HOW

Volume 33, Number 1, pages 43 - 67.

https://doi.org/10.19183/how.33.1.836

HOW
A Colombian Journal for English-language Teachers

AI-powered Natural Language Processing in Language Education: A Systematic Review

Procesamiento del Lenguaje Natural Basado en IA para la Educación en Lenguas: Una Revisión Sistemática

DANIEL MURCIA ✉

ORCID: https://orcid.org/0000-0002-3146-6116

✉ **Corresponding Author**

The Pennsylvania State University
234 Sparks Building, University Park, PA, 16802
danielmurcia@psu.edu

LUIS FELIPE JARAMILLO-CALDERÓN
Universidad tecnológica de Pereira

ORCID: https://orcid.org/0009-0002-8782-4047

Received:

March 31st, 2024

Accepted:

June 7th, 2025

Published

April 13th, 2026

REVIEW ARTICLE

https://doi.org/10.19183/how.33.1.836

How to cite this article (APA, 7th ed.):
Murcia, D., & Jaramillo-Calderon, L. F. (2026). AI-powered natural language processing in language education: A systematic review. *HOW, 33*(1), 43 - 67. https://doi.org/10.19183/how.33.1.836

Abstract

This systematic review explores the potential of Natural Language Processing (NLP) based Artificial Intelligence (AI) technologies to enhance language development in higher education. Following the PRISMA 2020 guidelines (Page et al., 2021), we reviewed (*n*=63) documents published between 2015 and 2023, exploring how NLP has been used in language education within processes of linguistic development, biliteracy instruction, and language assessment. The reviewed literature reveals exploratory integrations and empirical evidence of the impact of these technologies in language instruction, learning, and assessment, which sheds light on key application areas of AI-powered NLP tools, such as enhancing literacy instruction, providing personalized feedback, tracking students' progress and weaknesses, and promoting accessibility. Existing challenges and future directions include honesty concerns like plagiarism, factfulness, and hallucinations as ethical considerations that reveal the ongoing debates and efforts to leverage AI-powered technologies to current curricular approaches in higher education.

Keywords*:* artificial intelligence, digital literacy, language education, natural language processing, systematic literature review

Resumen

Esta revisión sistemática investiga el potencial del Procesamiento del Lenguaje Natural (PLN) basado en tecnologías de Inteligencia Artificial (IA) para mejorar el desarrollo lingüístico en la educación superior. A partir del protocolo PRISMA 2020 (Page et al., 2021), se revisaron (*n*=63) documentos publicados entre 2015 y 2023 para explorar el uso de PLN en la enseñanza de lenguas dentro de procesos de enseñanza de la (bi)literacidad y evaluación de lenguas. La literatura revisada revela integraciones exploratorias y evidencia empírica sobre el impacto de estas tecnologías en la enseñanza, el aprendizaje y la evaluación, lo cual permite identificar áreas de aplicación clave como el mejoramiento del desarrollo de la literacidad, la retroalimentación personalizada, el seguimiento del progreso de los estudiantes y sus respectivos aspectos de mejora, y la promoción de la accesibilidad. También se presentan desafíos existentes y direcciones futuras, así como consideraciones éticas que incluyen el plagio, veracidad de la información y las alucinaciones. Los trabajos revisados demuestran los debates y esfuerzos actuales por incorporar tecnologías impulsadas por IA en los enfoques curriculares de la educación superior.

Palabras clave: educación de lenguas, inteligencia artificial, literacidad digital, procesamiento del lenguaje natural, revisión sistemática de la literatura

Résumé

Cette revue systématique examine le potentiel des technologies d'intelligence artificielle (IA) fondées sur le traitement automatique du langage naturel (TALN) pour favoriser le développement linguistique dans l'enseignement supérieur. Conformément aux lignes directrices PRISMA 2020 (Page et al., 2021), nous avons analysé 63 documents publiés entre 2015 et 2023 afin d'explorer les usages du TALN dans l'enseignement des langues, notamment dans les processus de développement linguistique, d'enseignement de la bilittératie et d'évaluation des compétences langagières.

La littérature examinée met en évidence des intégrations encore exploratoires ainsi que des données empiriques attestant de l'impact de ces technologies sur l'enseignement, l'apprentissage et l'évaluation des langues. Ces travaux permettent d'identifier plusieurs domaines d'application majeurs des outils de TALN alimentés par l'IA, notamment l'amélioration de l'enseignement de la littératie, la fourniture de rétroactions personnalisées, le suivi des progrès et des difficultés des apprenants, ainsi que la promotion de l'accessibilité.

Par ailleurs, la littérature met en lumière divers défis et perspectives futures, notamment des enjeux liés à l'intégrité académique tels que le plagiat, la véracité de l'information et les phénomènes d'« hallucination » générés par les systèmes d'IA. Ces considérations éthiques témoignent des débats et des efforts actuels visant à intégrer de manière critique les technologies fondées sur l'IA dans les approches curriculaires de l'enseignement supérieur.

Mots-clés: Enseignement des langues, intelligence artificielle, littératie numérique, traitement automatique du langage naturel, revue systématique de la littérature.

