frontiers in Psychology

TYPE Original Research
PUBLISHED 04 November 2026
DOI 10.3389/fpsyg.2026.1995850

Check for updates

OPEN ACCESS

EDITED BY
Juan Pedro Martínez-Ramón,
University of Murcia, Spain

REVIEWED BY
Ahmadreza Mohebbi,
The University of Auckland, New Zealand
Shannon McCrocklin,
Southern Illinois University, United States

*CORRESPONDENCE
Olusiji Adebola Lasekan
✉ olasekan@uct.cl
Margot Teresa Godoy Pena
✉ margot.godoy@ufrontera.cl

RECEIVED 11 September 2026
REVISED 02 October 2026
ACCEPTED 13 October 2026
PUBLISHED 04 November 2026

CITATION
Lasekan OA, Godoy Pena MT and Mathew BS (2026) From chatbot to language-learning partner: learner–LLM configuration, affordances, and regulatory demands in foreign language learning. *Front. Psychol.* 17:1995850. doi: 10.3389/fpsyg.2026.1995850

COPYRIGHT
© 2026 Lasekan, Godoy Pena and Mathew. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

From chatbot to language-learning partner: learner–LLM configuration, affordances, and regulatory demands in foreign language learning

Olusiji Adebola Lasekan[1]*, Margot Teresa Godoy Pena[2]* and Blessy Sarah Mathew[3]

[1]Departamento de Educación e Innovación, Universidad Católica de Temuco, Temuco, Chile, [2]Languages Coordination, Dirección de Trayectoria Formativa (DITFO), Universidad de La Frontera, Temuco, Chile, [3]Department of Economics, Lovely Professional University, Phagwara, Punjab, India

Large Language Models (LLMs) are increasingly used as flexible resources for self-directed foreign-language learning, yet limited research explains how learners configure these systems as pedagogical partners. This study examined pedagogical role construction, modality–pedagogy alignment, affordance realization, breakdowns, and regulatory demands in learner–LLM interaction. Using a retrospective observational computational content-analysis design, 117 first-person public posts and comments, representing 80 pseudonymous participants, were analyzed from seven public forums and X. Structured multi-label coding and quantitative analyses identified eight pedagogical roles and examined relationships among interaction modalities, learning activities, perceived affordances, and breakdowns. Results showed that learners reported configuring LLMs as multifunctional pedagogical partners, most prominently as interlocutors, while 65.8% of accounts involved multiple roles. Voice interaction was significantly associated with conversation and pronunciation, whereas text was associated with vocabulary learning, supporting modality–pedagogy alignment. More orchestrated learning configurations were associated with greater perceived benefits, particularly personalization and convenience, although an apparent difference in activity breadth was substantially attenuated once the activities used to define orchestration were excluded from the count. However, 43.6% of accounts reported concerns, revealing function-specific breakdowns involving accuracy/language reliability, access/privacy/system constraints, pedagogical sustainability, interaction/continuity, and voice/technical reliability. The findings support an empirically informed framework conceptualizing learner–LLM interaction as an AI-mediated regulatory process involving role construction, modality and activity configuration, affordance realization, monitoring, breakdown detection, and regulatory demand. Psychological self-regulation and objective proficiency gains were not directly measured.

KEYWORDS
foreign-language learning, Large Language Models, learner–LLM interaction, pedagogical affordances, self-regulated learning

1 Introduction

The rapid integration of Large Language Models (LLMs) into foreign-language learning is transforming generative chatbots into increasingly versatile language-learning partners. Tools such as ChatGPT provide interactive, personalized, and readily accessible opportunities for language practice, supporting learner engagement both within and beyond formal instructional contexts (Bertolini Mendes and Panhan, 2024; Kohnke et al., 2023). This transformation has accelerated as LLMs have evolved from predominantly text-based systems toward multimodal and customizable environments incorporating voice interaction, speech technologies, and personalized agents (Mai et al., 2025; Wan and Moorhouse, 2025). Unlike conventional Computer-Assisted Language Learning (CALL) applications, which are typically designed around predetermined instructional functions, general-purpose LLMs can be configured by learners to perform multiple pedagogical roles, including conversational interlocutor, tutor, corrective-feedback provider, language explainer, and learning organizer (Sadigzade, 2025). Their pedagogical value therefore resides not only in their technological capabilities but also in how learners appropriate those capabilities in response to their individual goals and learning needs. This learner-driven configuration represents an important shift in technology-mediated language learning: rather than simply using a predefined educational tool, learners increasingly participate in constructing the pedagogical role that the technology assumes within their learning process.

The integration of Large Language Models into self-directed foreign-language learning marks a significant expansion of learner autonomy in digital environments. Traditionally, autonomous language learning involved selecting among available resources such as online courses, videos, mobile applications, and other authentic digital materials, often beyond formal institutional settings (Beaven, 2021; Chik, 2017; Godwin-Jones, 2019). LLMs extend this model by allowing learners not only to choose resources but also to configure how those resources behave pedagogically. Learners can set goals, determine the focus and difficulty of activities, select interaction strategies, request particular forms of feedback, and adapt the LLM to function as a tutor, conversational partner, or personalized learning assistant (Khadawardi, 2025; Ye et al., 2025). This configurability makes LLMs particularly suited to informal and self-regulated learning, where learners can access support flexibly and shape interactions according to their evolving needs. Thus, whereas earlier forms of digital autonomy largely centered on resource selection, LLM-based autonomy increasingly involves resource configuration, with learners actively constructing the pedagogical behavior of the technology itself.

The interaction between learners and Large Language Models creates a range of pedagogical affordances that extend beyond the models' technical capabilities alone. In language-learning contexts, LLMs can support personalization, immediate feedback, convenience, constant availability, engagement, flexibility, fluency practice, comprehension, and vocabulary development by adapting responses to learners' needs and providing on-demand support (Daberao, 2025; Zhang and Jiang, 2026). These affordances are especially relevant to self-directed learning because learners can determine when and how to seek feedback, adjust task difficulty, request clarification, and use the system for different linguistic purposes. However, such benefits should not be understood as inherent properties of the technology; rather, they are realized through the interaction between the learner, the task, and the way the LLM is configured and used. Thus, the pedagogical value of LLMs depends not simply on what the technology is capable of doing, but on how learners mobilize those capabilities to achieve specific language-learning goals.

Self-regulation provides an important psychological lens for understanding how learners manage their interactions with Large Language Models during foreign-language learning. Self-regulated learning involves processes such as goal setting, strategy selection, monitoring, evaluation, adaptation, feedback seeking, resource management, and metacognitive control, through which learners actively direct and adjust their own learning (Zimmerman and Schunk, 2011). LLMs can support these processes by helping learners formulate goals, suggest strategies, provide immediate feedback, monitor progress, recommend resources, and offer alternative approaches when difficulties arise (Sun et al., 2025). In learner–LLM interaction, self-regulation therefore extends beyond managing one's own cognitive and behavioral processes to include the active configuration and evaluation of the AI-supported learning environment itself. Learners may decide what role the LLM should perform, how it should respond, which modality to use, and whether its feedback is sufficiently accurate or useful to retain, revise, or reject. Accordingly, effective LLM-assisted language learning can be understood as a recursive process in which learners plan, interact, monitor, evaluate, and adapt both their own learning strategies and the way they engage with the LLM.

Despite their growing pedagogical potential, Large Language Models remain vulnerable to technological, linguistic, and pedagogical breakdowns that can constrain their effectiveness in foreign-language learning. These include hallucinated or inaccurate information (Dentella et al., 2025; Guerreiro et al., 2023; Marchisio et al., 2024). Such problems may be particularly consequential in language learning because learners can have limited ability to recognize inaccurate explanations or corrections, while context-memory and speech-processing failures can disrupt personalized and sustained interaction. Furthermore, reliance on LLMs may reduce opportunities for authentic human interaction and the cultural and pragmatic learning associated with it (Lowin Kropf, 2025). Importantly, these limitations may create regulatory demands by requiring learners to evaluate the accuracy, relevance, continuity, and pedagogical usefulness of AI support. The present study therefore examines not only perceived affordances and reported breakdowns but also the potential regulatory demands implicated by those breakdowns, without assuming that learners necessarily recognized or responded to them.

Despite the growing literature on Large Language Models in language education, important gaps remain in understanding how learners actually integrate these tools into self-directed foreign-language learning. Existing research is often tool-centric, emphasizing what systems such as ChatGPT can do rather than how learners appropriate and configure them for their own

learning purposes. Studies also tend to examine single functions, such as writing, speaking, vocabulary, or grammar, even though learner–LLM interaction can involve multiple pedagogical roles and activities within the same learning process. In addition, research frequently discusses LLM affordances and limitations as separate sets of benefits and challenges, with less attention to how technological or pedagogical breakdowns create demands for learner monitoring and adaptation. Consequently, there remains limited empirical modeling of learner–LLM interaction as an integrated process linking learner intention, AI role construction, modality selection, learning activity, affordance realization, breakdown, and regulatory demand. Addressing this gap is important because LLM-assisted language learning appears to involve not simply the use of an AI tool, but a dynamic and recursive process in which learners continually shape and evaluate the pedagogical function of the technology. This gap follows directly from the preceding framing of learner-configured pedagogical roles, affordance realization, self-regulation, and LLM breakdowns in the manuscript.

The study addresses five research questions: **(RQ1)** What pedagogical roles do foreign-language learners report assigning to LLMs in self-directed learning? **(RQ2)** How are reported LLM interaction modalities associated with reported language-learning activities? **(RQ3)** How do personalization, feedback, convenience, and other perceived affordances co-occur within reported self-directed learning configurations? **(RQ4)** What technological and pedagogical breakdowns do learners report across LLM-supported learning activities, and what potential regulatory demands are implicated by these reported breakdowns? **(RQ5)** How can the relationships evidenced across these first-person accounts be integrated into an empirically informed model of learner–LLM interaction? By focusing on reported practices and potential regulatory demands rather than assuming direct measurement of psychological self-regulation, the study seeks to identify patterns evidenced in learners' first-person accounts that can inform a theoretically grounded model of AI-mediated self-regulated language learning.

