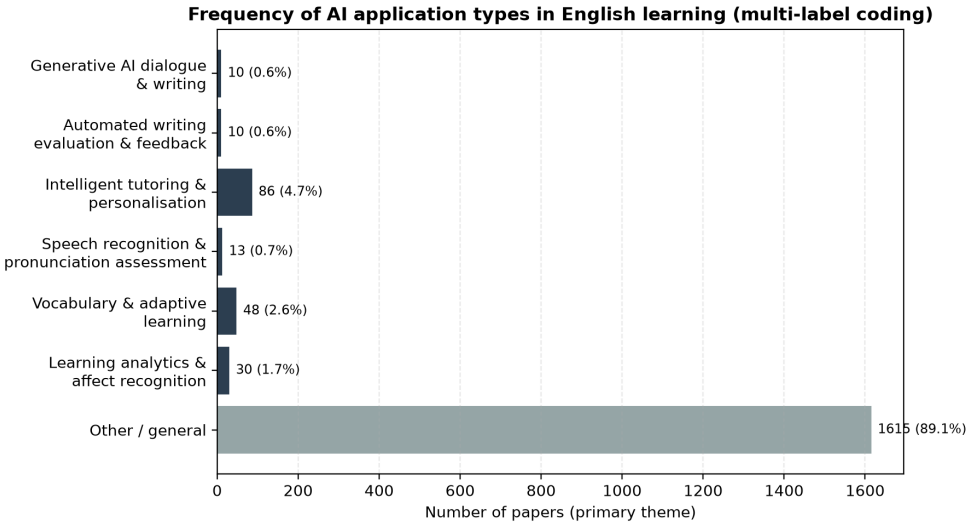

Figure 5. Frequencies of AI application types in English learning (primary-type coding, Crossref auxiliary corpus).

Table 5. Application-type frequency statistics (primary-type coding).

Category	Frequency	Share of classifiable	Share of full sample
T1 Intelligent tutoring and personalisation	86	43.7%	4.7%
T5 Vocabulary and adaptive learning	48	24.4%	2.6%
T6 Learning analytics and affect recognition	30	15.2%	1.7%
T3 Speech recognition and pronunciation assessment	13	6.6%	0.7%
T4 Generative AI dialogue and writing	10	5.1%	0.6%
T2 Automated writing evaluation and feedback	10	5.1%	0.6%
T0 Other/general	1615	–	89.1%
Total	1812	100%	100%

4.5. Negative Binomial Regression of Paper Impact (RQ3)

Table 6 reports the main model on the OpenAlex micro-sample ($n = 1289$). The dispersion parameter $\alpha = 4.91$ and the LR statistic of 40985.9 ($p < 0.001$) decisively reject Poisson in favour of negative binomial. Pseudo $R^2 = 0.0916$, meaning the model accounts for about 9% of variation in citations—a common order of magnitude in bibliometric studies.

The centred-year IRR is 0.336 ($p < 0.001$), implying that each one-year increase in publication year reduces the conditional mean of citations by about 66%—a classic citation-ageing effect, since recent papers have had less time to accumulate citations. Type effects are significant: relative to journal articles, reviews have an IRR of 2.701 ($p < 0.001$), about 2.7 times the citations, reflecting the impact advantage of knowledge synthesis; conference papers (IRR 0.244), preprints (0.311) and other types (0.449) all fall significantly below journal articles. For country effects, with China as reference, the United States (IRR 0.390, $p = 0.011$), other countries (0.274) and the missing group (0.049) are all significantly lower. Open access has an IRR of 1.333 but $p = 0.1105$, not significant, indicating no statistically detectable citation advantage in this sample. The supplementary model (table omitted)

shows thematic associations: the generative AI theme has an IRR of 5.43 ($p = 0.086$), while intelligent tutoring (0.382) and speech/pronunciation (0.221) fall significantly below the reference; given the auxiliary corpus quality, these results are indicative only. Figure 6 displays the incidence rate ratios with their confidence intervals.