Resumo

Essa revisão sistemática investiga o potencial do processamento de linguagem natural (PLN) baseado na tecnologia da Inteligência artificial (IA) para melhorar o desenvolvimento linguístico no ensino superior. A partir do protocolo PRISMA 2020 (Page et al., 2021), foram revisados (n=63) documentos publicados entre 2015 e 2023 para explorar o uso do PLN no ensino das línguas em processos de ensino do (bi) letramento e avaliação das línguas. A literatura revisada revela integrações exploratórias e evidência empírica sobre o impacto dessas tecnologias no ensino, aprendizagem, e avaliação, o que permite identificar as áreas de aplicação chave como o melhoramento do desenvolvimento do letramento, a retroalimentação personalizada, o seguimento do processo dos estudantes e seus respetivos aspetos a melhorar, e a promoção da acessibilidade. Também se apresentam desafios existentes e futuras direcções, bem como considerações éticas incluindo o plágio, a veracidade da informação, e as alucinações. Os trabalhos revisados demonstram os debates e esforços atuais por integrar tecnologias impulsadas pela IA nos enfoques curriculares dos ensino superior.

Palavras chave: Ensino das línguas, Inteligência artificial, Letramento digital, Processamento de linguagem natural, revisão sistemática da literatura.

Introduction

NLP as a subdiscipline of computational linguistics has been propelled by the use of AI-powered applications (Eisenstein, 2019; Meurers, 2012). This progress in the field has captured the attention of researchers and language educators (Bond et al., 2024; Chalmers et al., 2023; UNESCO, 2023), who are eager to explore the varying ways in which AI could enhance administrative and educational processes such as teaching, learning, assessment. We differentiated the NLP area from the intricate web of AI because it provides a more focalized compilation of studies that could build a specific rationale for researchers and those interested in the field.

Even though NLP is a subfield of AI, previous approaches did not embrace AI technologies. Among these methods, prevalent approaches included rule-based systems, statistical NLP, and text classification. Since these systems lack the flexibility and complexity of modern techniques at the service of language education and assessment, we have decided to include in the review only those that include deep learning and machine learning-based NLP, such as Duolingo English Test, OpenBook, arText, iSTART, AutoTutor, CyWrite, etc. (see Findings below).

NLP techniques can be used to analyze, understand, and generate text, speech, and other forms of human language, encompassing the analysis, comprehension, and generation of text, speech, and other forms of human communication. A typical example could be the use of Intelligent Tutoring Systems (ITS) to provide personalized feedback to students' responses within a digital course. NLP-based AI has had significant advancements in the last years, such as Generative Pre-trained Transformers (Vaswani et al., 2017), which have rendered powerful processors and generators of natural language. Even so, language researchers and teachers still explore the possibilities and potential of these technologies in university research and classroom routines. These include language development, accommodations, test item generation, personalized feedback, and adaptive instruction, among others, which will be characterized in the review of the papers.

Universities have an intricate relationship with the social and productive areas of society. This relationship has led to several advancements across various domains, and so is the case with AI, which follows that nature. NLP-based AI has emerged as a promising tool in the field of language development, allowing for state-of-the-art approaches and expanding the scope of educational practices.

Literacy is an essential skill for success in higher education (Olave et al., 2013). Its development, typically attributed to the school, also manifests in higher education as students face increasingly demanding texts or are requested to produce sophisticated written content. Nowadays, literacy goes beyond being able to read or write (New London Group, 1996), leading to a revised framework of [academic] digital literacies and AI Literacy (Chan & Colloton, 2024). This is, as stated in Pegrum (2019, p. 462), "the ensemble of skills needed to effectively manage meaning in an era characterized by digitally networked communications, including blended communications that exploit both analogue and digital channels …" As some students struggle with literacy, for instance, students not grasping what they read in complex expository texts (McNamara, 2022), traditional educational methods are often not effective enough for all learners or some higher education contexts. As an example, some students require an amount of ongoing feedback that traditional settings may not be able to provide regularly.

Additionally, literacy development has marked challenging prospects. Some of these are poor funding, lack of professional development programs, absenteeism, attrition rates, and diversity and

identity issues (Lippert et al., 2019). These converge in the need for more personalized assessment and large-scale marking, which has proven to be of high cost. Thus, the need for a link between NLP, [AI] Literacy Development, and Language Assessment appeals as the backbone that structures this paper.

NLP-based AI technologies offer fresh prospects for AI literacy development in higher education. These different applications are under development and studied across multiple functionalities as personalized feedback, identifying areas of weakness, and tracking student progress, among other functionalities, and through varied academic settings across the world (see Tables 2 and 3 below). Hence, the focus of this systematic review aims to provide a comprehensive snapshot and holistic view of the state of the art to offer valuable insights to classroom practitioners and pave the path for future research.

As NLP-based AI tools evolve, increasing their accuracy and sophistication, they will be more proficient at performing text understanding and generating tasks (Attali et al., 2023). This poses both challenges and opportunities for literacy development processes, which have been explored earlier by Murcia et al. (2025). The authors highlighted the increasing risk of Gen-AI dependency, attrition of academic trust, and loss of creativity and criticality while also presenting the possibility of critical uses to enhance student autonomy, assessment practices, and professional development processes. Yet, the possible applications are mostly unknown since the integration of these tools into language instruction, learning, and their incidence in the language assessment processes in classrooms are still underexplored. Thus, a solid conceptual and historical common ground should be established to open a potential research agenda.

Exploring these integrations in the academic sphere is of paramount importance, considering the opportunities, challenges, and ethical considerations. An active collaboration between researchers and educators should exist since this synergy may enable the development and implementation of more effective NLP-based AI interventions for literacy development. However, there seems to be scarce collaboration between AI developers and educators because of the perceived distance between their disciplines, which may limit the potential of such technologies for literacy development processes. As researchers and educators from different disciplines collaborate, they will be able to develop and implement more effective interventions for literacy development and language assessment. Thus, an important step is to explore the available research data about the application in higher education from 2015 up to 2023, to offer a baseline for further collaboration and exploration in the emerging interdisciplinary field. This review is constructed under the following research questions:

1. How has NLP been used in literacy instruction and language assessment in Higher Education institutions from 2015 to 2023?
2. What are the prevalent NLP software tools used in language education for higher education institutions in the period?
3. What challenges, future directions, and ethical concerns does the reviewed literature show about the implementation of NLP engines in language education?

