T I E Journal
Volume 4 Issue 3, 2026
http://ijtie.com/v403/n40373
ISSN: 2820-7521

The International Journal of Technology, Innovation, & Education

Low-Bandwidth AI Support for Chinese Teaching in Tanzanian Universities: A Design Synthesis

Jivitius Ching'oro Sabatho *[1,2], **Eliya Mbulika Ndiwu** [1,3], **Moabu Jimmy Chandafa** [1,4], **and Fang Huang** [1]

[1] School of Education, Huazhong University of Science and Technology (HUST), Wuhan, P.R. China
[2] Tanzania Institute of Education (TIE), Dar es Salaam, Tanzania
[3] Vocational Education and Training Authority (VETA), Tanzania
[4] Department of Technical Education, Mbeya University of Science and Technology (MUST), Mbeya, Tanzania

sabathojc@gmail.com

Received: August 15, 2026; **Accepted**: October 25, 2026; **Published**: October 31, 2026

Abstract

Artificial intelligence (AI) can expand oral practice and feedback in language learning, but its value depends on pedagogy, access, teacher capacity, assessment validity and governance. This conceptual design synthesis develops a low-bandwidth AI-supported pathway for beginner-to-intermediate Chinese teaching in Tanzanian universities. A bounded integrative review drew on 31 sources spanning AI-supported language learning, Mandarin pronunciation and speech recognition, listening-speaking assessment, instructional design, and Tanzanian digital-education policy. Evidence was interpreted through communicative can-do outcomes, Technological Pedagogical Content Knowledge, the Technology Acceptance Model, and human-centred AI principles. The synthesis produced six design requirements: pedagogy before platform, an offline-first core, targeted AI use for feedback-intensive practice, integrated listening-speaking tasks, human-in-the-loop assessment, and readiness before scale. These requirements were translated into a six-stage pathway, an oral-performance rubric, a 60-minute lesson shell, workload scenarios and an institutional readiness scorecard. The framework distinguishes direct Chinese-specific evidence from cross-language extrapolation and provides offline substitutes when connected AI is unavailable. It does not claim effectiveness: no field data were collected, and speech-recognition performance for Mandarin spoken with Tanzanian accents remains unestablished. The framework is intended for piloting, local validation and adaptation before scale.

Keywords: Chinese language teaching, artificial intelligence, low-bandwidth learning, speech recognition, language assessment, Tanzania

Sabatho, J. C., Ndiwu, E. M., Chandafa, M. J., & Huang, F. (2026). Low-bandwidth AI support for Chinese teaching in Tanzanian universities: A design synthesis. *The International Journal of Technology, Innovation, and Education*, *4*(3), 380–404. https://ijtie.com/v403/n40373

CC BY

1. Introduction

Artificial intelligence is moving rapidly from a specialist technology to an everyday layer of educational practice. In language education, conversational agents, automatic speech recognition (ASR), generative AI, automated feedback and learning analytics can create additional opportunities for rehearsal, feedback and interaction outside limited class time. Recent syntheses report positive potential for AI-supported second-language learning, but effects vary with task design, learner characteristics, feedback type, duration and the role of the teacher (Du & Daniel, 2024; Lai & Lee, 2024; Ngo et al., 2024; Wu, 2024). The practical issue is how AI can support language learning without displacing sound pedagogy.

That issue is especially relevant to Chinese language education. Mandarin requires learners to coordinate lexical tone, segmental pronunciation, listening discrimination, character recognition and production, vocabulary, grammar and interaction. Technology can assist repeated listening, spaced retrieval, speech rehearsal and immediate low-stakes feedback, but these functions are not interchangeable. ASR can help learners notice pronunciation problems, yet automated indicators do not fully represent communicative oral proficiency (Ngo et al., 2024; Shi et al., 2024). Mandarin tone perception and tone-word learning also require carefully sequenced practice (Li & Lee, 2021; Liu & Xiao, 2021).

Tanzania provides a useful context for this design problem. The 2023 Chinese Language Syllabus for Ordinary Secondary Education presents Chinese as a communicative subject involving listening, speaking, reading, writing, translation and interpreting and calls for learner-centred teaching, continuous assessment and appropriate use of information and communication technology (Tanzania Institute of Education [TIE], 2023). The syllabus is used only as downstream curriculum context for university learners who are preparing to become Chinese teachers. For general university elective or minor students, it is contextual rather than a direct curriculum source. The higher-education setting is grounded in the National Digital Education Guidelines for Universities and Tanzanian studies of technology adoption and readiness in higher education (Kafyulilo et al., 2016; Landa et al., 2021; Ministry of Education, Science and Technology [MoEST], 2025; Mtebe & Raisamo, 2014). Recent documentary work on Chinese curriculum reform also identifies teacher support, localisation, coordination and sustainable resourcing as continuing implementation concerns (Sabatho et al., 2026).

This article develops a pedagogy-first, low-bandwidth pathway for Chinese teaching and assessment in Tanzanian universities. It is a conceptual design-synthesis paper and reports no field data. The primary use case is beginner-to-intermediate undergraduate Chinese taught as

an elective, minor or additional subject. The framework can serve both general learners and future teachers; the school syllabus becomes directly relevant only for the latter group. No universal contact-hour or device profile is assumed because institutions differ in timetabling and learner access. The framework instead specifies the pedagogical functions technology should serve, the conditions under which those functions are defensible, and the points at which human judgement remains necessary.

To guide the study, three research questions were formulated:

- **RQ1:** What pedagogical and implementation principles should guide AI-supported Chinese teaching in resource-constrained Tanzanian higher education?
- **RQ2:** How can low-bandwidth and optional AI functions be added to pronunciation, vocabulary, listening, speaking and assessment without replacing teacher judgement?
- **RQ3:** Which readiness conditions should be checked before the pathway is scaled?

The objectives were to synthesise relevant evidence, translate it into a practical instructional architecture, and specify readiness and evaluation conditions for future piloting.

2. Literature Review

2.1. AI-Supported Second-Language Learning: Promise with Conditions

Recent AI-in-language-learning research supports careful augmentation more strongly than technological substitution. Wu's (2024) meta-analysis found an overall positive contribution of AI to second-language learning while also identifying contextual, instructional and social-emotional moderators. Reviews of conversational AI describe benefits for interaction, feedback, engagement and practice, but they also report uneven research designs and concerns about accuracy, dependency, privacy and the pedagogical quality of generated language (Du & Daniel, 2024; Lai & Lee, 2024). Kohnke et al. (2023) therefore treat generative AI as a language-learning resource whose value depends on digital competence and critical, ethical use.

These findings matter in resource-constrained settings because AI is most useful when it expands an activity learners otherwise receive too little of. Teachers cannot listen to every learner repeatedly, give individual pronunciation feedback after every attempt, or sustain one-to-one conversation across an entire class. AI can add practice attempts and formative signals. The educational aim, however, remains meaningful language use. A tool that consumes scarce data or produces feedback that teachers and learners cannot interpret may reduce effective learning time.