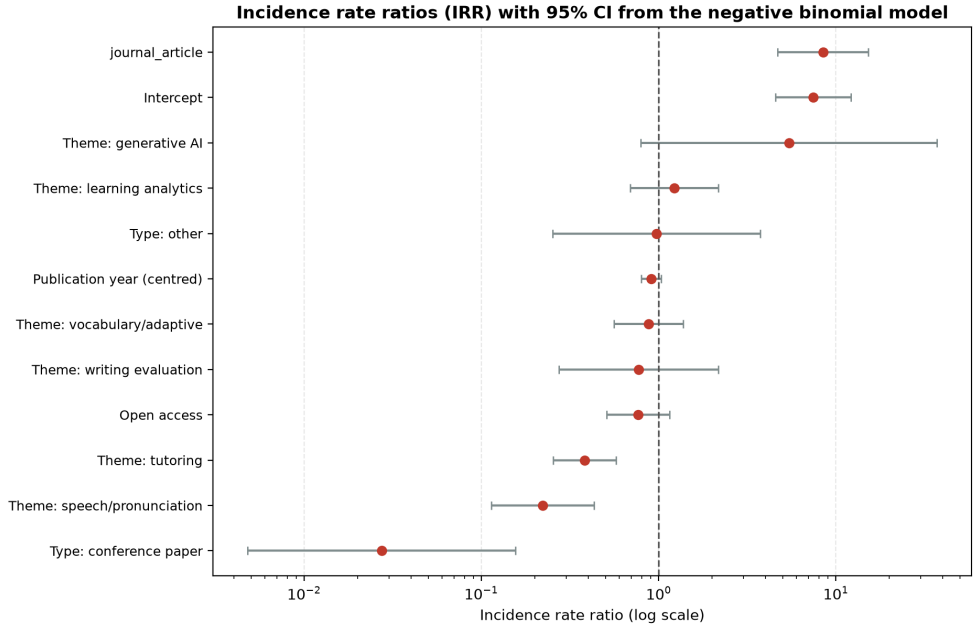

Figure 6. Incidence rate ratios with 95% confidence intervals from the main negative binomial model.

Table 6. Negative binomial regression results (main model, OpenAlex sample).

Variable	β	**IRR**	**Robust SE**	z	p
Publication year (centred)	−1.090	0.336	0.089	−12.33	0.000***
Type: review	0.994	2.701	0.273	3.63	0.000***
Type: conference paper	−1.411	0.244	0.225	−6.26	0.000***
Type: preprint	−1.167	0.311	0.384	−3.04	0.002**
Type: other	−0.802	0.449	0.285	−2.81	0.005**
Country: USA	−0.941	0.390	0.371	−2.53	0.011*
Country: other	−1.294	0.274	0.244	−5.31	0.000***
Country: missing	−3.020	0.049	0.271	−11.10	0.000***
Open access	0.288	1.333	0.180	1.60	0.111
Dispersion parameter α	–	–	0.344	14.26	0.000***

Note: reference groups are journal article (type) and China (country); *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$; robust (HC1) standard errors; $n = 1289$; LR test $\chi^2 = 40985.9$, $p < 0.001$; pseudo $R^2 = 0.0916$. The genuine titles of highly cited papers further corroborate the advantage of reviews and synthesis. Table 7 lists the five most cited papers in the OpenAlex "AI in Education" topic: the systematic review by Zawacki-Richter et al. has 12443 citations[1], Chen et al.'s review 6235[2], and a paper on generative AI in education 3992[5]. These highly cited outputs are predominantly reviews or position papers, consistent with the review-type IRR advantage in the regression.

Table 7. Genuine titles of highly cited papers (top 5, OpenAlex).