Methodology

The aim of conducting a systematic review is to address inquiries by employing a clear, systematic, and reproducible search strategy. This process involves establishing inclusion and exclusion criteria, facilitating the identification of studies (Gough, Oliver & Thomas, 2017). In this

sense, the review was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) 2020 guidelines (Page et al., 2021).

Data Sources and Literature Search

The initial searching stage was conducted through three databases: *Scopus, Web of Science (WoS),* and *SpringerLink*. Only studies from 2015 to 2023 were included to ensure the review incorporated the most up-to-date literature available at the time of writing this paper. The search included three main concepts: Natural Language Processing, Language Education, and Higher Education; yet, all of them have possible sub-terms that could be found in the literature. Thus, the sub terms are presented in Table 1.

Table 1.

Keywords for Database Search

Main term	Sub-terms with "OR" Boolean operators
Natural Language Processing	"Natural Language Processing" OR NLP OR "Natural Language Understanding" OR NLU OR "Natural Language Assessment" OR NLA OR "Intelligent Tutoring System" OR "Text Summarization" OR Chatbot
Language Education	"LANGUAGE EDUCATION" OR Literacy OR "Literacy Instruction" OR "Literacy Development" OR "Computer-assisted writing" OR "Intelligent Computer Assisted Language Learning" OR ICALL OR Reading OR Writing "Language Development"
Higher Education	"Higher Education" OR University OR College OR Education

For all the databases, the same search strategy was used:

("Natural Language Processing" OR NLP OR "Natural Language Understanding" OR NLU OR "Natural Language Assessment" OR NLA OR "Intelligent Tutoring System" OR "Text Summarization" OR Chatbot) AND ("LANGUAGE EDUCATION" OR Literacy OR "Literacy Instruction" OR "Literacy Development" OR "Computer-assisted writing" OR "Intelligent Computer Assisted Language Learning" OR ICALL OR Reading OR Writing "Language Development") AND ("Higher Education" OR University OR College OR Education).

Study Selection Process

With this search strategy, we found 554 documents in *Scopus*, 671 in WoS, and 1054 in SpringerLink, making it a total of (*n*=2279), which were later screened to a final compilation of (*n*=63) studies (see Table 3 in the Findings).

Before the screening process, we filtered the duplicate articles and deleted them (ending up with 1914 documents). During the screening and eligibility processes, the inclusion and exclusion criteria were implemented (see Table 2).

Table 2.

Inclusion and Exclusion Criteria

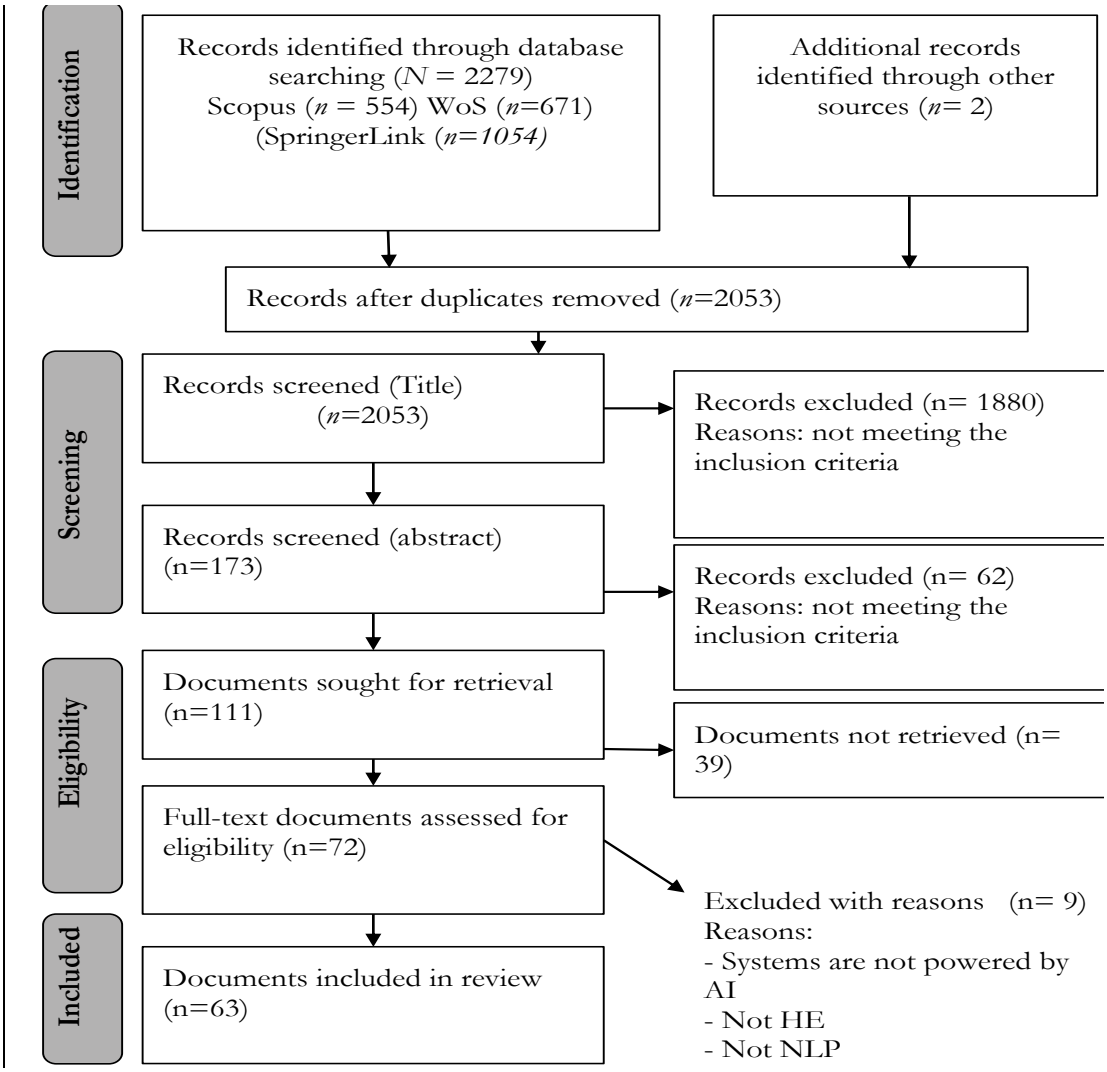
Inclusion criteria	Exclusion criteria
• The article is accessible to the researchers and more focused on the educational factors than the technical ones.	• The article develops statistical or numerical-based information that is not accessible to the readers.