2.2. Pronunciation, Tones and Automatic Speech Recognition

Pronunciation is a plausible target for selective AI support because it benefits from repetition and rapid feedback. Ngo et al. (2024), synthesising 15 ASR pronunciation studies and 38 effect sizes, reported a medium overall effect and stronger outcomes when feedback was explicit and practice extended beyond very short periods. Most evidence in that meta-analysis concerns English, however. Speech recognisers can accept non-target-like production, reject acceptable accents, or signal an error without explaining how a learner should improve.

Mandarin adds lexical tone. Perceptual research shows that tone difficulty varies across tones and phonological environments and that disyllabic training can be useful (Li & Lee, 2021). Liu and Xiao (2021) show the value of connecting tone-category learning with familiar words. This supports practice that moves from contrasts to meaningful words and short phrases and routes persistent or ambiguous errors to a teacher.

Chinese-specific technology evidence is growing but remains much smaller than the English-language evidence base. Bao et al. (2025) reported improved Chinese speaking scores among 113 international learners using an AI-driven ASR dialogue platform; willingness to communicate was positively associated with speaking performance, while speaking anxiety was negative. Wu et al. (2025) found that a computer-assisted pronunciation training system with ASR-supported feedback improved Mandarin lexical-tone learning in a self-directed setting. Liu (2026) likewise examined voice-based ChatGPT with Chinese-as-a-second-language learners and reported improvement in Mandarin pronunciation over time. These studies strengthen the case for targeted pronunciation support, but they were not conducted with Tanzanian learners. Published evidence on how Mandarin ASR handles Tanzanian-accented Chinese was not identified. That uncertainty is treated as a local validation requirement and is not assumed away.

2.3. Listening, Speaking and AI-Supported Assessment

AI can widen the repertoire of listening and speaking tasks. Runge et al. (2024) show that generative AI can support the creation of interactive listening-assessment content at scale when human review remains embedded. The transferable principle is that AI can assist task generation and administration while human quality assurance protects construct relevance.

The same caution applies to automated speech scoring. Shi et al. (2024), using 956 audio clips, found meaningful relationships between automatic speech-scoring indicators and pronunciation and fluency, but those indicators were not identical to broader complexity, accuracy and fluency constructs. Automated scores can support formative screening, progress visualisation or

flagging, but they should not replace an analytic rubric for consequential decisions involving interaction, meaning, task fulfilment or diverse accents.

2.4. Spaced Practice and Micro-Learning

Low-bandwidth design is easier to sustain when learning is organised into small reusable units instead of long synchronous sessions. Kim and Webb's (2022) meta-analysis supports spaced practice over massed practice for second-language learning, although optimal spacing varies with goals and retention intervals. Short cycles of retrieval, listening discrimination, character recall and oral rehearsal can be distributed across the week, with downloaded audio and compact task files available offline.

Micro-learning here is an implementation strategy, not a claim that complex language competence can be reduced to fragments. Each short activity has a clear communicative or linguistic purpose and contributes to an integrated task. A learner may practise a tone contrast, retrieve characters, rehearse a sentence frame and then use those resources in an interaction.

2.5. Teacher Knowledge, Acceptance and Human-Centred AI

Technology integration requires pedagogical judgement as well as technical access. Technological Pedagogical Content Knowledge (TPACK) argues that effective technology use depends on the interaction of technological, pedagogical and content knowledge, not technical skill alone (Mishra & Koehler, 2006). For Chinese teaching, a teacher needs knowledge of linguistic content such as tones or character structure, the pedagogy required to teach it, and the capabilities and limits of the selected tool.

The Technology Acceptance Model (TAM) adds an implementation lens. Davis (1989) emphasises perceived usefulness and perceived ease of use as determinants of adoption. In resource-constrained settings, usefulness must include connectivity, subscription cost and setup burden. Bao et al. (2025) also found that perceived usefulness and ease of use did not predict Chinese speaking gains in their study. Acceptance is therefore relevant to sustained use, but it is not evidence of learning effectiveness.

Human-centred AI principles add ethical boundaries. UNESCO's guidance on generative AI calls for privacy protection, human agency and institutional governance (Miao & Holmes, 2023). The AI Competency Framework for Teachers identifies human-centred mindset, AI ethics, foundations and applications, AI pedagogy and professional learning as core teacher competencies (Miao & Cukurova, 2024). In the present framework, teachers remain responsible for task purpose, quality assurance, interpretation of feedback and consequential assessment decisions.

2.6. Equity and Low-Bandwidth Implementation

Digital innovation can widen access, but it can also reproduce inequality when access to devices, connectivity, electricity, accessibility features or digital literacy is uneven. UNESCO's Global Education Monitoring Report stresses relevance, equity, scalability and sustainability in technology decisions (UNESCO, 2023). Tanzania's National Digital Education Guidelines for Universities provide a policy basis for systematic digital-education planning (MoEST, 2025).

An offline-first architecture therefore keeps essential explanations, audio models, vocabulary and character practice, task instructions and assessment criteria usable without continuous connectivity. Connected AI becomes an enhancement, not the only route to a learning outcome. This does not mean offline learners receive identical feedback speed or personalisation. Peer checking against downloaded models, self-record-and-compare routines and periodic teacher audio comments can reduce the gap, but they do not fully reproduce immediate ASR feedback. That residual difference should be measured during piloting.

Tanzanian educational-technology research reinforces this contextual logic. Mtebe and Raisamo (2014) identified unreliable connectivity, uncertainty about resource quality and copyright awareness as barriers to open-resource adoption in higher education. Kafyulilo et al. (2016) found that electricity, time, class size, tools and management support affected continued technology use. Landa et al. (2021) reported relatively low integration of innovative teaching and learning technologies in two Tanzanian public universities and linked readiness for organisational change with implementation. Sustainable AI integration therefore depends on infrastructure, institutional support, teacher readiness and manageable access modes.

The evidence base remains uneven. Direct studies of AI-supported Chinese speaking and pronunciation are now available (Bao et al., 2025; Liu, 2026; Wu et al., 2025), while much of the broader evidence on ASR, conversational AI and automated scoring comes from English-language settings. Table 1 therefore labels cross-language applications as indirect or extrapolated and does not present them as direct evidence for Tanzanian Mandarin learners.

3. Materials and Methods

3.1. Design Purpose and Target Setting

The study used a structured integrative review and conceptual design-synthesis approach. Integrative reviews are appropriate when findings from different forms of literature are connected to generate a conceptual or practical model instead of a pooled effect estimate (Snyder, 2019; Torraco, 2016). The intended use context is beginner-to-intermediate

undergraduate Chinese taught as an elective, minor or additional subject in Tanzanian universities. The framework does not assume Chinese-major status, a fixed weekly contact-hour model, or universal personal-device ownership. The lesson example uses one 60-minute guided session plus brief asynchronous practice. Core activities require access to audio playback and recording, which may be through personal or shared devices; connected AI remains optional. To avoid presenting one class size as typical, the workload sensitivity analysis in Table 4 shows 20-, 40- and 60-learner scenarios with two short oral samples per learner per week.