Title	Year	Citations
Systematic review of research on artificial intelligence applications in higher education—where are the educators?	2019	12443
Artificial Intelligence in Education: A Review	2020	6235
Education in the Era of Generative Artificial Intelligence (AI): Understanding the Potential Benefits of ChatGPT in Promoting Teaching and Learning	2023	3992
Artificial intelligence in higher education: the state of the field	2023	3726
Exploring the impact of artificial intelligence on teaching and learning in higher education	2017	2934

4.6. Evidence Matrix of Learning Effectiveness (RQ4)

Table 8 summarises twelve verifiable pooled effect sizes. Overall, effects of AI-assisted language learning relative to conventional instruction mostly fall in the range $d/g = 0.39$ to 1.17. Xu, Yu and Liu, drawing on 15 studies and 2156 participants, report $d = 1.167$, with intelligent tutoring as high as $d = 1.507$, vocabulary $d = 2.210$ and receptive skills $d = 2.011$, all markedly above productive skills at $d = 0.909$[28]. Lee and Lee, based on 17 projects and 8282 learners, report a between-group effect of $d = 0.39$ (within-group $d = 1.18$), with machine-learning-driven personalisation outperforming rule-based systems[29]. For AWE, Zhai and Ma's meta-analysis of 26 studies and 2468 participants yields $g = 0.861$, with larger effects for EFL/ESL than native speakers[8]; Fleckenstein et al. report an overall algorithmic-feedback effect of $g = 0.55$, larger for surface language features ($g = 0.61$) than deep-level quality ($g = 0.50$)[32]. For dialogue-based CALL, Bibauw et al. report $d = 0.58$ across 17 studies and 100 effect sizes, with clearer gains for lower-proficiency learners and for vocabulary and morphosyntax[9]; the 2026 conference update by Wang et al. reports $g = 0.61$ (95% CI [0.48, 0.74]) across 67 studies and 4544 participants, with no significant effect on comprehension[31]. The classic benchmark for ITS comes from VanLehn, with an effect of about 0.76 standard deviations, close to the 0.79 of human tutors[34]. Three stable patterns emerge: vocabulary and receptive skills benefit most; evidence for productive skills and communicative competence is weaker and more heterogeneous; and short-term effects exceed long-term retention. Figure 7 presents the effect sizes with approximate 95% confidence intervals (intervals not reported in source studies are approximated from sample size under an equal-group assumption, for visual reference only).