• Articles are in English, Spanish, French, or Portuguese.	• In languages other than English, Spanish, French, and Portuguese.
• Researchers can access the full document.	• Articles cannot be downloaded.
• Articles are within the 2015 and 2023 span.	• Older than 2015.
• Articles related to AI and subfields.	• Do not relate to AI.
• Belong to a higher education or adult education context.	• Not about a higher education or adult education context.

The screening process (see Figure 1 below) had two stages: title screening and abstract screening. For both stages, researchers screened the records individually, labeling them as classified, unclassified, and neutral. The process showed an Inter-rater agreement of 0.92 for the title screening stage. For the abstract screening, the two researchers analyzed the dataset and decided upon the inclusion and exclusion criteria to select the articles. Both researchers discussed discrepancies in labeling or neutral records. Based on the inclusion/exclusion criteria, we reached a consensus on whether the document was moved to the next stage.

Figure 1.

Screening Process



Data Analysis

To extract data, articles were processed through ATLAS.ti 23 data analysis software using a content analysis approach: a method to generate codes from text which may turn into themes and findings. Initial codes were established based on the research questions and the documents were initially coded by a research assistant. Later, researchers reviewed the initial coding process and established pertinent themes for the study. These themes were then categorized into five subsections.

Findings

General Characteristics of the Review

In total, this systematic review included (*n*=63) articles, most of which are first-hand research articles (*n*=47). Additionally, a small percentage of articles are systematic reviews (*n*=8) and second-hand articles (*n*=8), both of which help to better understand the overall approaches, practices, and concerns regarding the use of NLP-based AI in HE for language education and literacy development. Table 3 presents a summary of the articles reviewed, considering the authors, year, and topic.

Table 3.

Articles Reviewed

#	Authors	Year	Topic	Type	Literacy focus	Country
1	Allen et al.	2017	Analysis of student writing	Research Article	L1	USA
2	Allen et al.	2015	Student's reading analysis	Research Article	L1	USA
3	Alrashid et al.	2022	AWE for reflective writing	Research Article	L1	UK
4	Attali et al.	2022	Automatic Item Generation	Research Article	L2	USA
5	Barbu et al.	2015	Text simplification	Research Article	L1	Europe
6	Bauer et al.	2023	Diagnostic assessment	Research Article	L1	Germany
7	Bradác et al.	2022	Personalized study plans	Research Article	L2	Czech Republic
8	Burstein et al.	2017	AWE	Research Article	L1	USA
9	Cerga-Pashoja et al.	2019	Assisted reading for autism	Research Article	L1	Europe
10	Chen et al.	2022	Educational technology	Review Article	L1/L2	China
11	Chong et al.	2020	Reflective writing analysis	Research Article	L1	Malaysia

12	Contreras et al.	2019	Automated Essay Scoring (AES) Question Generator	Research Article	L1	Malaysia
13	Crompton & Burke	2023	AI in HE	Review Article	L1/L2	USA
14	Demir	2020	Intelligent tutoring system (ITS) in English Language Education	Research Article	L2	Turkey
15	Dergaa et al.	2023	AI-generated text	Research article	L1	Asia
16	Fang et al.	2019	Conversation-based ITS for reading	Research Article	L1	USA
17	Feng et al.	2016	Automated error detector for ESL learners	Research Article	L2	USA
18	Fryer et al.	2017	Chatbot vs humans on language course interest	Research Article	L2	Japan
19	Fu & Gu	2020	Automatic scoring effect on continuous learning	Research Article	L2	China
20	Fung et al.	2020	Question generation for reading comprehension	Research Article	L2/L1	China
21	Gao & Passonneau	2021	AWE	Research Article	L1/L2	USA
22	Houston & Corrado	2023	LLM on Academia	Second-hand	L1/L2	USA
23	Huang et al.	2023	AI in Language Education	Review Article	L2	China
24	Jeon & Lee	2023	AI and human teacher collaboration	Research Article	L2	South Korea
25	Jeon et al.	2023	Chatbots for language learning	Review Article	L2	USA
26	Katsarou et al.	2023	Intelligent Virtual agents (IVA) on EFL education	Review article	L2	Europe
27	Kuhail et al.	2022	Educational Chatbots	Review Article	L1/L2	Non-specified
28	Li & Graesser	2020	IVA on summary writing	Research Article	L2	USA
29	Li et al.	2018	Students' academic language within an ITS	Research Article	L1	USA
30	Liang et al.	2021	AI in language education	Review article	L2/L1	Taiwan
31	Lim et al.	2022	Neural AWE for L2 writing	Research Article	L2	China
32	Lippert et al.	2019	Adaptive ITS for literacy development	Research Article	L1-L2	USA