Tanzanian university learners are likely to be multilingual, using Kiswahili and English alongside other local languages. Standard Kiswahili is generally described as non-tonal, whereas some Tanzanian Bantu languages, including Sukuma, use lexical and grammatical tone (Contini-Morava, 1997; Richardson, 1956). Mandarin-tone difficulty may therefore vary with learners' language histories. Because no learner sample was collected, this is treated as a pilot hypothesis, not a uniform transfer pattern. ASR performance for Mandarin spoken with Tanzanian accents also remains an empirical question for local testing.

3.2. Search Strategy and Source Selection

Evidence was assembled in four domains: (a) AI-supported second-language learning and conversational AI; (b) ASR, pronunciation and Mandarin tone learning; (c) technology-mediated listening, speaking and assessment; and (d) digital-education, curriculum and AI-governance guidance relevant to Tanzania. Recent empirical and review literature was prioritised for January 2019 to 12 August 2026. Earlier foundational sources on technology adoption, teacher knowledge and instructional design were not time-limited, and official policy/curriculum documents were included according to current relevance.

Targeted searches were conducted through ScienceDirect, SpringerLink, Cambridge Core, SAGE Journals, Frontiers, Wiley Online Library, John Benjamins, UNESCO, the Council of Europe, and Tanzanian education websites. Reference chaining was used to locate relevant sources outside those platforms, including foundational work and studies published in journals such as Language Learning and MIS Quarterly. Search combinations included: ('artificial intelligence' OR chatbot OR 'generative AI') AND ('second language' OR language learning) AND (speaking OR pronunciation OR feedback); ('automatic speech recognition' OR ASR OR 'speech recognition') AND (Mandarin OR Chinese OR 'Chinese as a second language') AND (pronunciation OR speaking OR tone); ('listening assessment' OR 'speaking assessment') AND (AI OR automated OR interactive); and (technology OR digital OR AI) AND (Tanzania OR 'higher education') AND (access OR adoption OR teacher OR policy).

Sources were retained when they informed at least one design decision about pedagogy, feedback, assessment, teacher roles, adoption, access or governance. Sources focused mainly on technical model performance without an educational application, promotional claims without evidence, or contexts too remote to inform a transferable design principle were excluded. The corresponding author coordinated the criterion-based relevance check; the process was not independently double-screened. The article cites 31 sources in total. Twenty-three directly informed the evidence-to-design matrix and are represented in Supplementary Table S1; eight additional sources provide background, linguistic, theoretical, policy or discussion context. The review was designed as a purposive integrative synthesis, so platform hit counts and a complete found/screened denominator were not logged during the original search. Multidisciplinary bibliographic indexes such as Scopus, Web of Science and ERIC were not systematically searched, which may have biased retrieval toward the publisher platforms and sources reached through reference chaining. Retrospective counts are not reported because they cannot now be reconstructed reliably.

3.3. Analytical Lenses

Four lenses were used to read the evidence. First, communicative can-do outcomes from the Common European Framework of Reference for Languages (CEFR) Companion Volume provided an action-oriented way to state what learners should be able to do, without claiming that Tanzanian Chinese courses are formally equivalent to CEFR levels (Council of Europe, 2020). Second, TPACK tested whether a technology served a specific content and pedagogical need. Third, TAM provided a check on practical usefulness and ease of use. Fourth, human-centred AI principles defined ethical and governance boundaries.

Together these lenses produced a decision rule: an AI function should support a defined learning outcome, add value over a lower-technology alternative, remain usable under local resource conditions, and preserve human oversight where validity or learner welfare is at stake. This rule was used as an analytical filter, not an empirical finding.

3.4. Design Matrix and Contextual Stress Test

Evidence statements were entered into a design matrix with four fields: pedagogical affordance, limitation or risk, resource implication, and human responsibility. Supplementary Table S1 makes this matrix visible through source clusters. Recurring statements were then grouped into candidate design requirements.

The corresponding author applied five Tanzanian implementation checks to each candidate requirement: (1) Can the core learning outcome still be reached through an offline or low-data route? (2) Is the teacher role feasible where specialist time is scarce? (3) Does the design preserve assessment validity and human judgement? (4) Does it fit university-level digital policy and the school-level curriculum context relevant to future teachers? and (5) Are privacy, accessibility, support and recurrent-cost risks visible and mitigable? The 2023 secondary-school Chinese syllabus was used only for downstream alignment where future-teacher preparation was relevant; higher-education feasibility was checked against MoEST's university digital-education guidance and Tanzanian higher-education technology studies. This was a conceptual stress test and was not independently replicated or field-validated.

3.5. Ethics and Nature of Evidence

The evidence base consisted only of published literature and publicly available policy and curriculum documents; no primary data were collected. The workload scenarios and readiness thresholds below are design heuristics for planning and piloting, not empirical findings or validated cut-off scores. The purposive search and author-led relevance checking may have introduced selection bias, which is acknowledged in the conclusion.

4. Design Requirements

The synthesis produced six interdependent requirements. Table 1 maps the main evidence to each requirement and labels its strength for the present context. Three patterns cut across the table: learning outcomes should precede technology choice; access constraints require an offline core instead of an online-only course; and automated feedback should not displace teacher judgement when assessment consequences increase. The following subsections explain the practical meaning of each requirement.

4.1. Requirement 1: Pedagogy before Platform

The learning outcome should precede the technology choice. A course can begin with a communicative task such as introducing oneself, asking for directions, completing a simple service encounter, describing a routine or interpreting a short classroom exchange. The teacher identifies the vocabulary, characters, sentence patterns, listening discrimination and pronunciation features needed for that task before selecting any AI function.

This sequence avoids designing lessons around whatever a tool happens to do. A chatbot may be useful for generating additional role-play turns, but paired practice may already meet the outcome. ASR may help when learners need many pronunciation attempts, but it should not

score an interactional task that depends on meaning negotiation, appropriateness and partner response.

Table 1. *Evidence-to-Design Logic and Evidence Strength for the Conceptual Pathway*

Design requirement	Evidence base	Evidence strength	Tanzanian design implication
1. Pedagogy before platform	AI effects vary by task/context; TPACK emphasises fit (Mishra & Koehler, 2006; Wu, 2024).	Indirect/theoretical	Define outcomes first; add AI only where it improves practice, feedback or access.
2. Offline-first core	Technology can reinforce inequality; Tanzanian studies identify connectivity, electricity and readiness constraints (Kafyulilo et al., 2016; Landa et al., 2021; Mtebe & Raisamo, 2014; UNESCO, 2023).	Contextual/indirect	Keep essential content and assessment routes usable offline; acknowledge slower or less personalised feedback without AI.
3. Targeted AI for feedback-intensive practice	ASR evidence is strong in English; Chinese-specific studies show promising speaking / tone / pronunciation gains (Bao et al., 2025; Liu, 2026; Ngo et al., 2024; Wu et al., 2025).	Direct Chinese + indirect L2	Prioritise pronunciation rehearsal and low-stakes dialogue; validate recognition locally, especially for Tanzanian accents.
4. Integrated listening-speaking tasks	Interactive listening and task-based speaking can better represent communicative purposes (Runge et al., 2024).	Indirect/extrapolated	Use short real-life tasks that require listening, response and meaning negotiation.
5. Human-in-the-loop assessment	Automatic speech indicators only partially represent oral proficiency constructs (Shi et al., 2024).	Indirect/extrapolated	Use AI for formative signals and triage; teachers retain analytic and consequential judgement.
6. Readiness before scale	Teacher competence, usefulness, infrastructure, ethics and governance shape adoption (Davis, 1989; Miao & Cukurova, 2024; MoEST, 2025).	Contextual/theoretical	Pilot only when access, devices, teacher competence, assessment validity, privacy, accessibility, support and cost are workable.