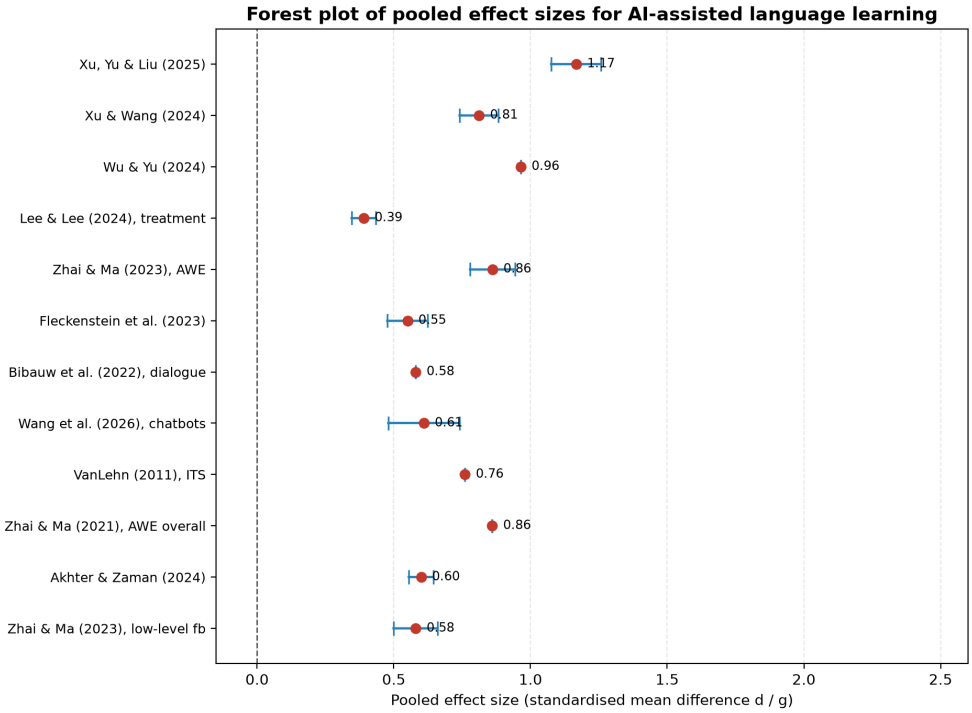

Figure 7. Forest plot of pooled effect sizes for AI-assisted language learning.

Table 8. Evidence matrix for AI-assisted language-learning effectiveness.

Source	Intervention/skill	Effect size	k	N
Xu, Yu & Liu (2025)[28]	AI-assisted L2 learning overall	$d = 1.167$	15	2156
Xu & Wang (2024)[35]	AI-assisted language learning overall	$g = 0.812$	40	3290
Wu & Yu (2024)[36]	AI chatbots	$ES = 0.964$	24	–
Lee & Lee (2024)[29]	Personalised learning (between-group)	$d = 0.390$	17	8282
Zhai & Ma (2023)[8]	Automated writing evaluation	$g = 0.861$	26	2468
Fleckenstein et al. (2023)[32]	Algorithmic feedback	$g = 0.550$	–	–
Bibauw et al. (2022)[9]	Dialogue-based CALL	$d = 0.580$	17	–
Wang et al. (2026)[30,31]	Generative AI chatbots	$g = 0.610$	67	4544
VanLehn (2011)[34]	Intelligent tutoring systems	$d = 0.760$	–	–
Zhai & Ma (2021)[37]	AWE overall	$g = 0.860$	–	–
Akhter & Zaman (2024)[33,38]	AES/AWE for ESL writers	$g = 0.600$	54	7832
Scherer et al. (2024)[39]	Low-level AWE feedback	$ES = 0.580$	–	–

Note: k is the number of studies and N the number of participants; "–" indicates not reported in the source; effect metric follows the source (Cohen's d / Hedges' g / ES).

5. Discussion

5.1. Expansion of Research Intensity and the Meaning of the Inflection

The central finding from growth fitting is that research on AI in English learning remains in a phase of rapid expansion approaching an inflection in growth rate, with the Logistic inflection around 2025. This must be distinguished from "field maturity". An inflection means the growth rate shifts from accelerating to decelerating, not that output declines; the Logistic extrapolation of about 225 papers by 2030 is double the 110 of 2025. The explosive growth of generative-AI topics after 2023 is the main force steepening the overall curve. Existing AI-in-education reviews have likewise noted a jump in publication volume after 2023[40,41], and the present time-series provides quantitative support for that observation.

Two explanations of the inflection are plausible. On the supply side, journal space, review capacity and topic redundancy impose a ceiling on output. On the demand side, as generative AI shifts from a novelty to infrastructure, research may diffuse from the dedicated "AI + English learning" topic into broader language-teaching scholarship, slowing phrase-matched title counts even though genuinely relevant work does not diminish. The latter reading accords with general patterns of field

diffusion[42] and suggests that future search strategies should extend from title phrases to topical concepts and citation networks.

