



(RESEARCH ARTICLE)



MITIGATING THE GERMAN LANGUAGE TEACHER AND SHORTAGE IN UNITED STATES K 16 EDUCATION THROUGH AI ASSISTED INSTRUCTIONAL DESIGN

Gideon Akinwale Oladayo *

Department of Comparative Literature, College of Liberal Arts, Purdue University.

* Corresponding Author

ORCID Details

Gideon Akinwale Oladayo: <https://orcid.org/0009-0005-4537-6962>

World Journal of Advanced Research and Reviews, 2026, 31(03), 1778–1785

Publication history: Received on 15 August 2026; revised on 24 September 2026; accepted on 26 September 2026

Article DOI: <https://doi.org/10.30574/wjarr.2026.31.3.2461>

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

German language education in the United States faces two connected pressures: a limited supply of qualified teachers and a contraction of programs and enrollments, especially at institutions with fewer resources. Generative artificial intelligence cannot certify teachers, improve salaries, or reverse institutional disinvestment. It can, however, help qualified educators design, differentiate, evaluate, and share instructional resources more efficiently. This article argues that AI should therefore be treated as infrastructure for increasing teacher capacity rather than as a substitute for professional labor. It proposes the German AI Assisted Instructional Design and Evaluation framework, or G-AIDE, a six-stage process that begins with a documented instructional need and proceeds through standards-based outcome design, constrained AI generation, expert review, classroom piloting, and version-controlled dissemination. The article analyzes the framework according to linguistic accuracy, pedagogical alignment, cultural validity, differentiation, workload, scalability, and ethics. It contends that a resource is not scalable merely because an AI system can generate it quickly; scalability requires evidence of quality, portability, accessibility, and responsible governance. Properly implemented, G-AIDE can reduce duplicated preparation, support novice and geographically isolated teachers, and preserve access to German while retaining teachers as the accountable designers of learning.

Keywords: German language education, Teacher shortage, Generative artificial intelligence, Instructional design, ACTFL, K-16 education, Teacher workload

1 INTRODUCTION

The contraction of German language education in the United States is visible in enrollment data, program availability, and the professional conditions under which teachers work. In higher education, the Modern Language Association reports that total enrollments in languages other than English fell 16.6 percent between 2016 and 2021, while German experienced the largest decline among the fifteen most commonly taught languages, falling 33.6 percent. During the same period, 172 fewer institutions reported German enrollments (Lusin et al. 3). These figures do not prove that every disappearing enrollment represents a formally closed department, but they do identify a substantial loss of instructional

presence. They also reveal a feedback loop: fewer programs produce fewer advanced students and prospective teachers; a smaller teacher pipeline makes programs harder to staff; unstable programs then become less able to recruit students.

K-12 data are less consistent, which is itself part of the problem. The National K-12 Foreign Language Enrollment Survey estimated that approximately 19.66 percent of school-age students were enrolled in a language course and that roughly 330,898 studied German in 2014-15 (American Councils 7-9). The report also warned that the absence of regularly collected, comparable K-12 data impedes national planning (5-6). More recent federal reporting indicates that foreign-language vacancies are particularly difficult to fill in rural schools. A shortage is therefore not one condition. It may mean that no qualified candidate applies, that an applicant lacks the required credential, that a school cannot retain its teacher, that a small program depends on one person, or that an institution declines to fund a position. These distinctions matter because an instructional technology can alleviate only some consequences of shortage.

Generative artificial intelligence has entered this vulnerable environment with an alluring promise of speed. A teacher can request a novice-level reading, a role-play, a rubric, or three differentiated versions of a worksheet and receive them in seconds. Yet rapid production does not establish linguistic accuracy, pedagogical value, cultural validity, or classroom usefulness. Large language models generate probable sequences of language from patterns in their training data; they do not possess a teacher's situated knowledge of students, curriculum, community, disability accommodation, or institutional responsibility. Their fluent output may conceal factual invention, awkward German, stereotyped cultural representation, or activities that bear only a superficial relationship to stated outcomes. UNESCO consequently recommends a human-centered approach in which educational uses of generative AI are ethically validated, age appropriate, privacy protective, and subject to human agency (Miao and Holmes 13, 26-7).