Note. *Direct = evidence from AI-supported Chinese/Mandarin learning; indirect = evidence from other second-language contexts; contextual = Tanzania-specific policy or implementation evidence; theoretical = established conceptual guidance; extrapolated = a cross-language or cross-context application requiring local validation.*

4.2. Requirement 2: An Offline-First Core with Optional Connectivity

The course should remain educationally functional when internet access is slow or unavailable. Each unit should have a transferable core package with concise explanations, compressed audio, vocabulary and character sets, practice prompts, model dialogues, an assessment rubric and retrieval activities. These materials may sit in a learning-management system but should also be downloadable to learner devices.

Connected AI can add speech recognition, conversational rehearsal, adaptive practice or teacher-facing analytics. Offline and AI-enhanced routes should target the same learning outcome, but the paper does not claim identical feedback conditions. Offline learners may rely on model recordings, peer checklists, self-record-and-compare routines and delayed teacher audio comments. This reduces but does not eliminate the immediacy gap created by lack of connectivity.

4.3. Requirement 3: Targeted AI for Pronunciation, Vocabulary and Character Practice

AI is most defensible where it increases meaningful attempts and shortens the delay to formative feedback. For pronunciation, learners can listen to a model, record a word or phrase, receive an ASR or pronunciation signal, compare the result and try again. Repeated failure on the same tone pair, syllable or phrase should lead to teacher-supported diagnosis instead of endless automated repetition. Mandarin tone practice should move quickly from isolated contrasts to familiar disyllabic words and short utterances (Li & Lee, 2021; Liu & Xiao, 2021).

Vocabulary and character practice can remain lighter technologically. Spaced retrieval can function offline; AI may optionally generate short examples or retrieval variations. Teacher checking is especially important at beginner levels. The newer Chinese-specific evidence is promising (Bao et al., 2025; Liu, 2026; Wu et al., 2025), but local recognition performance must be validated before feedback is treated as reliable for Tanzanian learners.

4.4. Requirement 4: Task-Based Listening and Speaking

Listening and speaking should converge in tasks that resemble communication. Learners might hear a short voice message and respond with requested information, complete a two-turn appointment dialogue, ask follow-up questions after a description, or role-play a campus interaction. AI can supply additional turns or variations, but the task remains defined by a communicative goal.

Low-bandwidth versions can use pre-recorded branching audio, peer role-play or teacher-led interaction. A generative conversational agent is an optional extension, not the only route to task completion.

4.5. Requirement 5: Automate First-Pass Feedback, not Final Judgement

Automated tools can detect some speech features, produce transcripts, identify missing responses or flag low recognition confidence. These functions can reduce undifferentiated teacher listening. Oral proficiency, however, also includes task fulfilment, intelligibility in context, interaction, repair, appropriateness and the functional effect of errors.

The assessment architecture therefore separates formative automation from consequential judgement. Learners may receive immediate AI-supported feedback during rehearsal. For graded oral tasks, teachers apply an analytic rubric. Automated indicators may be supplementary evidence or triage signals, but the teacher assigns the final grade. This is consistent with UNESCO's emphasis on human agency and accountable AI use (Miao & Cukurova, 2024; Miao & Holmes, 2023).

4.6. Requirement 6: Readiness before Scale

Scale should depend on readiness across eight dimensions: core access, devices and audio, teacher competence, assessment validity, privacy and data governance, accessibility, technical support, and cost and sustainability. Small pilots can begin with modest infrastructure, but a required AI function should not become compulsory while a critical gap remains in the conditions needed to use it fairly or safely.

Readiness is not binary. An institution may have strong teacher expertise but weak connectivity, or adequate devices but no clear data policy. The scorecard in Table 5 is therefore a planning aid rather than an institutional ranking.

5. The Pedagogy-First Low-Bandwidth AI Pathway

Figure 1 combines the six stages. The pathway starts with outcomes and ends with recycling, while implementation conditions constrain every stage. The iterative-feedback line indicates that performance evidence from Stage 6 informs the next Stage 1 learning cycle.

5.1. Stage 1: Outcome-Aligned Micro-Lessons

Each cycle begins with one observable outcome and a limited set of language resources. A beginner outcome might require a learner to introduce their name, place of origin, field of study and one preference and ask a partner two related questions. The teacher then identifies the pronunciation, vocabulary, characters and sentence patterns needed for that performance. This scope control matters because AI systems can generate more material than learners can use productively.

5.2. Stage 2: Offline-First Practice

Learners receive compressed audio models, a small phrase bank, characters with pronunciation support and retrieval prompts. Activities include listen-and-select, read-and-recall, short dictation, tone discrimination and model imitation. Materials should remain reusable without logging into an online service, and retrieval should be spaced across several short sessions.

5.3. Stage 3: Targeted AI Feedback with an Offline Substitute

Where connectivity permits, learners use speech recognition or a conversational function for extra rehearsal. Feedback is deliberately narrow: whether a phrase was recognised, where a likely tone or segment problem occurred, or whether a required semantic element was present. Learners treat this output as a formative signal, not authoritative judgement, and persistent errors are recorded for teacher review.

When connected AI is unavailable, learners follow a listen-record-compare routine using downloaded models, use a peer checklist for target words and tones, and receive periodic teacher audio comments on selected samples. This route targets the same outcome but provides slower and less individualised feedback. The framework therefore claims outcome equivalence as a design goal, not complete equality of feedback experience.