2 Theoretical background

Self-directed foreign-language learning in digital environments centers on learner autonomy, agency, strategic resource selection, and regulation of learning activities. Learners set goals, select resources, evaluate progress, and sustain informal engagement through digital activities (Khadawardi, 2025). Such environments expand access to authentic language and incidental learning (Dizon, 2023), while CALL and MALL can support self-monitoring, self-management, and independent learning (Noori, 2025). Agency therefore involves selecting and strategically using resources in relation to learning goals, while self-regulation entails planning, organizing, monitoring, and adjusting learning activities (Rowberry, 2022).

Self-regulated learning (SRL) provides a psychological framework for understanding this process through the cyclical phases of forethought, performance, and self-reflection. During forethought, learners establish goals and select appropriate tasks and resources (Pipattarasakul and Singhasiri, 2018; Zimmerman and Schunk, 2011). During performance, they apply cognitive, metacognitive, and motivational strategies, monitor learning, seek assistance, and use feedback (Pipattarasakul and Singhasiri, 2018; Torre and Daley, 2023). During self-reflection, learners evaluate performance and modify subsequent strategies (Fomina, 2022; Tseng et al., 2017). This recursive perspective is particularly relevant to LLM-supported learning because learners can configure AI-supported activities while monitoring both their own performance and the quality of AI-generated support.

Large Language Models can consequently be conceptualized as configurable pedagogical partners whose roles are shaped by learners' goals. LLMs may serve as conversational interlocutors supporting dialogue and fluency practice (Ye et al., 2025), corrective coaches providing feedback (Alyahyan et al., 2025; Xu et al., 2025), and vocabulary assistants (Cavojský et al., 2023). They can also generate customized reading, listening, and practice materials (Marquez-Carpintero et al., 2025), and assist with planning, progress monitoring, and assessment (Alyahyan et al., 2025; Marquez-Carpintero et al., 2025). Unlike narrowly designed technologies, therefore, LLMs can be configured and reconfigured across multiple pedagogical roles.

This configurability highlights the distinction between technological capability and pedagogical affordance. Capabilities describe what a system can perform; affordances emerge when learners mobilize those capabilities in relation to goals and contexts. Cognitive affordances include feedback, explanation, comprehension support, and vocabulary development (Nguyen and Doan, 2025). Interactional affordances include conversation, responsiveness, and fluency practice (Du and Reynolds, 2025), while personalization involves adapting content, difficulty, and feedback (Nguyen, 2022). Organizational affordances include convenience, availability, and planning support (Haines, 2017), whereas affective affordances include engagement, confidence, and reduced anxiety (Kim and Su, 2024; Zheng, 2024). Pedagogical value thus depends on affordances realized through specific learner–LLM configurations rather than technological capability alone.

Multimodality further conditions affordance realization because different modes support different linguistic processes. Multimodal environments combining text, audio, speech, visual input, and interaction can support vocabulary, listening, communication, and engagement (Khabirova, 2025). AI-supported learning extends this repertoire to voice, speech recognition, text-to-speech, custom agents, and integrations (Xabirova and Ro'zimov, 2025). Their value depends on modality–pedagogy alignment: text may better support reflection, grammar, vocabulary, and written correction, whereas voice may better support conversation, pronunciation, fluency, and spontaneous production. This reflects approaches that align technologies with communicative tasks and learner needs (Al-Jarf, 2024).

However, LLM-supported learning also involves epistemic, interactional, pedagogical, technical, and access-related breakdowns. Epistemic problems include hallucinations, inaccurate corrections, and unreliable explanations (Lau, 2024). Interactional breakdowns include robotic exchanges, excessive agreement, and loss of conversational context (Lau, 2024; Mahlow, 2023).

Pedagogical breakdowns include insufficient structure, weak sequencing, and difficulty sustaining engagement (Albedah, 2025), while speech recognition and synthesized-voice limitations can disrupt oral learning (Koç and Savaş, 2025). Costs, usage restrictions, privacy concerns, and platform changes can also constrain continuity (Lau, 2024). From an SRL perspective, these breakdowns create regulatory demands: learners must evaluate AI reliability and pedagogical usefulness, recognize when interaction no longer serves their goals, and determine whether the learning configuration requires adjustment. Thus, learner–LLM interaction can be understood as a dynamic relationship among pedagogical role construction, affordance realization, modality alignment, and the monitoring demands generated by technological and pedagogical breakdowns.

2.1 Proposed conceptual framework of learner–LLM interaction

Drawing together the preceding perspectives on self-directed learning, self-regulation, configurable LLM roles, pedagogical affordances, multimodality, and LLM-related breakdowns, this study proposes an initial conceptual framework of learner–LLM interaction. The framework (Figure 1) conceptualizes LLM-assisted foreign-language learning as a goal-directed process beginning with a learner goal, which informs the pedagogical role assigned to the LLM, such as interlocutor, corrective coach, explainer, vocabulary assistant, input generator, pronunciation partner, or planner/evaluator. The selected role subsequently influences the interaction modality and the language-learning activity through which the learner engages with the system, recognizing that different modalities may be better aligned with different linguistic and communicative demands. These configurations, in turn, create opportunities for affordance realization, whereby technological capabilities become pedagogically meaningful through their use in relation to learners' goals and activities. Accordingly, the proposed relationship can be represented as Learner Goal → LLM Role → Modality → Activity → Affordance. Rather than treating this sequence as strictly linear, the framework positions monitoring and regulation as processes surrounding the interaction, consistent with SRL perspectives in which learners monitor performance, evaluate feedback, and adjust subsequent actions. At this theoretical stage, the framework deliberately remains parsimonious and does not predetermine how breakdowns or regulatory adaptations operate within the sequence; instead, these relationships are examined empirically in the subsequent analysis. This distinction is important because the study identifies regulatory demands rather than directly measuring psychological self-regulation.

3 Method

3.1 Research design

This study employed a retrospective observational computational content-analysis design to examine naturally occurring public accounts of LLM-assisted foreign-language learning. Guided by content-analysis principles (Krippendorff, 2018) and a theory-informed approach consistent with directed content analysis (Hsieh and Shannon, 2005), the study analyzed public posts and comments rather than elicited survey, interview, or experimental data.

The use of naturally occurring public accounts was deliberate rather than a substitute for direct observation. Whereas screen recordings, interaction logs, and think-aloud protocols can capture moment-to-moment behavior or cognitive processing within researcher-defined tasks, public discussion posts provide unsolicited retrospective accounts of how learners configure, experience, and evaluate LLMs across diverse self-directed contexts. Such accounts are particularly suited to identifying learner-perceived pedagogical roles, affordances, breakdowns, and experiences that users themselves considered salient enough to report. The resulting evidence therefore concerns reported practices and retrospective sense-making rather than directly observed learning behavior or cognition.

Structured computational coding and quantitative analysis of multi-label categorical data were used to examine five dimensions of learner–LLM interaction: pedagogical role construction, interaction modality, learning activity, perceived affordance, and technological or pedagogical breakdown. The corpus was interpreted as evidence of reported learner practices and perceptions, not direct behavioral observation, proficiency measurement, or psychological self-regulation; accordingly, SRL served as a theoretical lens for interpreting textually evidenced configurations and reported or analytically inferred regulatory demands. Throughout this study, "observed" or "textually evidenced" refers only to features identifiable in the written accounts; the language-learning practices described by authors were self-reported and were not directly observed by the researchers. The unit of analysis was one public post or comment. Of 323 initially retrieved forum and X records, 15 duplicates were removed, leaving 308 unique records; 226 were identified as relevant to LLM-supported human-language learning, and 117 contained explicit first-person evidence of LLM use for language learning or teaching. These 117 records constituted the principal analytical sample.

3.2 Data sources and corpus construction

The source frame covered public discourse published between 1 January 2023 and 31 August 2026, although X data represented a recent-search window collected on 31 August 2026 rather than complete historical coverage. The final analytical sample comprised 117 first-person records from seven publicly accessible forums and X, including 101 forum records (86.3%) from LingQ ($n = 38$), Read lang ($n = 36$), Forobeta ($n = 12$), Hacker News ($n = 8$), Chinois Tips ($n = 3$), Practice Portuguese ($n = 3$), and the Open AI Community ($n = 1$), alongside 16 X records (13.7%). Data were collected through source-appropriate public or authorized routes, including public Discourse JSON representations, the official Hacker News Firebase API, publicly accessible Forobeta pages, and an approved multilingual X recent-search workflow; records were subsequently deduplicated and cleaned of HTML

and quoted prior messages. Of 71 unique X posts retrieved after deduplication, 16 met the first-person inclusion criteria, while X text and native identifiers were retained only in a restricted audit layer in accordance with redistribution restrictions. Reddit and YouTube were excluded because the appropriate authorized research/API access was unavailable. Accordingly, the corpus constitutes a purposive sample of accessible public discourse rather than an exhaustive or representative sample of social-media discussion about LLM-assisted language learning.

3.3 Inclusion and exclusion criteria

Records were screened at two levels. First, a record was considered in scope when it concerned the use, testing, evaluation, recommendation, or planned use of an LLM or LLM-powered conversational system for learning or teaching a human language; records involving programming languages, unrelated generative-AI applications, generic promotional content, duplicates, or discussions without an identifiable LLM-related language-learning component were excluded. Second, the primary analytical sample was restricted to explicit first-person use, requiring evidence that the author had personally used an LLM or LLM-powered tool for human-language learning or teaching; positive, negative, mixed, unsuccessful, and discontinued experiences were all eligible. The final sample comprised 117 records corresponding to 80 pseudonymous participant keys, as some individuals contributed multiple posts or comments. Accordingly, record-level percentages represent patterns within the collected discourse rather than unique-person prevalence or population estimates, and participant-level robustness checks were conducted where repeated contributions could affect inferential analyses.

3.4 Analytical sample

The final sample comprised 117 first-person public posts and comments and was coded using a structured codebook consistent with content-analysis principles (Krippendorff, 2018) and a theory-informed directed approach (Hsieh and Shannon, 2005). Text chat was the dominant modality (71.8%), followed by voice chat (15.4%), text-to-speech (8.5%), speech-to-text (6.0%), API/integration (4.3%), custom agents (3.4%), and image input (1.7%). Multi-label coding captured relevance, evidence type, first-person use, stance, reported outcomes, products, learning activities, modalities, perceived benefits, concerns, target languages, contextual information, promotional content, and model-assigned coding confidence. Learning activities included conversation, role-play, writing correction, grammar explanation, vocabulary, translation, reading and listening support, graded-input and exercise generation, pronunciation, assessment/progress monitoring, study planning, resource recommendation, and memory/customization. Benefits included personalization, immediate feedback, convenience, engagement, fluency practice, comprehension, and vocabulary gain, while concerns encompassed accuracy, correction, interaction, speech/voice, context-memory, privacy, access, dependency, cost, and pedagogical-structure problems. Target languages were heterogeneous, and demographic or geographic characteristics were not inferred from user information. Reported outcomes were coded conservatively as improved, helpful_unspecified, no_clear_change, abandoned, switched_to_human, or not_reported; importantly, self-reported improvement was treated as perceived experience rather than objective evidence of proficiency development or causal effectiveness.