33	Liu et al.	2017	Automated feedback generation	Research Article	L2	China
34	Liu et al.	2022	AI in Situational teaching	Review Article	L1/L2	Asia
35	Maqsood et al.	2022	AWE	Research Article	L2	Asia
36	McNamara et al.	2023	Adaptive comprehension and stealth literacy assessment	Research Article	L1	USA
37	McNamara et al.	2022	Adaptive comprehension and stealth literacy assessment	Research Article	L1	USA
38	Miaschi et al.	2021	Tracking L1 writing evolution	Research Article	L1	Italy
39	Minoofam et al.	2021	Dyslexic students' instruction	Research Article	L1	Iran
40	Miranda	2020	ITS for vocabulary acquisition	Research Article	L2	Philippines
41	Nehyba & Stefánik	2022	Deep learning and reflective writing	Research Article	L1	Czech Republic
42	Nguyen et al.	2022	Edtech for Codeswitching	Research Article	L2	UK
43	Nuñez & da Cunha	2021	Impact of gen-IA on the student writing process	Research Article	L1	Spain
44	Öncel et al.	2021	AWE on source-based writing	Research Article	L1	USA
45	Pengel et al.	2021	Chatbots for Self-regulated learning and automatic writing feedback	Research Article	L1	Germany
46	Pokrivcakova	2019	Teacher training for AI in EFL	Review Article	L2	Czech Republic
47	Rapp & Kauf	2018	Thesis Writing Instruction	Research Article	L1	Non-specified
48	Reyes R.	2019	Virtual assistants in education	Research Article	L1	Latin America
49	Rus & Ştefanescu	2016	Non-intrusive assessment with an intelligent tutoring system	Research Article	L1	USA
50	Salas-Pilco & Yang	2022	AI in HE	Review Article	L1/L2	Latin America
51	Shehab et al.	2016	AES (automated essay scoring)	Research Article	L1	Asia
52	Steuer et al.	2022	Automatic question generation for L2 reading	Research Article	L2	Egypt

53	Strobl et al.	2018	Digital support for academic writing	Review Article	L1/L2	Europe
54	Taele et al.	2020	Intelligent tutoring system for L2 writing	Research Article	L2	USA
55	Tyen et al.	2022	Chatbot for language practice	Research Article	L2	UK
56	Venegas	2021	Text classification	Research Article	L1	Latin America
57	Vitartas et al.	2019	ITS for assignment checking	Research Article	L1	Australia
58	Wambsganss et al.	2022	Automated feedback for argumentative writing	Research Article	L1	Europe
59	Whitelock & Bektik	2018	Automated scoring and Feedback for Large-scale assessment	Review Article	L1/L2	UK
60	Yan	2023	Chatbot impact on L2 writing	Research Article	L2	China
61	Yang & Stansfield	2022	Digital reading support through chatbots	Second-hand	L1/L2	UK
62	Zhang et al.	2023	chatbot-assisted learning	Review Article	L1	China
63	Ziegler et al.	2017	CALL, NLP, and second language acquisition	Research Article	L2	USA

All articles document the development and use of AI for education purposes with adult populations and higher education to promote language development. Some of these studies explore second language development, while others foster students' first language. The key findings of this review are classified into main application areas, these being NLP-based AI-assisted Language Instruction, Language Assessment and Feedback, as well as Challenges and Future Directions, and Ethical Considerations.

The geographical distribution of the review (eight countries/regions) is presented in Table 3. The data reveals that most of the articles are published in the USA (n=19) followed by Europe (n=13), whilst Australia only presents (n=1) document for consideration. The non-specified articles describe the study without disclosing their geographical location.

Overall, the analysis of the reviewed articles reveals a significant growing body of evidence supporting the implementation of NLP-based AI for Language Education. The following sections detail insights that highlight the advancements and critical areas presented in the studies.

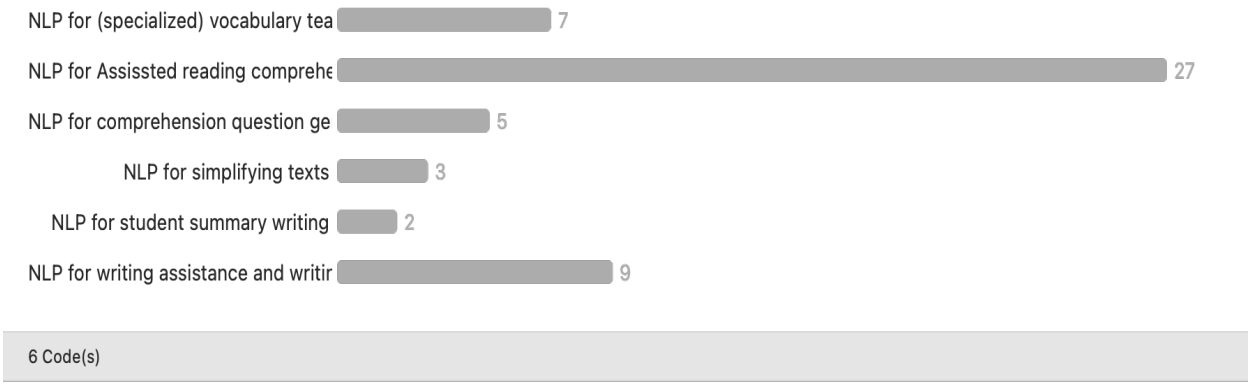
NLP-based AI and Language Instruction

One of the primary applications of AI technologies identified in our review was supporting language instruction, with (n=21) of the total reviewed documents discussing this area. Figure 2 shows the number of citations/quotes that were generated out of the 21 documents, which shows the relevance or occurrence of each specific strategy among the documents. The main ways in which this