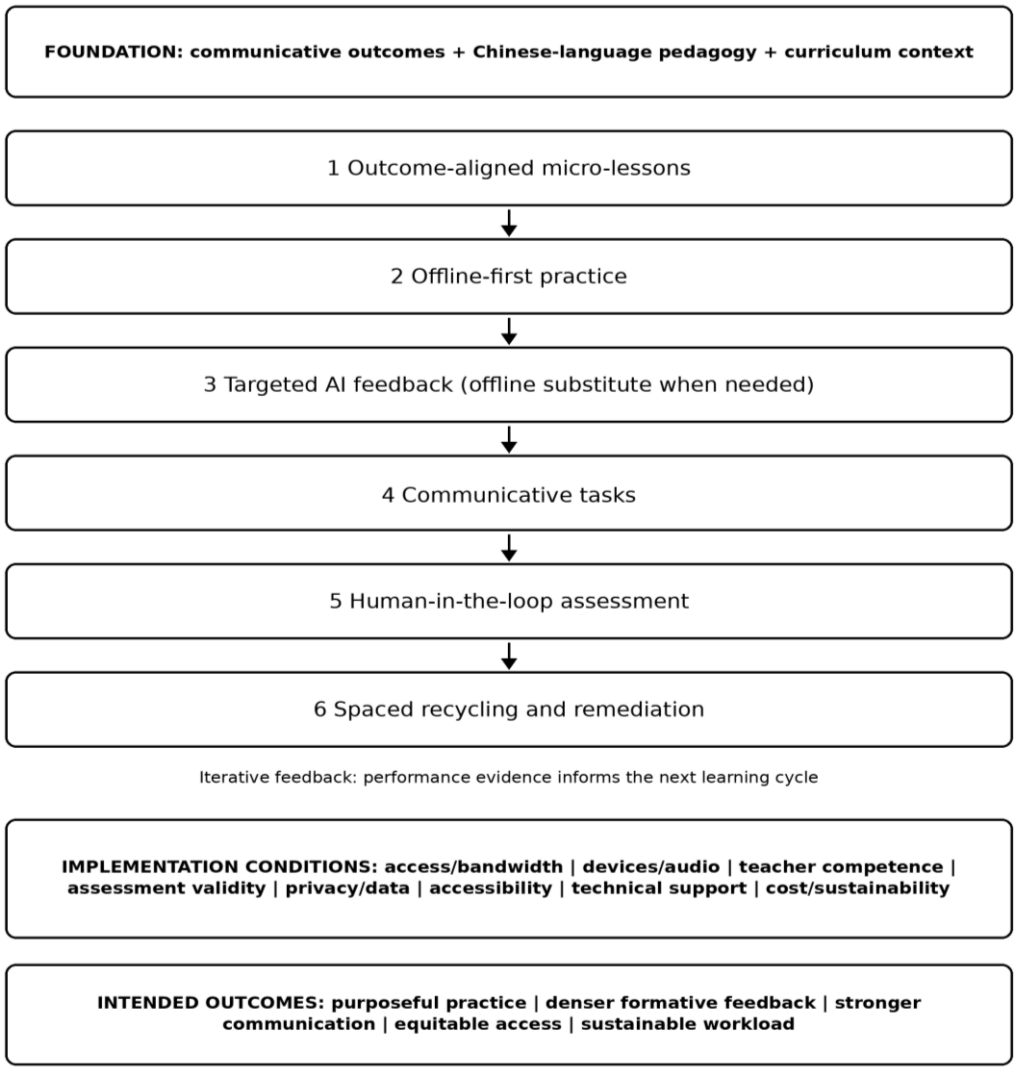


Figure 1. *Pedagogy-First, Low-Bandwidth AI-Supported Pathway for Chinese Language Learning*

5.4. Stage 4: Communicative Task

The learner then performs a task with a peer, teacher or approved conversational system. The task contains an information gap or communicative purpose and requires both listening and speaking. This stage checks whether controlled practice transfers into interaction.

5.5. Stage 5: Human-in-the-Loop Assessment

Automated feedback can organise evidence, but the teacher applies the analytic rubric in Table 2. The rubric separates task fulfilment, intelligibility, listening response, fluency/repair and

language resources so that a single pronunciation feature does not determine the entire score. Local programmes should adapt descriptors to course outcomes and familiarise raters before high-stakes use.

Table 2. *Illustrative Analytic Rubric for a Short Chinese Listening-Speaking Task*

Dimension	4 - Strong	3 - Adequate	2 - Developing	1 - Limited	AI role
Task fulfilment & interaction	Completes purpose; responds and follows up appropriately.	Completes main purpose with minor prompting.	Partial completion; interaction breaks at times.	Purpose largely unmet; minimal response.	May check required elements; never finalises interaction score.
Intelligibility, tones & segments	Generally clear; tone/segment errors rarely impede meaning.	Mostly clear; some errors but meaning remains accessible.	Frequent errors require listener effort or repetition.	Errors often prevent understanding .	Provides rehearsal feedback and flags low-confidence segments.
Listening response	Understands key details & responds accurately and promptly.	Understands main message with minor loss of detail.	Understands part of message; needs repetition / support.	Little evidence of understanding .	Can administer prompts or branch tasks; human checks construct-relevant performance.
Fluency & repair	Sustains short turns; pauses are manageable; repairs effectively.	Some hesitation but message continues.	Frequent hesitation and limited repair.	Fragmented output; unable to sustain turn.	May display timing /transcript indicators as supplementary evidence.
Language resources	Uses expected vocabulary/structures with useful range for level.	Mostly appropriate resources; errors do not obscure message.	Restricted range and recurring errors.	Insufficient resources for task.	May identify missing target items; teacher interprets quality & appropriateness.

5.6. Stage 6: Spaced Recycling and Remediation

Assessment feeds forward. Learners receive one or two high-priority targets instead of every detected error. A learner with persistent Tone 2/Tone 3 confusion may receive focused perception and production work; another may need more listening-response practice or a sentence frame for repair. Successful language resources reappear in later tasks at increasing intervals.

5.7. Illustrative Lesson Shell

Table 3 shows one 60-minute guided lesson plus asynchronous practice. It is an example, not a prescribed contact-hour model.

5.8. Illustrative Workload Simulation

The workload simulation is a planning aid, not a local class estimate. To make the class-size assumption transparent, Table 4 shows 20-, 40- and 60-learner scenarios, each with two short oral samples per learner per week. Full manual review assumes three minutes per sample. Hybrid A assumes that 75% of samples receive a 45-second verification and 25% a full three-minute review; Hybrid B assumes 80% receive a 30-second verification and 20% a full review. The resulting times are sensitivity estimates, not observed Tanzanian workloads.

Table 3. *Illustrative 60-Minute Plus Asynchronous Lesson Shell: Self-Introduction and Follow-up Questions*

Phase	Learning activity	Low-bandwidth core	Optional AI layer	Teacher responsibility
Before class	Retrieve eight target words, four characters and two sentence frames.	Downloaded audio, PDF/image card, offline flashcards.	Adaptive retrieval variations.	Select high-value language and check generated examples.
10 min	Tone and phrase discrimination.	Pre-recorded examples and peer checking.	ASR-supported listen-record-compare cycle.	Model contrasts; explain recurring error patterns.
15 min	Guided micro-dialogue.	Printed/downloaded prompt cards in pairs.	Conversational agent provides extra turns.	Monitor meaning and appropriateness.
20 min	Information-gap self-introduction task.	Peer task with different information cards.	Optional simulated partner for an extra attempt.	Observe transfer to genuine interaction.
15 min	Short performance sample and feedback.	Phone audio recorded locally or live teacher check.	Transcript/recognition-confidence flag.	Apply analytic rubric; retain final judgement.
After class	Spaced recycling and one remediation target.	Two 5-minute offline practice cycles.	Automated reminder or targeted extra practice.	Review persistent difficulties and plan next cycle.