The central argument of this article is that human-supervised, standards-aligned AI instructional design can mitigate some effects of the German teacher shortage by reducing duplicated preparation, supporting differentiation, and expanding access to reviewed resources. It cannot solve the structural causes of shortage, and it should never be used to redefine a teacher vacancy as a software opportunity. The relevant question is not whether AI can generate German teaching materials. It plainly can. The question is under what design and governance conditions AI-assisted materials become trustworthy enough to extend the capacity of qualified educators. To answer that question, this article proposes G-AIDE, the German AI Assisted Instructional Design and Evaluation framework. The framework joins backward design, technological pedagogical content knowledge, ACTFL-oriented proficiency teaching, human review, classroom evidence, and transparent dissemination. Its governing principle is augmentation with accountability: AI may accelerate drafting, but teachers retain authority over goals, selection, revision, implementation, and evaluation.

2 THE SHORTAGE AS A CAPACITY AND ACCESS PROBLEM

Describing the issue simply as a lack of German teachers risks obscuring its institutional dimensions. An absolute labor shortage occurs when schools advertise viable positions but cannot locate appropriately prepared applicants. A qualification shortage occurs when available applicants lack certification or sufficient German proficiency. A retention shortage emerges when workload, pay, isolation, or limited advancement drives teachers from the profession. Finally, an allocation problem occurs when institutions could sustain instruction but choose not to authorize full-time lines. AI addresses none of these causes directly. It does not create certification routes, negotiate salaries, mentor student teachers, or secure administrative investment. Any argument that presents automation as the solution would mistake the effects of scarcity for its causes.

Nevertheless, scarcity creates a genuine instructional-design problem. A single high-school German teacher may prepare German I through Advanced Placement, advise a club, recruit students, maintain exchanges, communicate with families, and defend enrollment. At a small college, one instructor may coordinate multiple levels, supervise adjuncts or graduate assistants, assess placement, organize cultural programming, and continually rebuild courses as textbooks and student needs change. Preparation is not incidental labor. It involves selecting authentic texts, sequencing input, anticipating difficulty, designing interaction, creating assessments, giving feedback, and accommodating differences. When this work is repeated independently by hundreds of teachers, limited professional capacity is consumed by parallel production.

Program contraction makes this burden more consequential. The MLA census found that German lost both enrollments and reporting programs, while two-year institutions experienced sharper overall language declines than four-year institutions. The report emphasizes that reductions at two-year colleges disproportionately restrict language access for first-generation, low-income, and other underserved students (Lusin et al. 10). Thus, the shortage is also an equity problem. Students at a well-funded suburban school or research university may choose among languages and course levels, while students at rural schools or community colleges may have no German option at all. A scalable resource system cannot supply the relational and institutional work of a teacher, but it can reduce the cost of sustaining varied course preparations and make quality materials available beyond well-resourced departments.

German is particularly suitable for a disciplined investigation of AI-assisted design. Its instruction requires careful control of grammatical complexity, including case, gender, adjective endings, verb placement, separable prefixes, and subordinate clauses. Proficiency also depends on pragmatics: the choice between *du* and *Sie*, regional vocabulary, modal particles, politeness, and genre conventions cannot be reduced to grammatical correctness. Cultural instruction must reach beyond a nation-centered catalogue of festivals and foods to represent Austria, Switzerland, multilingual regions, migration, racial diversity, reunification, colonial history, and contemporary everyday life. A system may produce flawless-looking German while reproducing a narrow image of Germanness. The validation problem is therefore simultaneously linguistic, pedagogical, and political.