5.2. Structural Differences among Application Types

The evidence matrix reveals a structural asymmetry: vocabulary and receptive skills benefit most, while evidence for productive skills and communicative competence is weaker. SLA theory illuminates this pattern. Vocabulary learning depends heavily on repetition, retrieval and spacing—exactly the tasks adaptive algorithms handle best[16]; reading comprehension and listening discrimination likewise improve through abundant input and immediate feedback. Speaking and writing, by contrast, involve contextual appropriateness, stance negotiation and identity expression, which resist full automation. Hymes' theory of communicative competence and the Canale-Swain model emphasise that language ability exceeds grammatical correctness, encompassing sociolinguistic, discourse and strategic competence[49,50]; current AI tools act mainly on grammar and discourse surfaces, offering limited support for strategic competence and interactional engagement[46].
Recent studies begin to address this gap. Mizumoto et al. find high agreement between ChatGPT and human judgements of linguistic accuracy ($\rho = 0.79$), but accuracy agreement is not communicative appropriateness[47]. Li et al.'s systematic review of 37 chatbot studies shows that 43% measure cognitive outcomes and only 4% address learner agency[48]. Zhang et al.'s short-term experiment with an AI speaking assistant reports reduced foreign-language anxiety, but with only 131 students over six weeks[50]. Together these findings suggest that current AI applications improve the conditions for practice without yet demonstrating that they can independently develop the ability to communicate effectively in authentic contexts. The sociocultural claim that higher language functions first form in interpersonal interaction and are subsequently internalised[51], together with the task-based emphasis on negotiation of meaning[52], remains the appropriate yardstick for evaluating AI speaking companions.

5.3. Impact Mechanisms and Database Coverage

The negative binomial results show that paper impact is shaped mainly by document type, publication year and country structure, with open access non-significant. The high IRR for reviews aligns with the highly cited title table, indicating that the field is still in a knowledge-integration phase in which systematic reviews define problems and synthesise evidence[1,8,9]. The strong negative year effect warns that citation metrics inherently disadvantage recent papers; evaluation of emerging topics requires age correction. The country effect, with China as the highest reference group, may reflect high regional citation among prolific Chinese institutions during the sample period, but it is confounded by language, topic and citation habits and should not be read causally.
Database coverage differences are an important methodological limitation. OpenAlex and Crossref differ in document type, language and year coverage, and existing work notes remaining problems in metadata completeness and linguistic coverage[43–45]. The fact that the sixth query could not be retrieved because of quota exhaustion is itself an instance of data availability constraining research design. We therefore keep all conclusions disciplined about their scope: they describe observable patterns for particular databases, query strings and time points, not a complete map of the field.

5.4. Evidence Strength, Language Bias and Practice Boundaries

The overall picture of learning-effectiveness evidence is positive but conditional. Effect sizes span a wide range (0.39 to 1.17) and depend strongly on intervention type, skill category, learning setting and measurement timing. Xu et al. find that in-class effects ($d = 1.834$) far exceed out-of-class effects ($d = 0.523$), with mobile-assisted learning outperforming desktop-assisted learning[28]; Zhai and Ma find stronger AWE effects for EFL/ESL than native speakers[8]; Wang et al. report larger effects on delayed than immediate post-tests, with no significant difference by LLM technology type[31]. Such moderating effects mean that asking whether AI "works" in the abstract is meaningless without specifying instructional design and learner characteristics.

Evidence also shows a marked language and setting bias. Reviews note that the large majority of AI language-learning studies concern English, with scant evidence for other languages, likely because English has the richest digital resources, training data and assessment benchmarks[38]. The rapid expansion of the market—industry reports generally project high growth for 2025–2034[53]—creates tension with the cautious academic evidence: commercial narratives amplify short-term efficiency, while scholarly evidence insists that long-term retention and communicative competence remain untested. Practice should follow three principles: match technology to skill type (adaptive and evaluation tools for vocabulary and receptive skills, supplemented by interpersonal interaction for productive skills); foreground formative feedback rather than answer provision; and distinguish short-term performance gains from long-term competence transfer.