goal was attained was through the teaching of (specialized) vocabulary (*n*=7 codes); to support reading comprehension and student's reading skills (*n*=27 codes); to generate comprehension questions for students to answer when engaging in reading (*n*=5 codes); to support writing and foster writing skills (*n*=9 codes) among others.

Figure 2.

Prevalent Categories of NLP-based AI for Language Development



Several studies in the review explored the use of NLP for vocabulary development, suggesting its potential effectiveness and diverse applications within higher education. To illustrate, Miranda (2020) investigated an intelligent language tutoring system for acquiring Kapampangan vocabulary, showcasing its potential for first language learners. Additionally, Huang et al. (2023) mentioned a context-aware vocabulary learning system that personalized word suggestions based on individual learning styles and available time. This aligns with the findings of Liang et al. (2021), who identified vocabulary acquisition as a core application domain within AI-powered language education (AILEd) research. These early explorations suggest that NLP can offer engaging and effective tools for vocabulary development in higher education settings.

Additionally, most studies suggest the potential of NLP for assisting reading comprehension and reading skills in diverse populations. For instance, Cerga-Pashoja et al. (2019) developed an open-source platform called OpenBook using NLP to specifically aid reading and comprehension for individuals with Autism Spectrum Disorder (ASD). OpenBook is a personalized platform that simplifies text by identifying language obstacles, adding definitions, synonyms, images, and formatting adjustments, and includes features tailored separately for independent users and caregivers. This aligns with the work of McNamara et al. (2022), who explored how interactive NLP technologies can support comprehension by prompting students to connect textual information with their prior knowledge, fostering deeper understanding.

Minoofam et al. (2021) investigated the Reading Assistant Learning Framework (RALF). This NLP-based tool demonstrated positive impacts on dyslexic students' reading fluency (i.e., 27% improvement among Persian primary students), comprehension, and word recognition. RALF is designed to support Persian-language reading skills through audio-visual tools, adaptive text generation, and virtual pen guidance. It incorporates simplified font structures, reinforcement learning strategies for personalized content, and interactive exercises and assessments. These encouraging findings

showcase the potential of NLP to personalize reading experiences, cater to individual needs, and ultimately improve reading skills for learners across various contexts.

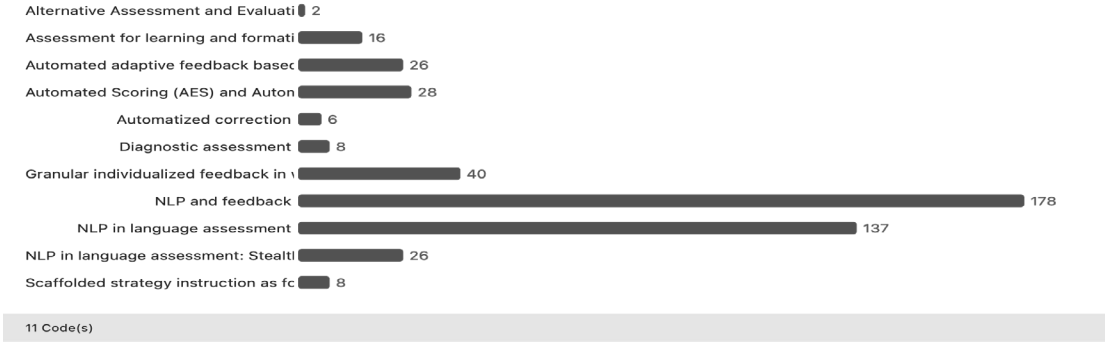
Enhancing writing skills through AI-powered assistance has also garnered attention in higher education research. Feng et al. (2016) developed CyWrite, a tool utilizing NLP to support not only testing but also the teaching and learning of second language (L2) writing. CyWrite analyzes various aspects of writing, from grammar and spelling to stylistic choices and discourse patterns, and provides targeted feedback throughout the writing process. Similarly, McNamara et al. (2022) explored the use of NLP-based algorithms for delivering personalized feedback during writing practice. For this purpose, researchers implemented the Interactive Strategy Training for Active Reading and Thinking (iSTART). This is an adaptive intelligent tutoring system designed to enhance students' comprehension of complex scientific texts in STEM. Guided by learning sciences principles, such as the generation effect, deliberate practice, and engagement strategies, iSTART delivers interactive instruction and personalized practice in literacy skill development. While these studies primarily focused on L2 writing, Miranda (2020) investigated the potential of an NLP-driven system for script writing skills in first language learners. These diverse applications suggest NLP's potential to provide personalized feedback, address specific writing challenges, and ultimately improve writing skills across various contexts and languages.

Language Assessment and Feedback

Assessment and feedback are paramount aspects of any education process, and they have been at the core focus of AI research and development (see Figure 3). In fact, (n=52) articles in this review were concerned to some extent about assessment and feedback opportunities of NLP-based AI technologies. Among the areas of interest that were identified are: granular individualized feedback in writing, Automated Essay Scoring (AES) and Automated Writing Evaluation (AWE), Automated adaptive feedback based on NLP and virtual writing assistants, and stealth assessment.