3 THEORETICAL AND PEDAGOGICAL FOUNDATIONS

G-AIDE begins from the principle that technology integration is a problem of teacher knowledge rather than tool availability. Mishra and Koehler's model of technological pedagogical content knowledge, commonly called TPACK, explains that effective integration emerges from the interaction of content, pedagogy, and technology rather than mastery of any one domain. They stress that technologies cannot be applied through a universal formula because teaching is situated and the relationships among knowledge domains are dynamic (1025-29). In German instruction, technological knowledge means more than knowing how to enter a prompt. It includes recognizing model limitations, protecting data, recording sources, comparing outputs, and deciding when AI is inappropriate. Content knowledge includes German grammar, pragmatics, cultural history, and variation. Pedagogical knowledge includes second-language acquisition, scaffolding, interaction, assessment, accessibility, and classroom management. A prompt is instructionally responsible only where these domains meet.

Backward design supplies the sequence. Wiggins and McTighe argue that curriculum should begin with desired understandings, determine acceptable evidence, and then plan learning experiences (17). Applied to AI, this order prevents a common error: generating an attractive activity and inventing an objective afterward. A teacher should first specify what learners will be able to interpret, negotiate, or present; define what successful performance looks like; and only then use AI to draft materials. If the intended outcome is that novice-high learners can ask and answer questions about weekend plans, for example, the output should support interpersonal exchange rather than merely provide a vocabulary matching exercise.

ACTFL's World-Readiness Standards provide an organizing account of language use through Communication, Cultures, Connections, Comparisons, and Communities. The Communication goal further distinguishes interpretive, interpersonal, and presentational modes (National Standards Collaborative Board 9-12). These categories can function as design constraints and validation criteria. An AI-generated unit should identify its communication mode, performance range, cultural context, and evidence of learning. Such alignment is not bureaucratic labeling. It exposes whether an activity actually creates purposeful language use. A dialogue completed by filling isolated endings may practice form, but it should not be represented as interpersonal communication unless learners exchange meaning, respond to one another, and manage unpredictability.

The final foundation is human-in-the-loop design. This phrase should not mean that a teacher briefly approves an AI product. It means that professional judgment governs the entire cycle. The teacher identifies the need, defines constraints, supplies or approves source material, audits output, pilots the resource, interprets evidence, and decides whether it can be shared. Bender and colleagues caution that large language models reproduce patterns and biases from vast datasets while generating text whose fluency can encourage users to attribute understanding that the system does not have (610-612). Human review is therefore not an optional final polish; it is the mechanism that converts probable text into accountable curriculum.

4 THE G-AIDE FRAMEWORK

G-AIDE is a six-stage cycle: identify the need; define standards-aligned outcomes; generate a constrained draft; validate through teacher and expert review; pilot in a classroom; and revise, document, and disseminate. The stages are cyclical because classroom evidence may reveal a misunderstood need or an inappropriate outcome. Scalability occurs only after the cycle produces a resource with a documented purpose, review history, and conditions of use.

4.1 Identifying the Instructional Need

The first stage asks what problem requires a resource and for whom. Useful information includes grade level, proficiency range, course sequence, class size, time available, technology access, learner backgrounds, accommodations, and institutional constraints. The need may be narrow: an interpretive task based on a train timetable for German I. It may be systemic: a rural teacher needs editable units for four preparations. A needs statement prevents generic generation and makes later evaluation possible.

At program level, needs analysis should combine teacher surveys, interviews, course inventories, job postings, state shortage designations, enrollment trends, and records of canceled sections. The data must distinguish demand for instruction from demand for labor. Low enrollment may reflect weak student interest, but it may also follow schedule conflicts, missing prerequisites, counselor advice, or repeated staff turnover. Similarly, teachers may report a lack of materials when the deeper issue is insufficient time to adapt commercial resources. Analysis should therefore ask which tasks consume time and which require expertise. AI is most defensible for high-volume, revisable drafting, not for decisions involving student welfare or high-stakes evaluation.