3.5 Codebook development and coding decision rules

The coding framework was organized into two analytical layers. First, a structured multi-label codebook operationalized the principal variable families relevant to the research questions, including learning activities, interaction modalities, perceived benefits, reported outcomes, and breakdowns. Individual codes were assigned according to phenomena explicitly described in each account, with multiple codes permitted when a record contained more than one analytically distinct phenomenon. Second, first-order breakdown indicators were grouped into higher-order breakdown domains for RQ4.

Breakdowns were classified according to the specific aspect of learner–LLM interaction reported. Accuracy/language reliability captured problems involving the correctness or linguistic appropriateness of generated content, whereas voice/technical reliability captured problems involving speech recognition, synthesized voice quality, latency, or related speech-system performance. Access/privacy/system constraints encompassed unavailable functionality, device or access restrictions, privacy issues, cost limitations, and related system conditions. Positive descriptions of system performance were not classified as breakdowns. The higher-order domains were not mutually exclusive; multiple classifications were permitted when an account explicitly described distinct types of failure. For example, an account describing both poor synthesized voice quality and an inappropriate regional variety could contribute to voice/technical reliability and accuracy/language reliability, respectively.

Example 1 – Voice/technical reliability: A learner reported that the system repeatedly misunderstood their speech and that the synthesized voice was poor. Codes: speech_recognition + voice_quality → Voice/technical reliability.

Example 2 – Accuracy/language reliability: A learner reported that a correct language form was marked as erroneous. Code: incorrect_correction → Accuracy/language reliability.

Example 3 – Access/system rather than voice reliability: A learner reported that microphone input could transcribe speech but the system did not provide spoken responses. Code: access_or_device_limits → Access/privacy/system constraints.

Example 4 – Legitimate cross-domain coding: A learner reported poor synthesized voice quality together with an unwanted regional accent. Codes: voice_quality + dialect_mismatch → Voice/technical reliability + Accuracy/language reliability.

3.6 Construction of higher-order analytical categories

Following primary coding, lower-level codes were organized into theoretically meaningful higher-order categories to facilitate interpretation (Krippendorff, 2018). For RQ1, 16 learning activities were consolidated into eight pedagogical roles: Interlocutor (conversation and role-play), Corrective coach (writing correction), Language explainer (grammar and translation), Lexical coach (vocabulary), Input scaffold (reading, graded input, and listening), Pronunciation/voice coach (pronunciation), Practice designer/evaluator (exercise generation and assessment/progress), and Learning manager/customizer (study planning, resource recommendation, and memory/customization). For RQ3, an orchestration indicator distinguished isolated task use from broader learning-management configurations, defined by the presence of study planning, memory customization, resource recommendation, or assessment/progress monitoring; activity, benefit, and concern breadth were also calculated as indicators of configuration complexity rather than psychometric measures of self-regulation. For RQ4, the audited first-order concern codes were consolidated into five multi-label breakdown domains: accuracy/language reliability (e.g., hallucination, incorrect correction, unnatural translation, and dialect mismatch), interaction/continuity (e.g., context-memory and interaction problems), voice/technical reliability (e.g., speech-recognition failure, poor synthesized voice quality, and latency/reliability problems), pedagogical sustainability (e.g., inadequate structure, sequencing, or sustained engagement), and access/privacy/system constraints (e.g., cost, privacy, unavailable functionality, device/access restrictions, and integration conditions).

3.7 Analytical strategy

Analysis proceeded in five stages corresponding to the research questions.

For **RQ1**, descriptive frequencies, percentages, multi-label breadth measures, and co-occurrence counts were used to characterize the pedagogical roles assigned to LLMs. Because multiple records could originate from the same contributor, role distributions were also examined after aggregation by pseudonymous participant identifier.

For **RQ2**, modality–activity associations were examined by grouping records into text-only, voice-only, text + voice, other, and missing-modality categories. Text-only and voice-only records were compared using Fisher's exact tests, with multiple comparisons controlled using the Benjamini–Hochberg false discovery rate (FDR) procedure (Benjamini and Hochberg, 1995); other modalities were examined descriptively. The text-only/voice-only comparisons were additionally examined at the participant level by aggregating records associated with the same pseudonymous contributor. The 16 X records without linkable participant identifiers were excluded from participant-level analyses.

For **RQ3**, orchestrated and other learning configurations were compared in activity breadth, perceived-benefit breadth, and reported-breakdown breadth using Mann–Whitney U tests with rank-biserial correlations. Individual affordances and activity–affordance relationships were examined using Fisher's exact tests, odds ratios, co-occurrence, lift, FDR-adjusted q values, and Spearman correlations. Because the orchestration indicator was defined by the presence of study planning, memory customization, resource recommendation, or assessment/progress monitoring, activity breadth was also calculated without these four defining activities. Orchestration–affordance relationships were additionally examined after aggregation at the participant level. Narrative elaboration was examined for the subset of forum records with original source text available in the restricted audit layer by calculating post/comment word counts and testing their association with perceived-benefit breadth using Spearman's rho.

For **RQ4**, the five higher-order breakdown domains were examined across learning activities to characterize activity-specific breakdown patterns and the potential regulatory demands associated with them. These analyses were descriptive and exploratory, particularly where activity-specific subsamples were small, and did not assume that learners recognized or responded adaptively to the reported breakdowns.

For **RQ5**, patterns across pedagogical roles, modalities, learning activities, perceived affordances, and breakdowns were integrated into an empirically informed framework of learner–LLM interaction. Exploratory agglomerative clustering was also examined as a possible means of identifying recurring configurations; however, the resulting solutions showed insufficient separation and stability to support cluster-based interpretation. All inferential analyses were interpreted as exploratory associations rather than causal effects.

3.8 Coding validation, reliability, and quality control

The coding workflow incorporated multiple quality-control procedures to enhance reproducibility and consistency (Krippendorff, 2018). Public text was cleaned by removing HTML, quoted messages, URLs, handles, email addresses, and phone-like strings; multilingual records were translated into English while retaining source-language information, and authors were represented using salted pseudonymous participant keys. Primary structured coding was conducted through an LLM-based pipeline (GPT-5 mini) using a fixed schema and codebook, with multilingual translation also performed by GPT-5 mini. A stratified sample of 60 records, drawn using a fixed random seed and stratified by platform, detected language, and in-scope status, was independently recoded by a second model (Claude Sonnet 4.6) to assess cross-model reproducibility. Exact agreement was 90.0% for in-scope classification, 75.0% for evidence type, 90.0% for first-person use, 83.3% for stance, 90.0% for reported outcome, and 95.0% for promotional status; mean Jaccard agreement for multi-label variables ranged from 0.610 for benefits to 0.942 for target languages. Records showing low model-assigned coding confidence, coding inconsistencies, out-of-scope leakage, or privacy-sensitive content were subsequently adjudicated and recoded. In the final 117-record sample, the mean model-assigned coding-confidence score was 0.852 (median = 0.86; minimum = 0.70). This score represents the primary model's self-assessed

confidence in its coding decision and should not be interpreted as an independently calibrated estimate of coding accuracy. These procedures were treated as quality-control and cross-model reproducibility checks rather than validation against human criterion judgments, particularly given the lower agreement for semantically complex multi-label categories such as benefits and learning activities. No independent human reliability coding was conducted. The complete coding prompts, codebook, decision rules, and adjudication criteria are provided in the Supplementary materials (S1).

3.9 Ethical considerations and data governance

The study analyzed only publicly accessible online discourse and involved **no recruitment, interaction, intervention, or collection of data directly from human participants**. Accordingly, formal ethical review approval was not required. Only publicly available content was analyzed, and data-minimization procedures were applied by removing usernames, handles, personal names, contact information, URLs, and other potentially identifying information. Forum contributors were assigned pseudonymous identifiers, and examples were paraphrased rather than quoted verbatim. X post text and native identifiers were excluded from the shareable dataset in accordance with platform restrictions. No sensitive demographic characteristics were inferred from user content. All findings are reported in aggregate and interpreted as patterns within publicly expressed experiences rather than as representative estimates of the broader population.

4 Results

4.1 Analytical sample and overall patterns of learner–LLM use

Before addressing the specific research objectives, the analytical sample and overall patterns of learner–LLM use were examined to establish the empirical context for the subsequent analyses. The analysis focused on 117 records containing explicit first-person accounts of LLM use for foreign-language learning, corresponding to 80 unique participant identifiers. Because some individuals contributed more than one account, the post or comment constituted the primary unit of analysis, while participant-level aggregation was used as a robustness check where appropriate.

Overall evaluations of LLM-assisted language learning were predominantly favorable. Of the 117 firsthand accounts, 74 (63.2%) expressed a positive stance, 30 (25.6%) a mixed stance, 7 (6.0%) a neutral stance, and 6 (5.1%) a negative stance. Regarding reported outcomes, 80 records (68.4%) described the LLM as helpful without specifying measurable improvement, whereas only 12 (10.3%) explicitly reported improvement. A further 18 records (15.4%) did not report an outcome, 6 (5.1%) reported no clear change, and one learner reported switching to human support.

Within the sampled public discourse, these findings indicate substantial reported perceived utility rather than demonstrated learning effectiveness. Because the corpus consists of naturally occurring self-reports rather than independently assessed proficiency outcomes, the subsequent analyses focus on how learners configured, experienced, and evaluated LLM-assisted learning rather than whether LLM use objectively improved foreign-language proficiency.

4.2 Objective 1: identifying the pedagogical roles assigned to LLMs

The first research objective was to identify the pedagogical roles that foreign-language learners assigned to LLMs in self-directed learning. To address this objective, 16 coded learning activities (Table 1) were consolidated into eight higher-order pedagogical roles representing different ways in which learners configured general-purpose LLMs as language-learning resources.