The scenarios show where time savings might arise by concentrating teacher attention on performances that require professional judgement. They assume locally validated triage rules and exclude initial training, configuration, calibration, troubleshooting and false-flag investigation. A pilot should measure baseline marking time, verification time, false-flag rates and setup costs before treating any simulated saving as an operational expectation.

Table 4. *Illustrative Weekly Workload Sensitivity Across 20-, 40- and 60-Learner Classes*

Class size	Scenario	Assumption	Approx. teacher time
20 learners	Fully manual	40 samples × 3 min	120 min
20 learners	Hybrid A	30 × 0.75 min + 10 × 3 min	52.5 min
20 learners	Hybrid B	32 × 0.5 min + 8 × 3 min	40 min
40 learners	Fully manual	80 samples × 3 min	240 min
40 learners	Hybrid A	60 × 0.75 min + 20 × 3 min	105 min
40 learners	Hybrid B	64 × 0.5 min + 16 × 3 min	80 min
60 learners	Fully manual	120 samples × 3 min	360 min
60 learners	Hybrid A	90 × 0.75 min + 30 × 3 min	157.5 min
60 learners	Hybrid B	96 × 0.5 min + 24 × 3 min	120 min

5.9. Institutional Readiness Scorecard

Table 5 lists eight readiness dimensions, each scored from 0 to 2. A score of 0 marks a critical gap, 1 partial readiness requiring mitigation, and 2 a workable minimum. For core access, 0 means learning depends on continuous internet, whereas 2 means every essential outcome has an offline or low-data route. The total score is a discussion aid, not a validated index. Any zero in core access, teacher competence, assessment validity or privacy should trigger corrective action before the relevant AI function becomes compulsory.

Table 5. *Illustrative AI-Supported Chinese Language Implementation Readiness Scorecard*

Dimension	0 - Critical gap	1 - Partial readiness	2 - Workable minimum
Core access	Learning depends on continuous internet.	Some offline alternatives exist but are incomplete.	All essential outcomes have an offline/low-data route.
Devices and audio	Many learners cannot reliably access recording/playback.	Shared or inconsistent access.	Most learners have dependable access; loan/shared plan covers gaps.
Teacher competence	No preparation in AI limits or language-specific use.	Basic tool familiarity only.	Teacher can align tool, pedagogy & Chinese content & audit outputs.
Assessment validity	Automated score treated as final without local checks.	Human moderation exists but criteria are unclear.	Analytic rubric, human judgement and local checking are explicit.
Privacy and data	No institutional guidance for learner data.	General policy exists but tool-level practice is unclear.	Data minimisation, consent/notice, account and retention rules are specified.
Accessibility	No alternative route for learners with access needs.	Case-by-case accommodations only.	Transcripts, alternative inputs and accessible materials are planned.
Technical support	Teachers troubleshoot alone.	Informal help is available.	Named support route and basic troubleshooting procedures exist.
Cost and sustainability	Recurring costs unknown or depend on personal payment.	Pilot costs known but scale financing uncertain.	Total cost, licensing & replacement /renewal plan are budgeted.

6. Discussion

The framework makes pedagogy and context the entry point for technology decisions. It begins with communicative outcomes, identifies where practice or feedback is scarce, and selects the least demanding technology that adds educational value. This is consistent with TPACK and with evidence that AI outcomes vary with instructional design, not technology presence alone (Mishra & Koehler, 2006; Wu, 2024).

This position differs from technology-integration models that imply a progression toward increasingly transformative uses of technology. The Substitution-Augmentation-Modification-Redefinition (SAMR) model, for example, is widely used to describe levels of technology integration, but Hamilton et al. (2016) criticise its limited attention to context, its hierarchical structure and its emphasis on products over process. The present framework does not assume that a more technologically complex activity is better. A downloaded audio prompt can be preferable to a conversational AI system when both meet the same outcome and the former is more reliable or equitable.

The pathway also differs from a general instructional-design cycle such as Analyze-Design-Develop-Implement-Evaluate (ADDIE) (Branch, 2009). ADDIE provides a broad process for instructional development. The present model operates at a narrower decision level inside that process: it specifies how low-bandwidth alternatives, AI feedback, communicative tasks, assessment authority and readiness checks should be organised for Chinese language teaching. In that sense it can sit within an ADDIE-style course-development process instead of replacing it.

Two practical implications follow from the comparison with broader technology-integration models. First, offline-first design is a robustness strategy, but it does not equalise feedback conditions: learners using peer checks and delayed teacher audio may receive slower feedback than learners using connected ASR. A pilot should therefore measure whether this difference affects learning or participation. Second, automated speech indicators are most defensible as formative aids. Chinese-specific studies are promising, yet no identified study validates Mandarin ASR for Tanzanian-accented speech; disagreement between a learner and the system should therefore trigger human review.

These points keep teacher judgement central without making the framework anti-technology. Teachers define the learning outcome, check generated language, decide when feedback needs intervention, apply the analytic rubric for consequential assessment and review whether a tool remains educationally worthwhile. This role is consistent with UNESCO's human-centred AI competency guidance (Miao & Cukurova, 2024).

7. Conclusion

7.1. Practical Implications for Tanzanian Higher Education

For Tanzanian universities, implementation should begin with a reliable pedagogical minimum: trained teachers, locally relevant outcomes, reusable audio/text resources, clear assessment criteria, basic recording access and an offline distribution route. AI can then be piloted for specific bottlenecks such as pronunciation rehearsal or additional conversational practice, with access and learning effects monitored before scale.

Institutional policy should specify learner accounts, personal data, audio storage and deletion, provider retention, cross-border processing, generated content and automated scores. Teacher development should be task-centred: staff can compare offline and AI-supported versions of the same activity, audit generated Chinese, identify privacy implications and decide what evidence would justify continued use. Re-use of classroom recordings or logs for research requires a separate ethical basis.

7.2. Limitations and Future Research

This is a conceptual design synthesis, not an effectiveness study. The pathway has not been field-tested, and the workload scenarios and readiness thresholds are not validated measures. The evidence base is uneven: a substantial share of ASR, conversational-AI and automated-scoring research concerns English, while Chinese-specific studies remain fewer and evidence

from African higher education is especially limited. No identified study establishes ASR accuracy for Mandarin spoken with Tanzanian accents.

The search was purposive and not systematic. The original platform searches did not log a complete denominator of hits, relevance checking was not independently double-screened, and Scopus, Web of Science and ERIC were not systematically queried. Relevant evidence may therefore have been missed, and reliable found/screened counts cannot be reconstructed retrospectively. This limitation is reported explicitly because approximate counts would give a false impression of systematic coverage. Sabatho et al. (2026), authored by members of the present team, is a self-citation used only for Tanzanian curriculum-implementation context; it is not evidence of AI effectiveness, and no dataset, findings table or text from that study is reused as primary evidence here.