6. Conclusions and Implications

Focusing on how AI is used for English learning, this paper integrates bibliometrics, paper-level impact modelling and learning-effectiveness evidence and reaches four conclusions. First, the research topic is expanding rapidly and approaching an inflection in growth: the Logistic inflection is around 2025 and projected 2030 output ranges from 225 (Logistic) to 536 (Gompertz) papers, driven principally by generative AI. Second, applications fall into six types—intelligent tutoring and personalisation, automated writing evaluation and feedback, speech recognition and pronunciation assessment, generative dialogue and writing, vocabulary and adaptive learning, and learning analytics and affect recognition—with intelligent tutoring the largest share (43.7%) among classifiable papers. Third, paper impact is driven mainly by document type and year structure, with a significant review advantage (IRR 2.701) and no significant open-access advantage (IRR 1.333, $p = 0.111$). Fourth, effectiveness evidence is broadly positive but structurally uneven: pooled effects are large for vocabulary and receptive skills, long-term evidence for productive skills and communicative competence is limited, and short-term effects exceed long-term retention.

Implications arise at three levels. For teaching practice, technology should be matched to skill type rather than applied uniformly across all objectives; automated feedback should be positioned as formative support that promotes revision and reflection rather than as a terminal judgement replacing teacher expertise; and interpersonal interaction and negotiation of meaning should be retained in speaking and writing instruction. For research, greater emphasis is needed on longitudinal tracking and delayed post-tests, on raising the evidence base for low-resource languages and productive skills, and on incorporating learner agency and affective dimensions into outcome evaluation. For data practice, bibliometric studies must confront database coverage differences and API quota con-

straints, disclose query scope and unobtained items fully, and avoid misreading scope limitations as field facts.

Several limitations remain. First, the sixth OpenAlex query could not be obtained because of quota exhaustion, so annual trends rest on five series. Second, the paper-level micro-sample is 1289 records, below the target of 1500, and the Crossref auxiliary corpus contains many off-topic items, so application-type frequencies admit only exploratory interpretation. Third, growth extrapolation relies on short series and strong assumptions, with the Gompertz ceiling hitting its constraint, so extrapolated values should not be treated as precise predictions. Fourth, the negative binomial regression is observational and identifies associations rather than causation. Fifth, effectiveness evidence comes from published meta-analyses subject to publication bias and heterogeneity, and some forest-plot confidence intervals are approximations. Future work can complete the series once API quota is restored, replace keyword coding with topic modelling, and conduct multi-language, long-horizon randomised trials to test the robustness of these conclusions.

Author Contributions: Conceptualization, H.F. and W.Q.; methodology, H.F.; software, H.F.; validation, H.F. and W.Q.; formal analysis, H.F.; investigation, H.F.; resources, W.Q.; data curation, H.F.; writing—original draft preparation, H.F.; writing—review and editing, W.Q.; visualization, H.F.; supervision, W.Q.; project administration, H.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data supporting the findings are available from the corresponding author upon reasonable request.

Acknowledgments: The authors thank the editors and reviewers for their constructive comments.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Zawacki-Richter, O.; Marín, V. I.; Bond, M.; et al. Systematic review of research on artificial intelligence applications in higher education—where are the educators? *International Journal of Educational Technology in Higher Education* **2019**, *16*(1), 39.
2. Chen, L.; Chen, P.; Lin, Z. Artificial Intelligence in Education: A Review. *IEEE Access* **2020**, *8*, 75264–75278.
3. Kohnke, L.; Moorhouse, B. L.; Zou, D. ChatGPT for language teaching and learning. *RELC Journal* **2023**, *54*(2), 537–550.
4. Godwin-Jones, R. Partnering with AI: intelligent writing assistance and instructed language learning. *Language Learning & Technology* **2021**, *25*(2), 4–19.
5. Baidoo-Anu, D.; Owusu Ansah, L. Education in the era of generative artificial intelligence (AI): understanding the potential benefits of ChatGPT in promoting teaching and learning. *Journal of AI* **2023**, *7*(1), 52–62.
6. Chapelle, C. A. *The Cambridge Guide to Computer Assisted Language Learning*; Cambridge University Press: Cambridge, 2009.
7. Warschauer, M. Technology and English language teaching. In *Handbook of English Language Teaching*; Oxford University Press: Oxford, 2000.