The most prominent role was interlocutor, appearing in 65.0% of firsthand accounts. However, learners did not restrict LLMs to conversational practice. Approximately one quarter of the records positioned the LLM as a language explainer (27.4%), corrective coach (26.5%), or lexical coach (26.5%). LLMs also functioned as input scaffolds (20.5%), learning managers/customizers (17.1%), practice designers/evaluators (16.2%), and pronunciation/voice coaches (6.8%).

Role construction was frequently multifunctional. The mean number of higher-order roles per record was 2.06 (median = 2), and 65.8% of records contained at least two pedagogical roles. The most frequent combinations were interlocutor + corrective coach ($n = 17$), interlocutor + lexical coach ($n = 16$), corrective coach + language explainer ($n = 16$), and language explainer + lexical coach ($n = 15$). The pattern remained evident at participant level: among the 80 unique participant identifiers, 78.8% showed evidence of LLM use as an interlocutor, while approximately one third used the LLM as a language explainer (31.2%), corrective coach (31.2%), or lexical coach (30.0%).

These findings indicate that the transition from chatbot to language-learning partner is characterized by learner-driven pedagogical role construction. Rather than using the LLM according to a single predetermined function, learners combined conversational, explanatory, corrective, lexical, scaffolding, and organizational functions according to immediate learning needs. The empirical pattern therefore supports conceptualizing the LLM as a multifunctional pedagogical partner whose role is partly constructed through learner interaction.

4.3 Objective 2: examining modality–pedagogy alignment

The second research objective was to determine whether different LLM interaction modalities were associated with different foreign-language learning activities. Text chat was the dominant modality, occurring in 84 of 117 accounts (71.8%), followed by voice chat (Table 2) ($n = 18$, 15.4%), text-to-speech ($n = 10$, 8.5%), speech-to-text ($n = 7$, 6.0%), API/integration-based interaction ($n = 5$, 4.3%), custom agents ($n = 4$, 3.4%), and image input ($n = 2$, 1.7%).

To examine modality specialization more directly, text-chat-only records ($n = 81$) were compared with voice-chat-only records ($n = 15$).

After correction for multiple comparisons, voice interaction was significantly more strongly associated with conversation practice ($q = 0.007$) and pronunciation practice ($q = 0.007$), whereas text interaction was significantly more strongly associated with vocabulary learning ($q = 0.024$).

The broader modality profiles reinforced this differentiation. All 18 voice-chat records involved conversation practice, and 33.3% involved pronunciation. Similarly, all seven speech-to-text accounts involved conversation, with 57.1% also involving pronunciation. By contrast, text-chat accounts were distributed across conversation (57.1%), vocabulary learning (34.5%), writing correction (32.1%), grammar explanation (28.6%), reading support (15.5%), and exercise generation (15.5%). Text-to-speech was associated particularly with conversation (80%), pronunciation (40%), reading support (30%), and exercise generation (30%). Although based on small samples, custom-agent use was concentrated around grammar explanation (75%), vocabulary learning (50%), and writing correction (50%), while API/integration use showed comparatively strong involvement in study planning (60%).

These results provide evidence of modality–pedagogy association rather than modality neutrality. Text interaction was associated with a broader range of reflective and form-focused activities, particularly vocabulary, grammar, correction, and reading, whereas voice and speech technologies were disproportionately represented among oral-interactional activities such as conversation and pronunciation. Modality selection may partly reflect learner intention—for example, learners may select voice precisely because they intend to practice conversation or pronunciation—rather than modality independently producing the activity; thus, modality appears to be associated with the pedagogical configuration of the LLM rather than functioning merely as an alternative interface.

Because the 117 records corresponded to only 80 pseudonymous forum participant identifiers (with 16 additional single-post contributions from X not linked to a tracked identifier), record-level associations could partly reflect repeated contributions from the same individuals. To assess this, the text-only/voice-only comparison was repeated after aggregating records to the 80 named participant identifiers (64 text-only, 12 voice-only participants). The pronunciation (1.9% vs. 33.3%; $p = 0.003$) and vocabulary (39.6% vs. 0.0%; $p = 0.006$) associations remained robust at participant level, while the conversation association was attenuated to a level just outside conventional significance (71.7% vs. 100.0%; $p = 0.054$). These results indicate that the pronunciation and vocabulary associations are unlikely to be artifacts of repeated contributions, whereas the conversation association should be interpreted with some caution.

4.4 Objective 3: examining learner orchestration and affordance realization

The third research objective was to examine how personalization, feedback, convenience, and other perceived affordances were combined within more complex self-directed LLM learning configurations. Because the dataset did not provide longitudinal measures of habitual self-regulation, learner orchestration was operationalized conservatively as the presence of at least one explicit learning-management behavior: study planning, memory customization, resource recommendation, or assessment/progress monitoring. Twenty-four records met this criterion, whereas 93 did not.

Orchestrated accounts contained a greater number of perceived benefits than other firsthand accounts. Personalization showed the strongest difference: 70.8% of orchestration records mentioned personalization compared with 15.1% of other records (OR $\approx$ 13.7, FDR-adjusted $q < 0.001$). Convenience was likewise more frequent among orchestrated records (75.0% vs. 36.6%; OR $\approx$ 5.21, $q = 0.007$). Because orchestration was itself defined by four learning-management activities (study planning, memory customization, resource recommendation, and assessment/progress monitoring), the raw activity-breadth comparison in Table 3 is partly circular. A sensitivity analysis recalculated activity breadth after excluding these four defining activities: the difference was substantially attenuated and no longer statistically significant (orchestrated M = 2.25 vs. other M = 1.90; Mann–Whitney U = 1,319.5, $p = 0.154$, rank-biserial $r \approx 0.18$). Broader activity breadth is therefore not treated as independent evidence supporting the orchestration construct; the construct is instead supported by its association with perceived-benefit breadth, personalization, and convenience. Across the complete firsthand corpus, activity breadth was positively associated with perceived-benefit breadth (Spearman's $\rho = 0.565$, $p < 0.001$), whereas activity breadth was not associated with concern breadth ($\rho = 0.054$, $p = 0.567$). As an additional robustness check using independent textual evidence, original post/comment word counts (rather than the length of the AI-generated paraphrase) were obtained for the subset of forum records whose source text remained available in the restricted audit layer (53 of the 101 forum records; word counts were not preserved for the remaining 48). Within this subsample, word count was significantly positively associated with perceived-benefit breadth (Spearman's $\rho = 0.428$, $p = 0.001$, $n = 53$), indicating that more elaborated original accounts tended to report a broader range of perceived benefits; this converges with the activity-breadth–benefit-breadth association above using a measure derived independently of the coding process.

As with the modality comparisons, the orchestration contrast was repeated at participant level using the 80 named forum participants (18 orchestrated, 62 other). The association with benefit breadth remained robust (M = 4.06 vs. 2.16; Mann–Whitney $p = 0.002$), as did personalization (83.3% vs. 17.7%; $p < 0.001$) and convenience (77.8% vs. 37.1%; $p = 0.003$). The orchestration construct is therefore supported primarily by its association with perceived benefits rather than by activity breadth, which is partly definitional (see above).

Specific activity–affordance relationships further demonstrated that pedagogical value varied according to the way the LLM was configured. Writing correction was strongly associated with perceived correction quality; grammar explanation with immediate feedback; reading support with comprehension; vocabulary learning with vocabulary gain; and study planning and memory customization with personalization. Notably, all seven memory-customization accounts identified personalization and convenience.

The results are conceptually consistent with a continuum from task-level use, through multi-task use, to more orchestrated learning configurations, although this continuum is a conceptual interpretation rather than an independently validated empirical progression, given the partly definitional relationship between orchestration and activity breadth noted above. At the more orchestrated end, learners reported combining planning, customization, resource-selection, progress-related, and multiple learning functions. These findings should not be interpreted as evidence that such learners possessed higher psychological self-regulation, which was not directly measured. Instead, the reported patterns provide textual evidence of AI-mediated regulatory practices associated with broader perceived-benefit realization, particularly personalization and convenience.

TABLE 1 Pedagogical roles assigned to LLMs in firsthand language-learning accounts.

Pedagogical role	Constituent functions	*n*	%
Interlocutor	Conversation practice; role-play	76	65.0
Language explainer	Grammar explanation; translation	32	27.4
Corrective coach	Writing correction	31	26.5
Lexical coach	Vocabulary learning	31	26.5
Input scaffold	Reading, graded input, listening support	24	20.5
Learning manager/customizer	Planning, resource recommendation, memory customization	20	17.1
Practice designer/evaluator	Exercise generation; assessment/progress	19	16.2
Pronunciation/voice coach	Pronunciation practice	8	6.8

Percentages exceed 100% because individual records could contain multiple pedagogical roles.

4.5 Objective 4: identifying breakdowns and their regulatory demands

The fourth research objective was to identify technological and pedagogical breakdowns across different LLM-supported learning activities and examine the regulatory demands associated with these breakdowns. Overall, 51 of 117 records (43.6%) (Table 4) contained at least one coded concern following the audited RQ4 coding described in Sections 3.5 and 3.8. These concerns were consolidated into five higher-order breakdown domains.

Accuracy/language reliability constituted the most frequent breakdown domain after the audit (12.8%), followed closely by access/privacy/system constraints and pedagogical sustainability (12.0% each) and interaction/continuity (10.3%). Importantly, however, breakdowns were not distributed uniformly across learning activities.

Writing correction and grammar explanation showed a relatively frequent association with accuracy-related problems within this exploratory breakdown, whereas pronunciation practice ($n = 8$) showed a relatively frequent association with access/system constraints combined with voice/technical limitations. Conversation practice and vocabulary learning

TABLE 2 Selected language-learning activities by text-only and voice-only LLM interaction.

Learning activity	Text only ($n = 81$)	Voice only ($n = 15$)	Fisher *p*	FDR-adjusted q	Odds ratio (voice vs. text) [95% CI]
Conversation practice	55.6% (45/81)	100.0% (15/15)	0.001	0.007	∞ [2.61, ∞]
Pronunciation practice	2.5% (2/81)	33.3% (5/15)	0.001	0.007	18.64 [2.65, 220.02]
Vocabulary learning	34.6% (28/81)	0.0% (0/15)	0.005	0.024	0.00 [0.00, 0.58]
Writing correction	32.1% (26/81)	6.7% (1/15)	0.060	0.239	0.15 [0.00, 1.11]
Grammar explanation	28.4% (23/81)	6.7% (1/15)	0.105	0.337	0.18 [0.00, 1.34]
Reading support	16.0% (13/81)	0.0% (0/15)	0.210	0.538	0.00 [0.00, 1.70]

TABLE 3 Learning-activity and perceived-affordance breadth in orchestrated and non-orchestrated LLM use.