Future research should begin with a feasibility pilot that records connectivity, device access, learner language backgrounds, task completion, teacher workload, recognition-error patterns and privacy procedures. A later mixed-methods or quasi-experimental study could compare the pathway with existing practice using pronunciation, listening and speaking outcomes, classroom observation and learner/teacher experience. The readiness scorecard also requires expert review and empirical validation.

7.3. Answers to the Research Questions and Final Statement

RQ1 is answered by six principles: pedagogy before platform, an offline-first core, targeted AI use, integrated listening-speaking tasks, human-in-the-loop assessment and readiness before scale. RQ2 is answered by a dual-route architecture in which connected AI can add rapid formative feedback while downloaded models, peer checking, self-recording and teacher comments preserve a low-bandwidth route without claiming identical feedback conditions. RQ3 is answered by eight readiness dimensions covering core access, devices/audio, teacher competence, assessment validity, privacy/data governance, accessibility, technical support and cost/sustainability.

The framework offers a practical route between ignoring AI because infrastructure is imperfect and adopting AI before pedagogical and governance conditions are ready. Its contribution is a testable design proposition. Local piloting must determine whether the pathway improves learning, access and workload in Tanzanian universities before wider scale is justified.

Ethics, Data Availability and AI-Assisted Writing Disclosure

Ethics Statement. This study used published literature and publicly available policy and curriculum documents. It did not involve human participants, personal records or primary data collection; therefore, human-subject ethical approval and informed consent were not applicable.

Data Availability. No new dataset was created. The evidence sources used in the synthesis are cited in the article, and the design matrix is provided as Supplementary Table S1 (See the Appendix).

Declaration of Generative AI and AI-Assisted Technologies. During manuscript preparation, the authors used ChatGPT (OpenAI) as an AI-assisted writing tool for language refinement. The authors revised the manuscript, and take full responsibility for the final text and conclusions. No primary data were generated, altered or analysed by generative AI.

Disclosure Statement

The authors declare that there is no conflict of interest regarding the publication of this article. No financial, personal, or professional relationships have influenced the research, analysis, or conclusions presented in this work.

Notes on Contributors

Jivitius Ching'oro Sabatho is a PhD candidate in Educational Economy and Management at HUST and Chinese Language Curriculum Developer at the Tanzania Institute of Education. His research focuses on curriculum management, Chinese language education, teacher education, educational policy and AI-supported education.

sabathojc@gmail.com

Eliya Mbulika Ndiwu is a PhD candidate in Educational Management at HUST and Principal Vocational Tutor at VETA, Tanzania. His research interests include TVET, educational management, technology-supported learning and workforce skills development.

ndiwuelliya@gmail.com

Moabu Jimmy Chandafa is a PhD candidate in Higher Education at HUST and Assistant Lecturer in Technical Education at MUST, Tanzania. His interests include educational technology, teacher professional development, curriculum implementation and educational leadership.

moabuchandafa95@gmail.com

Fang Huang is an Associate Professor in the School of Education at HUST. Her research interests include curriculum and instruction, science education and comparative research in primary and secondary education.

fanghuang@hust.edu.cn

ORCID

Jivitius Ching'oro Sabatho *https://orcid.org/0009-0007-7026-9979*

Eliya Mbulika Ndiwu *https://orcid.org/0009-0000-4138-1581*

Moabu Jimmy Chandafa *https://orcid.org/0009-0006-3513-6938*

Fang Huang *https://orcid.org/0000-0002-2826-7060*

References

Bao, W., Yusop, F. D., & Chang, X. (2025). Effects and contributing factors of an automatic speech recognition dialogue platform on international students' Chinese speaking ability. *Discover Computing, 28*, Article 265. https://doi.org/10.1007/s10791-025-09787-9

Branch, R. M. (2009). *Instructional design: The ADDIE approach*. Springer. https://doi.org/10.1007/978-0-387-09506-6

Contini-Morava, E. (1997). Swahili phonology. In A. S. Kaye (Ed.), *Phonologies of Asia and Africa* (Vol. 2, pp. 841–860). Eisenbrauns.

Council of Europe. (2020). *Common European framework of reference for languages: Learning, teaching, assessment—Companion volume*. Council of Europe Publishing. https://www.coe.int/en/web/common-european-framework-reference-languages/cefr-companion-volume-and-its-language-versions

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340. https://doi.org/10.2307/249008

Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. *Computers and Education: Artificial Intelligence, 6*, 100230. https://doi.org/10.1016/j.caeai.2024.100230

Hamilton, E. R., Rosenberg, J. M., & Akcaoglu, M. (2016). The substitution augmentation modification redefinition (SAMR) model: A critical review and suggestions for its use. *TechTrends, 60*(5), 433–441. https://doi.org/10.1007/s11528-016-0091-y

Kafyulilo, A., Fisser, P., & Voogt, J. (2016). Factors affecting teachers' continuation of technology use in teaching. *Education and Information Technologies, 21*, 1535–1554. https://doi.org/10.1007/s10639-015-9398-0

Kim, S. K., & Webb, S. (2022). The effects of spaced practice on second language learning: A meta-analysis. *Language Learning, 72*(1), 269–319. https://doi.org/10.1111/lang.12479

Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for language teaching and learning. *RELC Journal, 54*(2), 537–550. https://doi.org/10.1177/00336882231162868

Lai, W. Y. W., & Lee, J. S. (2024). A systematic review of conversational AI tools in ELT: Publication trends, tools, research methods, learning outcomes, and antecedents. *Computers and Education: Artificial Intelligence, 7*, 100291. https://doi.org/10.1016/j.caeai.2024.100291

Landa, E., Zhu, C., & Sesabo, J. (2021). Readiness for integration of innovative teaching and learning technologies: An analysis of meso-micro variables in Tanzanian higher education. *International Journal of Educational Research Open, 2*, 100098. https://doi.org/10.1016/j.ijedro.2021.100098

Li, Y., & Lee, G. (2021). The effect of perceptual training on teaching Mandarin Chinese tones. In C. Yang (Ed.), *The acquisition of Chinese as a second language pronunciation: Segments and prosody* (pp. 107–139). Springer. https://doi.org/10.1007/978-981-15-3809-4_5

Liu, J., & Xiao, C. (2021). Tone category learning should serve tone word learning: An experiment of integrating pronunciation teaching in the L2 Chinese curriculum. In C.