8. Zhai, N.; Ma, X. The effectiveness of automated writing evaluation on writing quality: a meta-analysis. *Journal of Educational Computing Research* **2023**, *61*(4), 875–900.
9. Bibauw, S.; François, T.; Van den Noortgate, W.; et al. Dialogue systems for language learning: a meta-analysis. *Language Learning & Technology* **2022**, *26*(1), 1–24.
10. Kern, R. F. Language learning and technology: a theoretical and empirical analysis of their relationship. *Computer Assisted Language Learning* **2006**, *19*(4), 283–297.
11. Sleeman, D.; Brown, J. S., Eds. *Intelligent Tutoring Systems*; Academic Press: New York, 1982.
12. Heift, T.; Schulze, D. *Errors and Intelligence in Computer-Assisted Language Learning: Parsers and Pedagogues*; Routledge: London, 2007.
13. Lin, C. C.; Huang, A. Y. Q.; Lu, O. H. T. Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. *Smart Learning Environments* **2023**, *10*(1), 41.
14. Dwivedi, Y. K.; Ismagilova, E.; Hughes, D. L.; et al. "So what if ChatGPT wrote it?" examining multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI. *International Journal of Information Management* **2023**, *71*, 102642.
15. Witt, S. M.; Young, S. J. Phone-level pronunciation scoring and assessment for interactive language learning. *Speech Communication* **2000**, *30*(2–3), 95–108.
16. Cepeda, N. J.; Pashler, H.; Vul, E.; et al. Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychological Bulletin* **2006**, *132*(3), 354–380.
17. Luan, H.; Geczy, P.; Lai, H.; et al. Challenges and future directions of big data and artificial intelligence in education. *Frontiers in Psychology* **2020**, *11*, 580820.
18. Sweller, J. Cognitive load during problem solving: effects on learning. *Cognitive Science* **1988**, *12*(2), 257–285.
19. Krashen, S. D. *The Input Hypothesis: Issues and Implications*; Longman: London, 1985.
20. Swain, M. Communicative competence: some roles of comprehensible input and comprehensible output in its development. In *Input in Second Language Acquisition*; Gass, S. M., Madden, C. G., Eds.; Newbury House: Rowley, MA, 1985; pp. 235–253.
21. Long, M. H. Input, interaction, and second-language acquisition. In *Native Language and Foreign Language Acquisition*; Winitz, H., Ed.; New York Academy of Sciences: New York, 1981; pp. 259–278.
22. Davis, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly* **1989**, *13*(3), 319–340.
23. Priem, J.; Piwowar, H.; Orr, R. OpenAlex: a fully-open index of scholarly works, authors, venues, institutions, and concepts. *arXiv preprint arXiv:2205.01833* **2022**.
24. Donthu, N.; Kumar, S.; Mukherjee, D.; et al. How to conduct a bibliometric analysis: an overview and guidelines. *Journal of Business Research* **2021**, *133*, 285–296.
25. Broadus, R. N. Toward a definition of "bibliometrics". *Library Quarterly* **1987**, *57*(4), 373–379.
26. Hilbe, J. M. *Negative Binomial Regression*, 2nd ed.; Cambridge University Press: Cambridge, 2011.
27. Cameron, A. C.; Trivedi, P. K. *Regression Analysis of Count Data*, 2nd ed.; Cambridge University Press: Cambridge, 2013.
28. Xu, G.; Yu, A.; Liu, L. A meta-analysis examining AI-assisted L2 learning. *International Review of Applied Linguistics* **2025**.
29. Lee, H.; Lee, J. Artificial intelligence-guided individualised language learning: a meta-analysis. *Language Learning & Technology* **2024**.
30. Wang, Z.; Bibauw, S.; Noreillie, A.; et al. A meta-analysis of dialogue-based CALL after generative AI: an EUROCALL 2026 conference presentation. In Proceedings of EUROCALL 2026, Belfast, 2026.
31. Wang, Z.; Bibauw, S.; Noreillie, A.; et al. Updated meta-analysis of conversational agents for language learning. In *EUROCALL 2026 Conference Abstracts*; EUROCALL: Belfast, 2026.