Measure	Orchestrated use ($n = 24$)	Other use ($n = 93$)	Mann–whitney *U*	*p*	Rank-biserial *r*
Mean number of learning activities	3.54	1.90	1,846.5	<0.001	0.655
Median activities	3	2			
Mean number of perceived benefits	3.29	1.96	1,581.5	0.001	0.417
Median benefits	3	2			
Mean number of concerns	0.96	0.76	—	0.536	0.075

Orchestration is defined by the presence of at least one of four activities (study planning, memory customization, resource recommendation, assessment/progress monitoring), which are also included in the "number of learning activities" count above; the activity-breadth comparison is therefore partly definitional. A sensitivity analysis excluding these four defining activities found a substantially attenuated, non-significant difference (M = 2.25 vs. 1.90 non-defining activities; Mann–Whitney $U = 1{,}319.5$, $p = 0.154$, rank-biserial $r \approx 0.18$; see Section 4.4), so activity breadth is not treated as independent evidence for the orchestration construct. The benefit-, concern-, and modality-breadth comparisons above do not share this circularity.

TABLE 4 Higher-order breakdown conditions in learner–LLM interaction.

Breakdown category	Examples	*n*	%
Accuracy/language reliability	Hallucination, incorrect correction, unnatural translation, dialect mismatch	15	12.8
Access/privacy/ system constraints	Cost, privacy, unavailable functionality, device/access restrictions, and integration conditions	14	12.0
Pedagogical sustainability	Inadequate structure, sequencing, or sustained engagement	14	12.0
Interaction and continuity	Context/memory limitations, robotic responses, excessive agreement, loss of human interaction	12	10.3
Voice/technical reliability	Speech-recognition failure, poor synthesized voice quality, and latency/reliability problems	6	5.1

were more frequently associated with pedagogical-sustainability problems. These activity-specific patterns should be read as exploratory rather than as stable rankings: several subgroups are small (memory customization, $n = 7$; pronunciation practice, $n = 8$; exercise generation, $n = 14$), and the corresponding Wilson 95% confidence intervals in Table 5 are wide. Memory customization showed the highest point estimate (85.7%, 95% CI [48.7%, 97.4%]), with access/privacy/system constraints and pedagogical sustainability as the principal breakdown pattern, but the wide interval indicates substantial uncertainty around this estimate.

These findings reveal an affordance–breakdown duality in learner–LLM interaction. The same functionality associated with perceived pedagogical value may also constitute a corresponding point of vulnerability: immediate correction depends partly on correction accuracy; personalization depends partly on system and contextual continuity; oral practice depends partly on the availability and reliability of speech functionality; and open-ended conversation can expand interaction while simultaneously weakening pedagogical structure. Consequently, affordance realization does not remove the need for learner monitoring; instead, it can generate regulatory demands requiring learners to evaluate accuracy, relevance, continuity, and the adequacy of pedagogical support.

4.6 Objective 5: integrating role, modality, affordance, and breakdown

The fifth research objective was to integrate the relationships identified across pedagogical role, modality, learning activity, affordance, and breakdown into an empirically informed model of learner–LLM interaction. To address this objective, activity-specific profiles were constructed linking the immediate learning function with its dominant interaction modality, salient perceived affordances, and principal breakdown conditions.

Three integrative patterns emerged (Table 6). First, pedagogical role and modality were interdependent: conversation and pronunciation increasingly recruited voice and speech technologies, whereas form-focused and explanatory practices remained predominantly text based. Second, different configurations realized different affordances: immediate feedback was particularly salient in grammar, correction, translation, and vocabulary activities, whereas personalization and convenience were especially prominent in study planning and memory customization. Third, breakdowns tended to correspond to the affordances on which particular configurations depended: correction-based activities were vulnerable to accuracy/language-reliability failure, voice-based activities to both access/system and technical failure, conversational and vocabulary-oriented activities to pedagogical-sustainability failure, and personalization-oriented activities such as memory customization to access/system and pedagogical-sustainability failure.

The integrated results therefore indicate that pedagogical value does not increase automatically with greater LLM functionality. Instead, perceived value appears conditional on the configuration linking learner purpose, pedagogical role, interaction modality, learning activity, affordance realization, and the breakdown conditions associated with that configuration.

4.7 Empirically informed framework of learner–LLM configuration and regulatory demands

Building on Objective 5, the relationships observed across the five analyses were synthesized into a process-oriented framework explaining how a general-purpose LLM may be configured as a personalized foreign-language learning partner. The framework does not treat psychological self-regulation as a directly measured variable; instead, it identifies reported regulatory practices and points at which the accounts imply potential regulatory demands.

As shown in Figure 2, the process begins with learner need or immediate learning purpose. Although standardized learner goals were not measured, activities explicitly described in the accounts were treated as textual indicators of immediate learning purposes. These purposes informed pedagogical role construction, whereby learners positioned the LLM as an interlocutor, corrective coach, language explainer, lexical coach, input scaffold, pronunciation coach, practice designer/evaluator, or learning manager/customizer.

The next stage involves modality selection and alignment. Text, voice, speech technologies, custom agents, and integrations supported different forms of pedagogical interaction. The concentration of voice around conversation and pronunciation and the stronger association of text with vocabulary and other form-focused activities demonstrate that modality conditions how the pedagogical relationship is enacted. This is followed by learning-activity configuration and affordance realization, through which technological capabilities become perceived pedagogical value in the form of immediate feedback, personalization, convenience, engagement, flexibility, comprehension, and fluency opportunities.

TABLE 5 Dominant breakdown profiles across selected LLM-supported activities.

Learning activity	*n*	Any concern, %	Wilson 95% CI	Principal breakdown pattern
Memory customization	7	85.7	[48.7, 97.4]	Access/privacy/system constraints and pedagogical sustainability
Pronunciation practice	8	62.5	[30.6, 86.3]	Access/privacy/system constraints and voice/technical reliability
Exercise generation	14	57.1	[32.6, 78.6]	Access/privacy/system constraints and accuracy/language reliability
Conversation practice	76	46.1	[35.3, 57.2]	Pedagogical sustainability
Writing correction	31	38.7	[23.7, 56.2]	Accuracy/language reliability
Reading support	15	33.3	[15.2, 58.3]	Accuracy/language reliability
Vocabulary learning	31	32.3	[18.6, 49.9]	Pedagogical sustainability
Grammar explanation	27	29.6	[15.9, 48.5]	Accuracy/language reliability
Study planning	11	18.2	[5.1, 47.7]	Access/privacy/system constraints

TABLE 6 Selected learner–LLM interaction configurations.

Immediate learning function	Dominant modality	Salient affordances	Principal breakdowns
Conversation	Text + voice	Convenience, engagement, immediate feedback	Pedagogical sustainability
Vocabulary	Text	Immediate feedback, convenience, personalization	Pedagogical sustainability
Writing correction	Text	Immediate feedback, correction quality	Accuracy/language reliability
Grammar explanation	Text	Immediate feedback, personalization	Accuracy/language reliability
Reading support	Text/TTS	Convenience, personalization, comprehension	Accuracy/language reliability
Exercise generation	Text/TTS	Convenience, feedback, personalization	Access/system constraints and accuracy
Study planning	Text/API	Personalization, convenience, engagement	Access/system constraints
Pronunciation	Voice/STT	Fluency, engagement, convenience	Access/system constraints and voice/technical reliability
Translation	Text	Immediate feedback, personalization	Accuracy/language reliability
Memory customization	Text/custom agent	Personalization, convenience	Access/system constraints and pedagogical sustainability

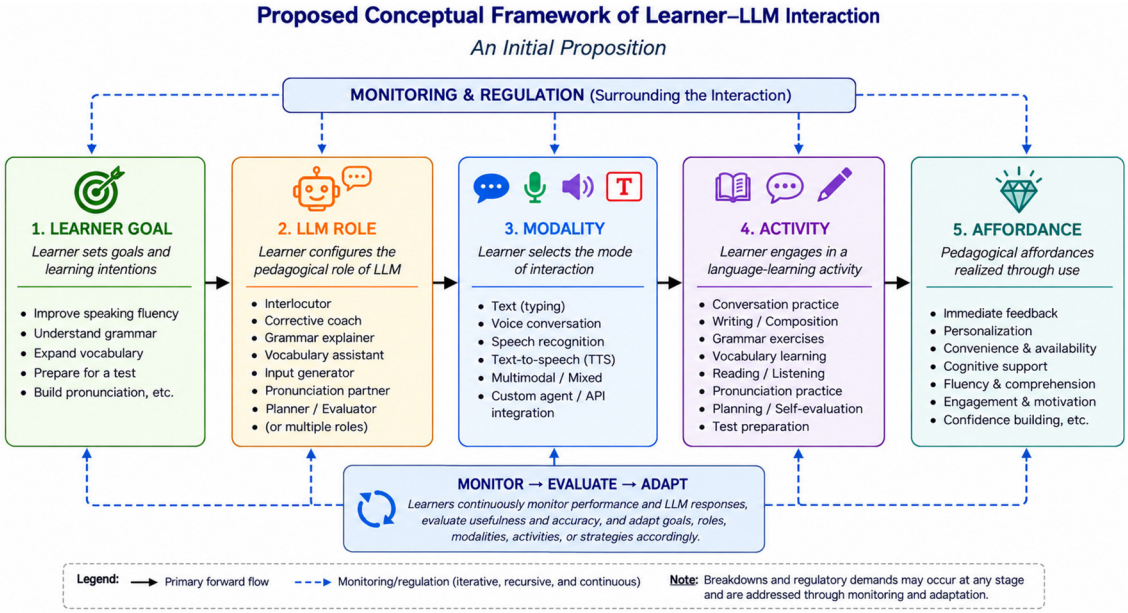


FIGURE 1
Proposed conceptual framework of learner–LLM interaction: an initial proposition.