4.2 Defining Outcomes and Evidence

The second stage translates the need into observable language performance. Every resource should record the targeted proficiency range, communication mode, Can-Do statement, cultural objective, key language functions, evidence of success, and accessibility requirements. A novice-mid resource might aim for learners to exchange basic information about meals using practiced questions and phrases. An intermediate-mid resource might require learners to narrate a past dining experience and handle an uncomplicated problem. The difference is not merely longer vocabulary; it concerns text type, time frame, autonomy, and communicative function.

Assessment evidence must be identified before generation. If the outcome is interpretive, evidence may include identifying main ideas, supporting details, text purpose, and cultural perspective from an authentic menu or advertisement. If the outcome is interpersonal, a static scripted dialogue is insufficient evidence because it eliminates negotiation. This stage makes AI serve curriculum rather than allowing the system's output to determine the curriculum.

4.3 Generating a Constrained Draft

The third stage uses a structured prompt containing the design record. The prompt should state the learners, outcomes, mode, source restrictions, length, language variety, prohibited stereotypes, accessibility requirements, and requested teacher notes. Where factual or cultural accuracy matters, the model should work from approved sources rather than inventing content. The prompt and model version should be retained because reproducibility is part of responsible design.

Generation may produce readings at several levels, role cards, comprehension tasks, rubrics, image descriptions, or alternative examples. Its chief advantage is not autonomous creativity but rapid variation. A teacher can request three versions of a text, compare them, and select the one best suited to learners. Yet every draft should be labeled unvalidated. Kohnke, Moorhouse, and Zou observe that conversational AI may support explanations, practice, feedback, and material creation in language education, while its inaccuracies and biases require teachers to cultivate AI literacy and verify outputs (540-541). The G-AIDE workflow formalizes that verification instead of relying on individual caution.

4.4 Teacher and Expert Validation

The fourth stage evaluates the draft with a common rubric. Linguistic review examines grammar, vocabulary, idiomaticity, register, discourse coherence, pronunciation guidance where relevant, and proficiency fit. Pedagogical review examines alignment, cognitive demand, scaffolding, interaction, assessment validity, accessibility, and feasibility. Cultural review examines factual accuracy, diversity, perspective, and stereotypes. Technical and ethical review checks sources, copyright status, personal data, platform requirements, and transparency.

Validation should involve more than one perspective for resources intended for broad distribution. A panel might include a K-12 German teacher, a postsecondary instructor, a highly proficient or native German speaker, and an instructional-design specialist. Their ratings can be summarized using means or medians, but agreement also matters. Inter-rater reliability may be assessed with an intraclass correlation coefficient for scaled ratings or Cohen's kappa for categorical decisions. Written comments should be coded thematically so that repeated problems, such as unnatural dialogue or overuse of stereotypes, inform prompt revision. A resource should advance only after critical errors are corrected and reviewers agree that it is safe to pilot.

4.5 Classroom Pilot

The fifth stage asks whether expert approval translates into classroom value. A small pilot can compare AI-assisted materials with teacher-designed or existing materials while avoiding unsupported claims of superiority. Evidence should include learning performance, teacher preparation time, revision time, student participation, accessibility, and teacher observations. Pre- and post-performance tasks may measure interpretive comprehension, vocabulary retention, or speaking and writing. Rubrics should match the learning outcome, and scorers should be trained when possible.

Student satisfaction alone is inadequate. Learners may enjoy a visually attractive worksheet that produces little language use. Conversely, a demanding interpersonal task may generate lower immediate satisfaction while producing

stronger evidence of communicative growth. Pilot analysis should triangulate outcomes: performance data indicate what students learned; observations show how the resource functioned; interviews or surveys explain participant experience; and design logs reveal the actual labor required.

4.6 Revision Documentation and Dissemination

The final stage converts a successful local resource into a reusable one. Each item should include title, intended learners, proficiency range, outcomes, communication mode, estimated time, prerequisites, accessibility notes, sources, prompt provenance, model and date, reviewers, revision history, licensing, and known limitations. Editable formats are essential because contexts differ. Version control allows errors to be corrected without leaving multiple unmarked copies in circulation.