Figure 3.

Prevalent Categories of NLP-based AI for Language Assessment and Feedback



One particularly exciting application involves granular individualized feedback, where feature-based models of AI provide students with detailed, personalized insights into their writing. This approach offers several advantages over traditional methods; for instance, AI can pinpoint specific errors in grammar, style, or mechanics in a matter of seconds. Students can then utilize these insights

to identify their weaknesses and target areas for improvement, fostering a more self-directed learning process (Demir, 2020). Furthermore, Vitartas et al.'s (2019) AI-powered feedback tool can be delivered promptly, allowing students to check their work iteratively before final submission. The software digitally assessed students' assignments and provided descriptive dashboard-based feedback based on categories such as word count, grammar errors, formatted references, performance scores, readability, and spelling errors, among others. The system uses content analysis with customizable evaluation criteria established by the instructors, which highlights areas of improvement. This is especially beneficial for instructors with large class sizes, as it can provide personalized feedback to all students, even when human resources are limited (Pengel et al., 2021).

According to Wambsganss et al. (2022), students have also reported positive experiences with AI feedback, appreciating the detailed analysis in structural composition, argument support, and its ability to motivate them to write more effectively. Hence, granular individualized feedback powered by AI presents a promising path for enhancing writing experiences and fostering literacy development.

AI can also be developed for Automated Essay Scoring (AES) and Automated Writing Evaluation (AWE). These systems analyze student writing and provide feedback, potentially reducing workload for instructors and offering students immediate assessment. AES focuses on assigning a score or grade to an essay, like a human rater; these systems were identified in works like Taele et al. (2020) and Chong et al. (2020). This allows for efficient grading, particularly for large writing assignments. AWE, however, expands on this by offering not just a score, but also formative feedback that helps students improve their writing. For instance, in the study of Rapp & Kauf (2018), the AWE is called: Thesis Writer. This software provided writing tutorials, phrase banks, examples, and linguistic support that accompanied students' writing of their texts.

Another significant area of interest AI offers is stealth assessment. This approach embeds assessment tasks seamlessly within learning activities, often in the form of games or simulations (McNamara et al., 2022). Students remain focused on games such as Balloon Bust, a balloon popping game where students play to demonstrate understanding; meanwhile, they are unaware that their actions and choices are being analyzed to gauge their understanding. This method not only reduces test anxiety but also allows for continuous, data-driven assessment, as presented by Rus and Ştefanescu (2016). By analyzing student interactions within the game, AI can identify areas of strength and weakness, providing valuable insights for both students and teachers. Furthermore, NLP techniques can be used to analyze student language use within the game, offering a deeper understanding of their comprehension (Allen et al., 2015).

Challenges

Most articles (n=48) identified several challenges of NLP-based AI in language education. For instance, several researchers faced limited performance in their systems, which indicates that there is still room for improvement. A key area demanding attention is the generation of effective assessment and feedback. Despite the potential of NLP techniques, several studies point to limitations in their ability to provide nuanced and individualized feedback. For example, one specific challenge identified by Allen et al. (2017) is the reliance on aggregate metrics to evaluate student language. Their study reveals that many assessment tools calculate factors like lexical sophistication based on averages, neglecting word distribution within the text. This approach overlooks valuable insights into the student's writing style and coherence. Nuñez and da Cunha (2022) provide another example, highlighting a

student's dissatisfaction with arText's recommendation to shorten sentences. The student perceived the suggested solution as compromising clarity and information density. This instance underscores the need for feedback that considers contextual factors and aligns with individual writing styles.

Despite the surge in AI-powered literacy tools, educational research on their impact and the presence of education in AI research lags behind. For instance, a lack of studies analyzing evidence-based chatbot-learner interaction design techniques in education, despite the growing popularity of chatbot-assisted learning. Similarly, Fu and Gu (2020) point to the limited exploration of how AI-enabled automatic scoring applications affect learning persistence and motivation. These knowledge gaps hinder our understanding of the specific mechanisms through which AI tools influence learning outcomes and student engagement. Furthermore, Demir (2020) emphasizes the difficulty of designing effective negative feedback mechanisms in educational AI software. Identifying and addressing student mistakes requires a nuanced understanding of learning processes, a challenge that highlights the need for closer collaboration between researchers and educators.

Despite impressive advancements, concerns emerge about the quality and accuracy of AI-generated language in educational contexts. Rapp and Kauf (2018) and Liu et al. (2014) highlight problems with automatic question generation, noting its dependence on the quality of the provided text and susceptibility to technical issues. Similarly, Katsarou et al. (2023) found no significant improvement in learners' speaking skills due to the simplistic nature of dialogue content in AI-powered language interactions. Perhaps more concerning is the illusion of fluency presented by models like ChatGPT. While capable of generating seemingly coherent text, this fluency often lacks the deeper logic, judgment, and creativity expected from human writers (Houston & Corrado, 2023). This raises concerns about potential misuse and the need for critical evaluation skills when encountering AI-generated language.