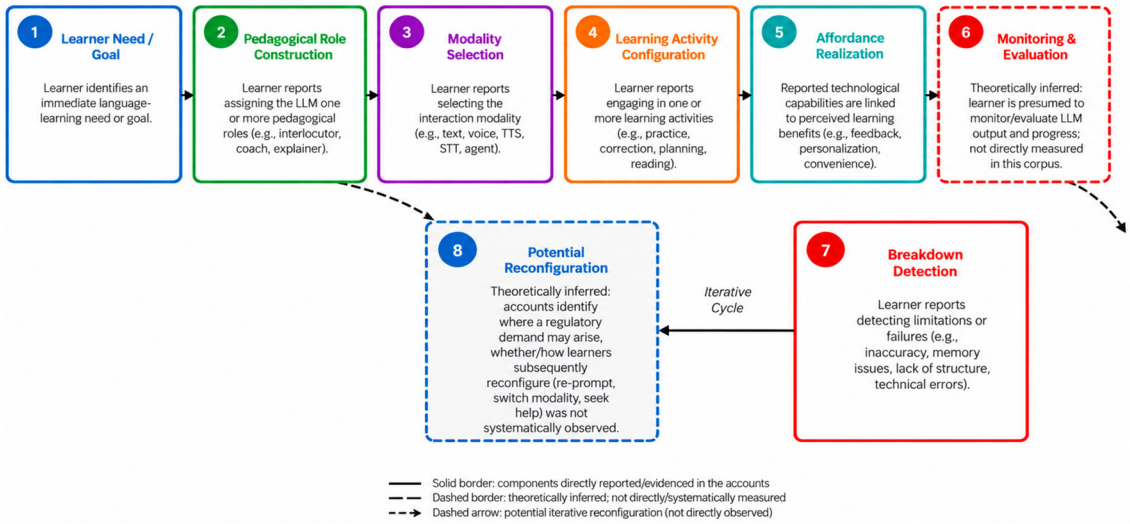


FIGURE 2
Learner-LLM configuration and regulatory demand process: from chatbot to language-learning partner. The framework is best interpreted as recursive rather than linear. Learner–LLM interaction can potentially be reconfigured when a particular combination of role, modality, activity, or affordance does not provide the expected pedagogical value. The empirical contribution of the framework is therefore not the claim that all learners complete every regulatory stage, but the identification of the points within LLM-assisted language learning at which regulatory practices and regulatory demands emerge.

The subsequent stage involves monitoring and breakdown detection. Nearly half of the firsthand accounts identified at least one constraint, and these constraints varied systematically according to pedagogical function. Accuracy, contextual memory, pedagogical structure, interaction quality, and voice reliability therefore represent conditions under which continued use may require learner evaluation rather than passive acceptance of LLM output. Reported breakdowns identify points at which regulatory demands may arise within learner–LLM interaction. The corpus, however, did not provide sufficient systematic evidence to quantify specific adaptive responses such as verification, reprompting, modality switching, or seeking human assistance; these should therefore not be interpreted as empirically measured self-regulatory outcomes.

4.8 Overall results synthesis

Across the five research objectives, the findings reveal a progression from isolated chat bot use toward a more configurable form of learner–LLM language-learning partnership. Three empirical transitions summarize this progression.

First, learners moved from single-purpose tool use toward multifunctional pedagogical role construction. Nearly two thirds of firsthand accounts contained multiple pedagogical roles, indicating that learners frequently configured the same general-purpose technology as interlocutor, coach, explainer, lexical assistant, scaffold, and learning organizer. Second, this role construction was modality dependent: text interaction predominantly supported form-focused, explanatory, and corrective functions, whereas voice interaction specialized in conversational and pronunciation-oriented activities. Third, more orchestrated configurations were associated with greater perceived affordance realization, particularly personalization and convenience; the accompanying difference in activity breadth did not survive a non-circular sensitivity analysis and is not treated as independent evidence for the orchestration construct.

At the same time, these affordances were accompanied by function-specific breakdowns. Correction-oriented activities were principally associated with accuracy/language-reliability concerns; conversational practice was most frequently associated with pedagogical-sustainability problems; pronunciation practice combined access/system constraints with voice/technical reliability problems; and memory customization showed particularly high concern rates involving access/system constraints and pedagogical sustainability.

Collectively, the findings indicate that the distinctive feature of self-directed LLM-assisted language learning is not simply access to generative AI, but the learner's role in configuring the conditions under which the technology becomes pedagogically useful. The theoretical implications of this pattern, including its relationship to self-regulated learning, are developed in the Discussion.

5 Discussion

The first major finding is that learners did not engage with LLMs as tools with a single predetermined instructional function; rather, they constructed multiple pedagogical roles around their immediate language-learning needs. Although the interlocutor role was the most prevalent, learners also positioned LLMs as language explainers, corrective coaches, lexical coaches, input scaffolds, learning managers/customizers, practice designers/evaluators, and pronunciation coaches. More importantly, the frequent co-occurrence of these roles indicates that learner–LLM interaction

is inherently multifunctional rather than organized around isolated uses. This finding extends previous research showing that generative AI can support conversational practice, corrective feedback, grammar explanation, vocabulary development, customized input, pronunciation practice, and learning planning (Marquez-Carpintero et al., 2025; Ye et al., 2025). Whereas such studies largely document the pedagogical functions that LLMs *can* perform, the present findings suggest a further distinction: the pedagogical identity of a general-purpose LLM is partly constructed by the learner through interaction. This differentiates LLM-supported learning from more functionally bounded CALL environments, because learners can move the same system between interlocutor, coach, explainer, scaffold, and learning-management roles as their immediate purposes change. This pattern is also consistent with the broader conception of digital learner agency as the strategic selection and use of technological resources in relation to individual learning goals (Khadawardi, 2025; Rowberry, 2022). With LLMs, however, such agency appears to extend beyond selecting a resource to configuring the pedagogical role that the resource assumes. Pedagogical role construction may therefore represent a defining mechanism in the transition from a responsive chatbot to a personalized language-learning partner.

The second major finding concerns the relationship between interaction modality and pedagogical function. Voice-only interaction was significantly more strongly associated with conversation and pronunciation practice, whereas text-based interaction showed a stronger association with vocabulary learning and was descriptively prominent across writing correction, grammar explanation, and reading support. These findings suggest that modality is not simply an interface through which the same pedagogical process occurs; rather, it helps shape the type of language-learning activity that becomes salient. This interpretation is consistent with multimodal approaches to language learning, which emphasize that text, audio, speech, visual information, and other representational modes can make different contributions to language processing, comprehension, communication, and learner engagement (Al-Jarf, 2024). Speech-enabled systems may particularly facilitate interactive and spoken-language practice by approximating conversational exchange (Wan and Moorhouse, 2025), while text-based interaction allows learners greater opportunity to inspect, revisit, and manipulate linguistic forms. The present findings therefore support the concept of modality–pedagogy alignment: the pedagogical usefulness of an interaction mode depends partly on its correspondence with the linguistic and cognitive demands of the learning activity. Text may be especially compatible with activities requiring reflection, noticing, vocabulary work, grammatical analysis, and written correction, whereas voice and speech technologies may be more compatible with spontaneous production, pronunciation, fluency, and conversational practice. This does not imply that one modality is intrinsically superior or that particular activities can occur only through one mode. Rather, it suggests that multimodality becomes pedagogically meaningful when learners align available interaction modes with particular learning purposes. Consequently, the increasing multimodality of LLMs should be evaluated not simply in terms of how many modes a system provides, but in terms of how effectively those modes can be configured around the demands of specific language-learning activities.

A third important finding is the distinction between isolated task-level use and more complex configurations in which learners combined planning, customization, resource recommendation, progress-related functions, and multiple language-learning activities. Records displaying such orchestration were associated with a greater breadth of perceived benefits, with particularly strong associations with personalization and convenience; a sensitivity analysis removing the four activities used to define orchestration found that the accompanying difference in activity breadth was substantially attenuated and no longer statistically significant ($M = 2.25$ vs. 1.90 non-defining activities; Mann–Whitney $U = 1{,}319.5$, $p = 0.154$, rank-biserial $r \approx 0.18$), so broader activity breadth is not treated here as independent evidence for the orchestration construct. This pattern extends established conceptions of self-directed digital language learning, in which learner autonomy involves setting goals, strategically selecting resources, organizing learning opportunities, and monitoring learning activities (Khadawardi, 2025; Rowberry, 2022). In conventional digital environments, such autonomy has largely involved navigating among external resources—such as applications, websites, videos, courses, and online communities—to assemble an individualized learning environment (Dizon, 2023; Noori, 2025). LLMs appear to introduce an additional form of agency because learners can configure the behavior of a single resource by changing its role, prompts, activities, feedback, level of personalization, and potentially its persistent settings. The present findings are conceptually consistent with a continuum from task-level use → multi-task use → orchestrated learning configuration, which can be conceptualized as AI-mediated learner orchestration, though this should be read as a conceptual interpretation rather than an independently validated empirical progression. At the more orchestrated end of this continuum, the LLM functions less as an on-demand answer generator and more as a configurable component of a broader self-directed learning system. Importantly, this interpretation should not be equated with evidence that these learners possessed higher psychological self-regulation. Self-regulated learning involves forethought, strategy use, monitoring, evaluation, and adaptation (Zimmerman and Schunk, 2011), none of which was measured psychometrically in the present corpus. Rather, the findings reveal reported AI-mediated regulatory practices that are compatible with SRL theory and suggest that learner autonomy in LLM environments increasingly involves not only selecting learning resources but also organizing and configuring how those resources behave.

The fourth theme concerns the nature of pedagogical affordances. The results showed that perceived benefits were systematically associated with particular learning configurations rather than distributed uniformly across LLM use. Writing correction was associated with correction quality, grammar explanation with immediate feedback, reading support with comprehension, vocabulary learning with vocabulary gain, and planning and memory customization particularly with personalization. These patterns reinforce the theoretical distinction between a technological capability and a pedagogical affordance. A capability describes what an LLM can technically perform, whereas an affordance emerges when that capability becomes actionable and pedagogically meaningful in relation to a learner's goal, activity, and context. This interpretation is consistent with

previous research identifying immediate feedback, personalization, conversational responsiveness, vocabulary support, engagement, and flexible access as important opportunities associated with AI-supported language learning (Nguyen and Doan, 2025). The present study extends this literature by showing that these affordances appear to be realized through particular configurations of learner purpose, pedagogical role, modality, and activity. For example, the capacity to generate language does not itself constitute vocabulary-learning support; that capability becomes a pedagogical affordance when the learner configures the LLM as a lexical coach and uses it for definitions, examples, practice, or feedback. Similarly, personalization becomes meaningful when learners configure planning, memory, input, or feedback functions around their individual needs. The findings therefore support a relational rather than technologically deterministic account of LLM affordances, represented conceptually as Learner Goal × LLM Role × Modality × Activity → Affordance Realization. In this sense, pedagogical value resides neither exclusively in the learner nor in the LLM but emerges from the configuration of the learner–technology relationship.