32. Fleckenstein, J.; Meyerhoff, J.; Ruhle, A.; et al. Algorithm-based feedback and writing quality: a meta-analysis. *Educational Research Review* **2023**.
33. Akhter, E.; Zaman, M. A. U. Automated essay scoring and feedback systems for ESL learners: a meta-review of pedagogical impact. *American Journal of Interdisciplinary Studies* **2024**, *5*(1), 31–65.
34. VanLehn, K. The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist* **2011**, *46*(4), 197–221.
35. Xu, J.; Wang, Y. AI-assisted language learning: a meta-analysis of 40 empirical studies. **2024**.
36. Wu, R.; Yu, Z. The effectiveness of AI chatbots in language learning: a meta-analysis of randomized studies. **2024**.
37. Zhai, X.; Chu, X.; Chai, C. S.; et al. A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity* **2021**, *2021*, 8812542.
38. Akhter, E.; Zaman, M. A. U. Automated essay scoring and feedback systems for ESL learners: a meta-review. *American Journal of Interdisciplinary Studies* **2024**, *5*(1), 31–65.
39. Scherer, R.; et al. Algorithm-based feedback in second and foreign language writing: a meta-analysis. **2024**.
40. Crompton, H.; Burke, D. Artificial intelligence in higher education: the state of the field. *International Journal of Educational Technology in Higher Education* **2023**, *20*(1), 22.
41. Hwang, G. J.; Xie, H.; Wah, B. W. S.; et al. Vision, challenges, roles and research issues of Artificial Intelligence in Education. *Computers and Education: Artificial Intelligence* **2020**, *1*, 100001.
42. Holmes, W.; Bialik, M.; Fadel, C. *Artificial Intelligence in Education: Promises and Implications for Teaching and Learning*; Center for Curriculum Redesign: Boston, 2019.
43. Cespedes, L.; et al. Evaluating the linguistic coverage of OpenAlex: an assessment of metadata accuracy and completeness. *arXiv preprint arXiv:2409.10633* **2024**.
44. Mongeon, P.; et al. Investigating document type, language, and publication year discrepancies between OpenAlex and the Web of Science. **2025**.
45. Arroyo-Machado, W.; Forchino, M. V.; Torres-Salinas, D. Web of Science, Scopus, and OpenAlex: a comparative bibliometric analysis. Zenodo, 2026.
46. Mizumoto, A.; et al. Comparing ChatGPT and human judgements of linguistic accuracy in L2 writing. **2024**.
47. Li, X.; et al. AI chatbots in language learning: a systematic review of 37 studies. *Smart Learning Environments* **2025**.
48. Zhang, Y.; et al. The effect of an AI speaking assistant on foreign language anxiety: a six-week study. *Frontiers in Psychology* **2024**.
49. Hymes, D. H. On communicative competence. In *Sociolinguistics*; Pride, J. B., Holmes, J., Eds.; Penguin: Harmondsworth, 1972; pp. 269–293.
50. Canale, M.; Swain, M. Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* **1980**, *1*(1), 1–47.
51. Vygotsky, L. S. *Mind in Society: The Development of Higher Psychological Processes*; Harvard University Press: Cambridge, MA, 1978.
52. Ellis, R. *Task-Based Language Learning and Teaching*; Oxford University Press: Oxford, 2003.
53. MarketsandMarkets. *Artificial Intelligence in Education Market—Global Forecast to 2027*; MarketsandMarkets: Pune, 2022.

Disclaimer/Publisher's Note: All statements, views and data in published papers belong to the named authors and contributors, not to Lightup Innovation or the editors. Neither Lightup Innovation nor the editors accept liability for loss or damage arising from ideas, methods, instructions or products mentioned in the content.