A repository should permit teachers to report problems and should assign responsibility for periodic review. Portability across AI vendors is also important. A framework tied to one commercial platform may disappear with pricing or policy changes. The durable product is therefore not only the generated worksheet; it is the combination of design specification, approved content, validation record, and adaptation guidance.

5 ANALYTICAL CRITERIA FOR EVALUATING THE FRAMEWORK

The framework's value depends on a multidimensional analysis. Speed is only one outcome and should never stand as a proxy for quality.

5.1 Linguistic Accuracy and Pragmatic Appropriateness

German outputs should be coded for morphological, syntactic, lexical, semantic, and pragmatic errors. The analysis must distinguish severity. A misplaced comma is different from an incorrect case that changes meaning; an unusual but possible phrase is different from an invented idiom. Reviewers should also examine whether the language is natural for the represented speaker and context. AI often produces formally correct but textbook-like exchanges in which every turn is unnaturally complete. For beginners, simplified language may be pedagogically necessary, but simplification should not become distortion.

Proficiency appropriateness must be analyzed separately from correctness. A passage may contain no error yet overload novice learners with subordinate clauses, low-frequency vocabulary, and implicit cultural knowledge. Conversely, an advanced task may be too controlled to elicit sustained discourse. The rubric should examine sentence and text length, lexical frequency, cohesion, time frames, inferential demand, and expected learner autonomy. Comparing expert ratings with readability indicators can be useful, but automatic formulas cannot replace judgment because German compound nouns and inflection can distort surface measures.

5.2 Pedagogical Alignment and Transfer

Alignment analysis traces a direct chain from objective to activity to assessment. Reviewers should be able to state what learners do, what evidence demonstrates the outcome, and how the task prepares them for later independent performance. Materials that only rehearse recognition should not claim to assess spontaneous production. Likewise, a cultural objective cannot be assessed by recalling isolated facts if the intended learning concerns perspective or comparison.

Transfer is a stronger test than immediate completion. A resource is educationally valuable when learners can use language beyond the practiced item. A pilot might therefore include a parallel task with a different context: after practicing how to solve a restaurant problem, learners handle a problem at a hotel. Such analysis resists the temptation to equate successful worksheet completion with communicative competence.

5.3 Cultural Validity and Representation

Cultural validation is central rather than supplementary. AI-generated German materials frequently gravitate toward recognizable symbols such as Oktoberfest, castles, sausages, and Christmas markets. These topics are not inherently illegitimate, but their repetition can reduce complex societies to consumable traditions. A coding scheme should record geographic range, historical period, social identities, source perspective, authenticity, and stereotype. It should ask whether Germany is treated as synonymous with the German-speaking world and whether German identity is implicitly represented as white, Christian, monolingual, and ethnically homogeneous.

Reviewers should also distinguish inclusion from tokenism. Adding a migrant character's name does not make a unit culturally responsive if the character has no perspective or if migration appears only as a social problem. Cultural

materials should allow learners to interpret products, practices, and perspectives and to recognize disagreement within communities. This criterion is an original contribution of G-AIDE: linguistic fluency is not evidence of cultural validity.

5.4 Differentiation Accessibility and Equity

AI's capacity to generate variants is valuable only if adaptations preserve intellectual purpose. Simplifying a text for a learner with lower reading proficiency should not automatically remove cultural complexity. Accessibility analysis should examine document structure, alternative text, font and contrast, captioning, screen-reader compatibility, cognitive load, and clarity of directions. Teachers should also inspect whether an adaptation stigmatizes learners by making their version visibly inferior.

Equity includes technological access. A resource that requires each student to purchase an AI subscription or disclose personal information is not scalable for public education. Core materials should remain usable without continuous access to a commercial chatbot. AI may assist teachers behind the scenes, while students receive accessible outputs through established school systems. This arrangement also reduces the risk of transferring instructional responsibility to families with unequal technological resources.