The fifth theme arises from the finding that perceived affordances were accompanied by function-specific breakdowns. Almost half of the firsthand accounts contained at least one concern, including problems involving interactional continuity, pedagogical sustainability, language accuracy, access, and voice or technical reliability. More importantly, the type of breakdown frequently corresponded to the pedagogical function on which the learner depended. Writing correction and grammar explanation were relatively frequently associated with accuracy-related problems within this exploratory breakdown; pronunciation practice was relatively frequently associated with speech and voice limitations; conversational practice encountered continuity and structural problems; and memory customization, despite its strong association with personalization, showed substantial context-memory and pedagogical-sustainability concerns, though this pattern is based on a small subgroup ($n = 7$) and should be read as exploratory. These findings are consistent with previous reports of hallucination and unreliable output (Lau, 2024), inappropriate language correction (Ye et al., 2025), weak pedagogical sequencing (Qin et al., 2025). However, the present study suggests that these limitations should not simply be cataloged separately from LLM affordances. Instead, they reveal an affordance–breakdown duality: the capability from which pedagogical value is derived can simultaneously create a corresponding point of vulnerability. Immediate feedback can create dependence on feedback accuracy; personalization can depend on contextual and system continuity; voice interaction depends partly on the availability and reliability of speech functionality; and open-ended flexibility may create challenges of structure and progression. From an SRL perspective, such breakdowns are important because they create regulatory demands requiring learners to monitor output, judge its relevance and reliability, and recognize when an interaction may no longer support the intended learning goal (Torre and Daley, 2023; Zimmerman and Schunk, 2011). Thus, a central theoretical proposition emerging from the findings is that the pedagogical affordances of LLMs generate corresponding regulatory demands. The value of an LLM-supported learning configuration therefore depends not only on the affordance it provides but also on the learner's capacity to recognize and manage the vulnerabilities associated with that affordance.

Taken together, the findings suggest that LLM-assisted foreign-language learning can be understood as an AI-mediated regulatory process in which learners potentially regulate not only their own learning but also aspects of the technological environment through which learning occurs. Traditional SRL theory conceptualizes learning as a recursive cycle involving forethought, performance, monitoring, self-reflection, and subsequent adaptation (Torre and Daley, 2023; Zimmerman and Schunk, 2011). The present findings suggest that LLM-supported learning introduces additional objects of regulation within this cycle. Learners may establish an immediate learning purpose, assign the LLM a pedagogical role, select an interaction modality, configure a learning activity, and experience particular affordances; the resulting interaction can then require monitoring not only of one's own language performance but also of the accuracy, relevance, continuity, and pedagogical usefulness of AI-generated support. When a breakdown is encountered or recognized, it may create a regulatory demand requiring reconsideration of the existing configuration. The present data identify where such demands may arise but do not establish whether learners recognized each breakdown or how they subsequently regulated their behavior. This extends the conventional SRL sequence toward a learner–LLM process represented as Goal → Role Construction → Modality/Activity Configuration → Affordance Realization → Monitoring/Evaluation → Breakdown Detection → Regulatory Demand → Potential Reconfiguration. The framework therefore complements rather than replaces established SRL theory: the familiar processes of planning, monitoring, evaluation, and adaptation remain central, but the AI-supported environment itself becomes something that may need to be configured and evaluated. This interpretation is also more appropriately described as AI-mediated regulatory orchestration than as human–AI co-regulation, because the present data do not establish that the LLM acts as an autonomous regulatory agent. Furthermore, the corpus did not directly measure metacognition, motivation, SRL capacity, or systematic adaptive responses following breakdowns. Accordingly, the contribution is not evidence that LLM users are more self-regulated, but an empirically informed account of where regulatory practices and regulatory demands arise within learner–LLM interaction. This distinction is theoretically important: the evolution from chatbot to language-learning partner appears to depend not simply on increasingly capable AI, but on the learner's capacity to construct pedagogical roles, align modalities and activities, realize affordances, monitor system performance, and potentially reconfigure the interaction as learning needs and technological conditions change.

6 Conclusion

This study examined how foreign-language learners configure LLMs as self-directed learning resources and how pedagogical roles, interaction modalities, learning activities, affordances, and breakdowns combine within learner–LLM interaction. The findings show a progression from single-purpose chatbot use

toward a more multifunctional language-learning partnership. Learners configured LLMs as interlocutors, corrective coaches, language explainers, lexical coaches, input scaffolds, pronunciation coaches, practice designers, and learning managers. This role construction was also modality dependent, with text supporting a broader range of form-focused activities and voice being particularly associated with conversation and pronunciation. More orchestrated configurations were associated with greater perceived personalization and convenience, while the affordances realized through LLM use were accompanied by function-specific breakdowns. Collectively, these findings support conceptualizing learner–LLM interaction as an AI-mediated regulatory process rather than a simple relationship between technology use and learning outcomes.

The study contributes to understanding LLM-assisted language learning by proposing that learner autonomy in AI-supported environments extends beyond resource selection to resource configuration. Learners actively shape the pedagogical role of the LLM, align interaction modalities with particular activities, and realize different affordances through these configurations. The findings further suggest an affordance–breakdown duality, whereby capabilities associated with perceived pedagogical value may also create corresponding vulnerabilities and potential regulatory demands. These relationships are integrated into a recursive framework moving from learner goal → role construction → modality/activity configuration → affordance realization → monitoring/evaluation → breakdown detection → regulatory demand → potential reconfiguration. The framework therefore extends established SRL perspectives by identifying the AI-supported learning environment itself as an additional object of learner configuration, monitoring, and evaluation.

The findings suggest that effective integration of LLMs into foreign-language learning requires more than providing learners with access to generative AI. Language education should develop learners' capacity to purposefully configure LLMs for particular learning goals, select modalities that correspond to specific linguistic activities, critically evaluate AI-generated explanations and corrections, and recognize when technological or pedagogical limitations require alternative resources or human support. AI literacy for language learning should therefore incorporate regulatory competence, including purposeful prompting, monitoring output quality, evaluating feedback, managing personalization, and maintaining appropriate pedagogical structure. Such preparation may help learners move from unstructured chatbot use toward more purposeful and critically monitored language-learning partnerships.

Several limitations delimit these findings. The corpus comprised a purposive and platform-uneven sample of public self-reports, with forums accounting for most records and Reddit and YouTube absent; consequently, the results cannot be generalized as prevalence estimates for the broader population of language learners. The corpus is additionally subject to voluntary-publication bias: learners who choose to discuss their experiences publicly may differ from those who use LLMs without posting about them, and particularly salient positive or negative experiences may be more likely to be reported. Platform composition may compound this bias: contributors drawn from LingQ and Readlang are members of dedicated language-learning communities and may differ systematically from LLM users more generally. The findings therefore characterize patterns within publicly expressed experiences rather than the distribution of experiences among LLM-assisted language learners generally. The study examined reported experiences rather than directly observed learner behavior, and perceived usefulness or reported improvement cannot establish objective language-learning effectiveness. In addition, psychological self-regulation, metacognition, motivation, and adaptive responses following breakdowns were not directly measured, while automated translation and LLM-assisted coding may have introduced some semantic or interpretive error. Future research should therefore test the proposed framework using larger cross-platform datasets and complementary process-based methods, including screen recordings, interaction logs, and think-aloud protocols, alongside validated SRL measures and objective proficiency outcomes. Such designs could directly examine how reported configurations correspond to actual learner behavior and how learners respond to breakdowns through verification, reprompting, modality switching, reconfiguration, or human support. Lastly, although cross-model reproducibility was assessed using an independently coded stratified sample, no independent human coder evaluated the same records. Consequently, the reported agreement statistics demonstrate consistency across two LLM-based coding systems rather than agreement with expert human judgments. LLM-assisted coding and translation may therefore have introduced systematic interpretive biases that cross-model agreement alone cannot detect. Future research should incorporate independent human coding of a stratified subset to provide an additional criterion for evaluating automated classifications.

Ultimately, the transition from chatbot to language-learning partner appears to depend not simply on increasingly sophisticated AI capabilities, but on how learners configure those capabilities for pedagogical purposes. LLMs become meaningful learning partners when learners construct appropriate pedagogical roles, align modalities with learning activities, realize relevant affordances, and critically monitor the quality and limitations of the resulting interaction. Thus, the future pedagogical value of LLMs in self-directed foreign-language learning may depend as much on learners' capacity to configure, evaluate, and regulate their interaction with AI as on what the technology itself is capable of doing.

Data availability statement

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

Author contributions

OL: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Writing – original draft, Writing – review & editing. MG: Conceptualization, Data curation, Funding acquisition, Investigation, Project administration, Supervision, Validation, Writing – review & editing. BM: Conceptualization, Data

curation, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – review & editing.

Funding

The author(s) declared that financial support was not received for this work and/or its publication.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was used in the creation of this manuscript. Generative AI tools were used to support structured coding, multilingual translation, and data-processing quality control. All analytical decisions, interpretation of findings, and final manuscript content remained the responsibility of the author(s).