Yang (Ed.), *The acquisition of Chinese as a second language pronunciation: Segments and prosody* (pp. 141–162). Springer. https://doi.org/10.1007/978-981-15-3809-4_6

Liu, S. H. J. (2026). Effects of voice-based ChatGPT on Chinese-as-a-second-language learners' Mandarin pronunciation. *International Journal of Applied Linguistics*. Advance online publication. https://doi.org/10.1111/ijal.70213

Miao, F., & Cukurova, M. (2024). *AI competency framework for teachers*. UNESCO. https://doi.org/10.54675/ZJTE2084

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535

Ministry of Education, Science and Technology. (2025). *National digital education guidelines for universities*. Government of the United Republic of Tanzania. https://www.moe.go.tz/

Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record, 108*(6), 1017–1054. https://doi.org/10.1111/j.1467-9620.2006.00684.x

Mtebe, J. S., & Raisamo, R. (2014). Challenges and instructors' intention to adopt and use open educational resources in higher education in Tanzania. *The International Review of Research in Open and Distributed Learning, 15*(1), 249–271. https://doi.org/10.19173/irrodl.v15i1.1687

Ngo, T. T.-N., Chen, H. H.-J., & Lai, K. K.-W. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL, 36*(1), 4–21. https://doi.org/10.1017/S0958344023000113

Richardson, I. (1956). *The role of tone in the structure of Sukuma* (Doctoral dissertation, University of London). SOAS Research Online. https://eprints.soas.ac.uk/29083/

Runge, A., Attali, Y., LaFlair, G. T., Park, Y., & Church, J. (2024). A generative AI-driven interactive listening assessment task. *Frontiers in Artificial Intelligence, 7*, 1474019. https://doi.org/10.3389/frai.2024.1474019

Sabatho, J. C., Huang, F., Mshana, G. G., & Rwiza, F. B. (2026). Sustaining Chinese language curriculum reform in Tanzania: A documentary analysis of policy alignment, teacher support, localisation, and institutional coordination. *International Journal of Research*

and Scientific Innovation, 13(6), 4447–4462. https://doi.org/10.51244/IJRSI.2026.1306000332

Shi, X., Wang, X., & Zhang, W. (2024). Exploring the relationships between ASS indices and CAF and the impact on Chinese college students' oral English performance. *Language Testing in Asia, 14,* Article 30. https://doi.org/10.1186/s40468-024-00306-z

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research, 104,* 333–339. https://doi.org/10.1016/j.jbusres.2019.07.039

Tanzania Institute of Education. (2023). *Chinese language syllabus for ordinary secondary education Form I–IV*. Ministry of Education, Science and Technology, United Republic of Tanzania. https://www.tie.go.tz/uploads/documents/sw-1727193068-CHINESE%20LANGUAGE%20SYLLABUS%20FOR%20ORDINARY%20SECONDARY%20EDUCATION%20FORM%20I-IV.pdf

Torraco, R. J. (2016). Writing integrative literature reviews: Using the past and present to explore the future. *Human Resource Development Review, 15*(4), 404–428. https://doi.org/10.1177/1534484316671606

UNESCO. (2023). *Global education monitoring report 2023: Technology in education—A tool on whose terms?* UNESCO. https://doi.org/10.54676/UZQV8501

Wu, X.-Y. (2024). Artificial intelligence in L2 learning: A meta-analysis of contextual, instructional, and social-emotional moderators. *System, 126,* 103498. https://doi.org/10.1016/j.system.2024.103498

Wu, Y., Su, H., & Yin, H. (2025). Enhancing Mandarin tone acquisition: Computer-assisted pronunciation training in self-directed L2 learning. *International Journal of Chinese Linguistics, 12*(2), 286–314. https://doi.org/10.1075/ijchl.00040.wu

Appendix

Supplementary Material

This matrix summarizes 23 design-bearing sources; eight additional article references provide background, linguistic, theoretical, policy or discussion context. Source clusters group publications informing the same design decision. The matrix is an audit trail for the conceptual synthesis, not a claim of systematic-review exhaustiveness.

Supplementary Table S1. Design matrix used in the evidence synthesis

Evidence source/cluster	Pedagogical affordance	Limitation or risk	Resource implication	Human responsibility
Wu (2024); Du & Daniel (2024); Lai & Lee (2024)	AI can expand interaction, feedback and practice when task design is appropriate.	Effects vary by context; accuracy, dependence and privacy concerns remain.	Connected tools may increase data / device demands.	Teacher defines the learning purpose and audits output.
Ngo et al. (2024)	ASR can improve pronunciation, especially with explicit sustained feedback.	Evidence base is dominated by English; recognition is not mastery.	Needs microphone / recording and often connectivity.	Teacher interprets persistent errors and avoids treating recognition as final judgement.
Bao et al. (2025)	ASR dialogue practice can improve Chinese speaking performance.	One institutional / platform context; affective factors strongly matter.	Requires speech-enabled platform and repeated access.	Teacher / peer interaction remains important; validate transfer to local learners.
Wu et al. (2025)	CAPT/ASR feedback can support Mandarin lexical-tone acquisition.	Not conducted with African or Tanzanian learners.	Speech-enabled device and software access required.	Validate tone feedback locally and route ambiguity to a teacher.
Liu (2026)	Voice-based ChatGPT can support Mandarin pronunciation practice.	Participants were from Asian contexts; no Tanzanian-accent evidence.	Voice AI access may require stable connectivity/account access.	Teacher checks whether feedback is linguistically and pedagogically appropriate.
Li & Lee (2021); Liu & Xiao (2021)	Tone practice benefits from contextualised, word-linked sequencing.	Not AI-specific and not Tanzania-specific.	Can be implemented offline with recordings and teacher models.	Teacher selects contrasts and links pronunciation to lexical meaning.
Runge et al. (2024)	AI can support interactive listening task generation / administration.	Generated content may threaten construct relevance without review.	Generation may be online; delivery can be pre-generated/offline.	Human quality assurance protects task purpose and validity.
Shi et al. (2024)	Automatic speech indicators can support screening / feedback.	Indicators only partially represent broader oral proficiency.	Requires audio processing and scoring infrastructure.	Teacher retains consequential judgement.
Kim & Webb (2022)	Spaced practice supports retention in L2 learning.	Not specific to Chinese AI learning.	Low resource demand; suitable for offline scheduling.	Teacher designs spacing and links micro-practice to integrated tasks.

Evidence source/cluster	Pedagogical affordance	Limitation or risk	Resource implication	Human responsibility
Mtebe & Raisamo (2014); Kafyulilo et al. (2016); Landa et al. (2021)	Tanzanian evidence identifies adoption / readiness constraints and enabling support.	Studies predate current GenAI tools and do not test Chinese learning.	Connectivity, electricity, class size, tools & organisational readiness matter.	Institutions plan support and provide low-data alternatives.
UNESCO (2023); MoEST (2025)	Technology should be relevant, equitable, governed and sustainable.	Policy/guidance does not establish learning effects.	Requires institutional planning, access and governance capacity.	Institutions define acceptable use and equity safeguards.
Miao & Holmes (2023); Miao & Cukurova (2024)	Human-centred AI and teacher competence provide ethical/role boundaries.	Guidance is normative rather than efficacy evidence.	Requires professional learning and data-governance procedures.	Human agency, privacy and accountability remain explicit.
Mishra & Koehler (2006); Davis (1989); Branch (2009); Hamilton et al. (2016)	Frameworks help test fit, adoption and instructional-design logic.	General frameworks do not resolve Mandarin-specific pronunciation validity.	No direct resource requirement; used as analytic lenses.	Authors use them to justify outcome-first, context-sensitive selection rather than technology hierarchy.