5.5 Workload and Efficiency

Efficiency should be measured as net labor rather than generation time. The appropriate calculation is manual design time minus the combined time for prompting, generation, fact-checking, revision, formatting, and documentation. Researchers should record each component with design logs rather than asking teachers to estimate retrospectively. They should also count revision cycles and critical errors.

A comparative study could use three conditions: teacher-designed materials, unreviewed AI-generated materials, and AI-assisted teacher-reviewed materials. The same experts would score all conditions blindly using the validation rubric. Analysis could compare median quality scores, error rates, and total preparation time. The unreviewed condition is analytically important even if it is never used with students, because it isolates the value added by professional review. If AI-assisted materials save drafting time but require extensive correction, the result may still be useful, but the claim must be precise. Efficiency exists only where total workload falls without unacceptable loss of quality.

5.6 Scalability and Sustainability

Scalability is often invoked vaguely. Within G-AIDE it has at least five dimensions: adaptation across learner levels, use across institutions, affordability, technical portability, and governance over time. A resource that works in one class is transferable, not yet scalable. Scaling requires clear metadata, editable formats, training, quality thresholds, error-reporting procedures, and maintenance.

Sustainability also demands professional learning. Teachers need not become machine-learning engineers, but they must understand prompt constraints, hallucination, bias, data privacy, source verification, and the difference between linguistic form and communicative performance. Professional development should use actual flawed outputs so that teachers practice diagnosis rather than merely learn prompt formulas. Shared repositories can then function as communities of practice in which teachers contribute revisions and explanations, not as warehouses of decontextualized worksheets.

6 ETHICAL BOUNDARIES AND INSTITUTIONAL POLICY

The most serious danger is institutional substitution. An administrator may conclude that if AI can generate lessons, a qualified German teacher is less necessary. This conclusion confuses content production with teaching. Teachers establish relationships, interpret hesitation, respond to identity and emotion, manage interaction, diagnose misconceptions, create motivation, make ethical decisions, and accept responsibility. AI-generated material can support these activities but cannot assume their accountability. G-AIDE should therefore be adopted with an explicit non-substitution principle: efficiency gains are used to improve instructional quality and program reach, not to eliminate certified positions.

Privacy is equally important. Student names, writing, disability information, grades, immigration histories, and behavioral records should not be entered into public AI systems without an authorized agreement and informed policy. UNESCO emphasizes data protection, human agency, transparency, and age-appropriate use as central governance concerns (Miao and Holmes 22). Districts and universities should specify approved platforms, prohibited data, retention policies, accessibility requirements, and procedures for reporting harmful output. Teachers should disclose meaningful AI involvement in shared materials, while students should learn when and why AI was used.

Copyright and provenance require attention because models may produce text resembling protected material and may obscure the origin of factual claims. G-AIDE therefore favors teacher-supplied, licensed, public-domain, or openly licensed sources. Generated output should be checked against sources, and repository metadata should identify them. A prompt record is useful for audit but does not itself establish ownership.

Finally, cultural and linguistic authority must remain plural. Native-speaker review is valuable, but nativeness alone does not guarantee pedagogical skill or freedom from bias. Panels should include teachers from different educational levels and, where appropriate, people familiar with the represented communities. German is pluricentric, and validation should not incorrectly label Austrian or Swiss usage as defective simply because it differs from northern German norms.

7 IMPLICATIONS FOR RESEARCH AND POLICY

The framework suggests a research agenda more modest and useful than claims that AI will transform education. First, researchers should examine where preparation time is actually spent across German programs. Second, controlled comparisons should test whether AI assistance reduces net workload while maintaining expert-rated quality. Third, studies should identify recurrent German-specific errors and cultural biases across models and over time. Fourth, classroom pilots should measure language performance rather than satisfaction alone. Fifth, implementation research should compare well-resourced and under-resourced settings to determine whether AI narrows or widens inequity.