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyg.2026.1995850/full#supplementary-material

References

Albedah, F. (2025). Artificial intelligence in language education: a systematic review of multilingual applications, large language models, and emerging challenges. *Lang. Teach. Res. Q.* 49, 247–268. doi: 10.32038/ltrq.2025.49.13

Al-Jarf, R. (2024). Multimodal teaching and learning in the EFL college classroom. *J. Engl. Lang. Teach. Appl. Linguist.* 6, 68–75. doi: 10.32996/jeltal.2024.6.4.7

Alyahyan, E., Ada, M. B., and Lever, J. (2025). "A systematic mapping of large language models as feedback provider in higher education," in *2025 IEEE international conference on advanced learning technologies (ICALT)*, (Los Alamitos, CA: IEEE Computer Society), 261–262. doi: 10.1109/ICALT64023.2025.00082

Beaven, A. (2021). *Learning without a Teacher: Self-Directed Language Learning in the Digital Wilds*. Dublin: Research-Publishing. Net.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Bertolini Mendes, G., and Panhan, A. M. (2024). ChatGPT as a foreign language assistance and learning tool. *Int. J. Hum. Sci. Res.* 4, 2–7. doi: 10.22533/at.ed.558422415019

Cavojský, M., Bugár, G., Kormaník, T., and Hasin, M. (2023). "Exploring the capabilities and possible applications of large language models for education," in *2023 21st international conference on emerging elearning technologies and applications (ICETA)*, (New York, NY: IEEE), 91–98. doi: 10.1109/ICETA61311.2023.10344166

Chik, A. (2017). "Learner autonomy and digital practices," in *Autonomy in language learning and teaching: New research agendas*, eds. A. Chik, N. Aoki, & R. Smith (London: Springer), 73–92. doi: 10.1057/978-1-137-52998-5_5

Daberao, L. (2025). "A context-aware intelligent tutoring framework for grammar modeling using large language models," in *2025 IEEE conference on pervasive and intelligent computing (PICom)*, (New York, NY: IEEE), 200–204. doi: 10.1109/PICom68402.2025.00034

Dentella, V., Guenther, F., and Leivada, E. (2025). Language *in vivo* vs. *In silico: size matters but larger language models still do not comprehend language on a par with humans due to impenetrable semantic reference. PLoS ONE* 20:e0327794. doi: 10.1371/journal.pone.0327794

Dizon, G. (2023). Foreign language learning in the digital wilds. *CALICO J.* 40, 238–256. doi: 10.1558/cj.22360

Du, Y., and Reynolds, B. L. (2025). Rooted in and beyond interaction: a systematic review of interactive affordances of chatbots for language learning amidst the rise of large language models. *Acta Psychol.* 259:105307. doi: 10.1016/j.actpsy.2025.105307

Fomina, T. G. (2022). Conceptual approaches to the analysis of self-regulated learning in foreign psychology. *J. Mod. Foreign Psychol.* 11, 27–37. doi: 10.17759/jmfp.2022110303

Godwin-Jones, R. (2019). Riding the digital wilds: learner autonomy and informal language learning. *Lang. Learn. Technol.* 23, 8–25. doi: 10.64152/10125/44667

Guerreiro, N. M., Alves, D. M., Waldendorf, J., Haddow, B., Birch, A., Colombo, P., et al. (2023). Hallucinations in large multilingual translation models. *Trans. Assoc. Comput. Linguist.* 11, 1500–1517. doi: 10.1162/tacl_a_00615

Hsieh, H.-F., and Shannon, S. E. (2005). Three approaches to qualitative content analysis. *Qual. Health Res.* 15, 1277–1288. doi: 10.1177/1049732305276687

Haines, K. (2017). *Contextualising the learning affordances of technology: an in-depth look at the developing practice of two modern language teachers.* Occasional and discussion paper series, 2017(6). ISSN 2324-3635. Available online at: http://www.unitec.ac.nz/epress. doi: 10.34074/ocds.62017 (Accessed October 20, 2026).

Khabirova, Z. (2025). *Teaching English With New Methodologies And Multimodal Technologies. Collection of Scientific Papers ≪SCIENTIA≫*, Dallas, TX: NGO International Center of Scientific Research / Primedia eLaunch LLC.

Khadawardi, H. A. (2025). Autonomous digital language learning strategies of EFL students beyond the classroom. *AWEJ* 16, 493–514. doi: 10.24093/awej/vol16no4.27

Kim, A., and Su, Y. (2024). How implementing an AI chatbot impacts Korean as a foreign language learners' willingness to communicate in Korean. *System* 122:103256. doi: 10.1016/j.system.2024.103256

Koç, F. Ş., and Savaş, P. (2025). The use of artificially intelligent chatbots in English language learning: a systematic meta-synthesis study of articles published between 2010 and 2024. *ReCALL* 37, 4–21. doi: 10.1017/S0958344024000168

Kohnke, L., Moorhouse, B. L., and Zou, D. (2023). ChatGPT for language teaching and learning. *RELC J.* 54, 537–550. doi: 10.1177/00336882231162868

Krippendorff, K. (2018). *Content analysis: An introduction to its methodology.* Sage publications. Available online

at: https://books.google.cl/books?hl=esandlr=andid=nE1aDwAAQBAJandoi=fndandpg=PP1anddq=Krippendorff,+K.+(2013).+Content+Analysis:+An+Introduction+to+Its+Methodology+(3rd+ed.).+Sage+Publications.andots=y_9eUofQ7xandsig=2JB5wS74lTlkKzUOib6bhKnDXIw (Accessed October 20, 2026).

Lau, M. (2024). Is ChatGPT taking over the language classroom? How language ideologies of large language models impact teaching and learning. *Work. Pap. Appl. Linguist. Linguist. York* 4, 1–11. doi: 10.25071/2564-2855.36

Lowin Kropf, B. (2025). "Besser sprechen dank KI!" *DaF/DaZ Im Wandel (Gesamtschweizerische DaF/DaZ-Tagung: Tagungsakten* 2024), (Bern: AkDaF and Ledafids), 61–70. doi: 10.51363/unifr.lrp.cqhc9q.05

Mahlow, C. (2023). large language models and artificial intelligence, the end of (language) learning as we know it—or not quite. *Int. J. Econ. Manag. Syst.* 7, 1–14. doi: 10.35542/osf.io/da2rm

Mai, N. T., Cao, W., and Fang, Q. (2025). A study on how LLMs (eg GPT-4, chatbots) are being integrated to support tutoring, essay feedback and content generation. *J. Comput. Electron. Inf. Manag.* 18, 43–52. doi: 10.54097/6r6yhn67

Marchisio, K., Ko, W.-Y., Bérard, A., Dehaze, T., and Ruder, S. (2024). Understanding and mitigating language confusion," in LLMs *proceedings of the 2024 conference on empirical methods in natural language processing*, (Miami, FA: Association for Computational Linguistics), 6653–6677.

Marquez-Carpintero, L., Lopez-Sellers, A., and Cazorla, M. (2025). Simulation of teaching behaviours in intelligent tutoring systems: a review using large language models. *Artif. Intell. Rev.* 59:56. doi: 10.1007/s10462-025-11464-8

Nguyen, Q. N. (2022). "Affordance-based learning trajectories: a deviation from the teachers-should-know-all model in language education," in *1st International conference on education 2022, Thailand, Vol.* 132, 131.

Nguyen, Q. N., and Doan, D. T. H. (2025). Idiographic self-regulated affordance uptake in AI-mediated language learning. *Cogent Educ.* 12. doi: 10.1080/2331186X.2025.2581414

Noori, A. (2025). Digital technologies and self-regulated language learning: a systematic review. *J. Soc. Sci. Kabul Univ.* 8, 243–263. doi: 10.62810/jss.v8i1.323

Pipattarasakul, P., and Singhasiri, W. (2018). Metastrategies employed by science and engineering EFL learners in a speaking task. *J. Asia TEFL* 15:66. doi: 10.18823/asiatefl.2018.15.1.5.66

Qin, Y., Zhou, Y., Huang, X., and Huricha, B. (2025). Empowering foreign language education with large language models: applications and insights. *J. High. Educ. Teach.* 2, 100–101. doi: 10.62517/jhet.202515114

Rowberry, J. (2022). Agency in and beyond the classroom: learners' selection of resources for self-directed language learning. *J. Learner Dev.* 6, 51–66. doi: 10.37546/JALTSIG.LDJ6-4

Sadigzade, Z. (2025). Generative AI in call: enhancing self-regulated L2 learning through adaptive chatbots. *Glob. Spectr. Res. Humanit.* 2, 25–45. doi: 10.69760/gsrh.0250206002

Sun, D., Xu, P., Zhang, J., Liu, R., and Zhang, J. (2025). How self-regulated learning is affected by feedback based on large language models: data-driven sustainable development in computer programming learning. *Electronics* 14:194. doi: 10.3390/electronics14010194

Torre, D., and Daley, B. J. (2023). Cyclical phases and processes of self-regulated learning. *Med. Teach.* 45, 464–465. doi: 10.1080/0142159X.2023.2186342

Tseng, W.-T., Liu, H., and Nix, J.-M. L. (2017). self-regulation in language learning: scale validation and gender effects. *Percept. Mot. Skills* 124, 531–548. doi: 10.1177/0031512516684293

Wan, Y., and Moorhouse, B. L. (2025). Using Call Annie as a generative artificial intelligence speaking partner for language learners. *RELC J.* 56, 489–498. doi: 10.1177/00336882231224813

Xabirova, Z., and Ro'zimov, U. (2025). Multimodal integratsiya va sun'iy intellekt orqali maxsus maqsadlar uchun ingliz tilini (ESP) QAYTA KO 'RIB *CHIQISH*. 1, 170–175. . Available online at: https://d-pressa.com/index.php/spc/article/view/202

Xu, P., Yin, J., Wang, H., Miao, Y., and Zhang, Y. (2025). A systematic review of large language models in education dual roles, user perspectives, and cross-disciplinary impacts.

Ye, J., Wang, S., Zou, D., Yan, Y., Wang, K., Zheng, H.-T., et al. (2025). Position: Llms can be good tutors in foreign language education. *arXiv E-Prints, arXiv*-2502. doi: 10.18653/v1/2025.emnlp-main.885

Zhang, A., and Jiang, C. (2026). Beyond grammar checking: the impact of llm-based instruction on writing strategies for self-regulated learning in EFL contexts. *Cogent Educ.* 13:2574333. doi: 10.1080/2331186X.2025.2574333

Zheng, S. (2024). The effects of chatbot use on foreign language reading anxiety and reading performance among Chinese secondary school students. *Comput. Educ. Artif. Intellig.* 7:100271. doi: 10.1016/j.caeai.2024.100271

Zimmerman, B. J., and Schunk, D. H. (2011). "Self-regulated learning and performance: an introduction and an overview," in *Handbook of Self-Regulation of Learning and Performance*, eds. B. J. Zimmerman and D. H. Schunk (New York, NY: Routledge), 15–26.