Policy makers should treat resource infrastructure as one component of a broader workforce strategy. Strong programs still require scholarships for teacher preparation, alternative certification that preserves rigor, mentoring, competitive compensation, exchanges, and institutional commitments to retain small-enrollment courses. The MLA report's case studies emphasize funding, administrative support, assessment, curricular relevance, and external partnerships as conditions of program strength (Lusin et al. 4-5). AI cannot replace these conditions. It can complement them by making reviewed materials and adaptation guidance available to teachers who lack local colleagues.

Professional associations are well positioned to host or certify repositories because they can convene expertise beyond a single vendor. A national German resource initiative might partner K-12 teachers, universities, German cultural organizations, disability specialists, and instructional designers. Resources could receive provisional, validated, or classroom-tested status so users understand the strength of evidence. Such a system would also turn validation into professional recognition: teachers who review, pilot, and improve materials would be credited as intellectual contributors rather than invisible correctors of machine output.

8 CONCLUSION

The German language teacher shortage is a structural problem expressed through labor scarcity, fragile programs, unequal access, and excessive workload. Generative AI does not resolve its central causes. It cannot certify or retain teachers, restore eliminated positions, or persuade institutions to value language education. Its legitimate contribution is narrower but still important: it can help qualified educators produce variations, differentiate materials, document alignment, and share resources with less duplicated effort.

G-AIDE defines the conditions under which that contribution becomes educationally defensible. The framework begins with need rather than novelty, outcomes rather than output, and professional accountability rather than automation. It subjects AI-assisted resources to linguistic, pedagogical, cultural, accessibility, workload, and ethical analysis before treating them as scalable. Most importantly, it locates teachers at every consequential decision point.

The framework, therefore, reframes the policy question. The goal is not to ask how many teacher functions an AI system can perform. It is to ask how carefully designed infrastructure can preserve teacher time for the work that requires human judgment and relationship. If implemented with investment, transparency, and a non-substitution commitment, AI-assisted instructional design can help vulnerable German programs sustain quality and reach. If implemented as a shortcut around professional labor, it will intensify the very shortage it claims to solve.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Disclosure of conflict of interest

There is no competing interest in this work.

REFERENCES

- [1] American Councils for International Education. The National K-12 Foreign Language Enrollment Survey Report. American Councils for International Education, June 2017, www.americancouncils.org/sites/default/files/FLE-report-June17.pdf.
- [2] Bender, Emily M., et al. "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Association for Computing Machinery, 2021, pp. 610-23, doi:10.1145/3442188.3445922.
- [3] Kohnke, Lucas, Benjamin Luke Moorhouse, and Di Zou. "ChatGPT for Language Teaching and Learning." RELC Journal, vol. 54, no. 2, 2023, pp. 537-50, doi:10.1177/00336882231162868.
- [4] Lusin, Natalia, et al. Enrollments in Languages Other Than English in United States Institutions of Higher Education, Fall 2021. Modern Language Association, 2023, www.mla.org/content/download/191324/file/Enrollments-in-Languages-Other-Than-English-in-US-Institutions-of-Higher-Education-Fall-2021.pdf.
- [5] Miao, Fengchun, and Wayne Holmes. Guidance for Generative AI in Education and Research. UNESCO, 2023, unesdoc.unesco.org/ark:/48223/pf0000386693.
- [6] Mishra, Punya, and Matthew J. Koehler. "Technological Pedagogical Content Knowledge: A Framework for Teacher Knowledge." Teachers College Record, vol. 108, no. 6, 2006, pp. 1017-54, doi:10.1111/j.1467-9620.2006.00684.x.
- [7] National Standards Collaborative Board. World-Readiness Standards for Learning Languages. 4th ed., ACTFL, 2015.
- [8] Wiggins, Grant, and Jay McTighe. Understanding by Design. Expanded 2nd ed., Association for Supervision and Curriculum Development, 2005.