Future Directions

Regarding future directions (n=44), while AI's potential to enhance existing educational practices is evident, the future may hold more possibilities. For instance, Liang et al. (2021) urge exploring new learning paradigms and teaching strategies specifically designed for AI-powered language learning (AILEd). This involves identifying ways AI can go beyond simply automating existing approaches to truly transform how to teach and learn. Similarly, Crompton and Burke (2023) call for research exploring unmapped affordances of AI in higher education. This includes investigating how AI can assist both instructors and learners, potentially revolutionizing roles and responsibilities within the educational ecosystem. In this line, research like that of Yang and Stansfield (2022) points towards integrating AI chatbots into augmented reality environments, creating richer, multi-sensory learning experiences that transcend traditional classroom settings. Katsarou et al. (2023) highlight the potential of advanced natural language processing and machine learning to create more natural and engaging Intelligent Virtual Agents (IVAs) for language learning. These agents could provide personalized feedback, facilitate interactive dialogues, and offer a more immersive learning experience.

Another crucial future direction lies in enhancing adaptability to individual users. This goes beyond tailoring generated content and interfaces to different languages but also involves dynamically adjusting the complexity and difficulty of the AI-powered interactions. Tyen et al. (2022) call for generating messages at different difficulty levels, like A1 for beginner learners, and suggest adjusting complexity through sentence structure and paraphrasing. Lippert et al. (2019) emphasize the potential of adaptive learning systems that cater to individual needs and learning styles, drawing upon AutoTutor's

ability to adjust instruction based on learner behavior and background.

Furthermore, Minoofam et al. (2021) address the linguistic challenges of representing different writing systems, particularly complex characters like Chinese, and propose innovative solutions such as using serifs to differentiate similar letters. This highlights the need for AI tools that can effectively handle the nuances of diverse languages. Yan (2023) identifies a gap in implementing advanced NLP tools for L2 writing instruction, suggesting that further research is needed to leverage these technologies for effective language learning across different languages. Additionally, Nguyen et al. (2022) point out the scarcity of multilingual NLP tools that can handle simultaneous processing of multiple languages, hindering truly bilingual interactions.

By addressing these challenges and focusing on adaptability to individual needs and languages, AI-powered literacy tools can move beyond generic approaches and deliver truly personalized and effective learning experiences for all learners, regardless of their proficiency level or native language. This will require ongoing research and development to create multilingual, adaptive systems that can cater to the diverse needs of users in a globalized learning landscape.

Ethical Considerations

While the potential benefits of AI in language instruction are undeniable, ethical concerns regarding its implementation cannot be ignored. Interestingly, only (n=12) documents mention ethical concerns or considerations related to the use of AI in education. Nonetheless, the documents that treat it offer significant insights into the major concerns that scholars should focus on when working with AI and having students interact directly with these technologies. The main ethical concerns, as evidenced in Figure 4, include academic honesty concerns, including issues on plagiarism, concerns on automatic writing processes, and factfulness and hallucinations in generated content.

Figure 4.

Prevalent Categories of Ethical Considerations

One of the most pressing ethical concerns surrounding AI-powered literacy tools centers on academic honesty. Studies like those by Yan (2023) highlight student anxieties about the potential for these tools to facilitate plagiarism. The ease with which AI can generate essays and other academic content raises concerns about the integrity of student work and the erosion of essential learning skills. Dergaa et al. (2023) further explore these concerns, pointing to the risk of overreliance on Generative AI Models like ChatGPT by researchers and students. This overdependence can lead to the creation of inaccurate and unoriginal work, potentially constituting plagiarism.

Another significant ethical concern involves the use of AWE systems that analyze and grade student work. Studies like Whitelock and Bektik (2018) highlight the opposition from writing professionals who fear AWE may lead students to prioritize writing for machines rather than human audiences. They argue that the social aspect of writing is crucial, and human readers can provide more nuanced feedback that fosters critical thinking and growth. In a similar line, Strobl et al. (2018) delve deeper into concerns regarding the reliability and pedagogical value of AWE systems. While efficiency is undeniable, questions arise about the accuracy of scoring engines and whether they can adequately capture the complexities of good writing. Additionally, Cotos (2015, as cited in Strobl et al., 2018) warns against using AWE solely for grading, highlighting the importance of human interaction and formative feedback. AWE systems struggle to provide the kind of scaffolding and rhetorical feedback that can truly guide students in developing their writing skills.

Another ethical concern arises from the potential for AI-powered tools to generate factual inaccuracies or hallucinations. Studies like Tyen et al. (2022) acknowledge that large language models can sometimes fabricate information during text generation. While ongoing research aims to mitigate this issue (Zhao et al., 2020; Komeili et al., 2022), it highlights the importance of user awareness. Learners must be equipped to critically evaluate the information presented by AI tools and distinguish between factual content and potential fabrications. Dergaa et al. (2023) raise a broader concern, suggesting that the impressive capabilities of NLP technology could make it difficult for even reviewers to discern fake scientific articles from genuine ones. This possibility underscores the importance of transparency and responsible development practices within the field of AI. Developers and educators alike must prioritize methods for ensuring the factual accuracy of AI-generated content and empowering users to critically evaluate the information they encounter. After all, this is a strong call for AI literacy development.

As a final note and the subsequent ethical considerations, authors like Yan (2023) suggest that unequal access to these technologies because of infrastructure or subscription fees may further increase educational inequalities since the ones with the best access and knowledge about these technologies have "enormous advantages to outperform" their peers (p. 14). Regarding data privacy, like Salas-Pilco and Yang (2022), the review does not show it as a pressing issue in most of the reviewed articles. One of the few articles was Wambsganss et al. (2022), which highlighted the risk of exposing students' argumentation data to large tech companies and called for ways to balance accessibility with data protection.

Conclusions

This systematic review has explored the diverse applications of AI-powered NLP tools for enhancing language instruction across various higher education contexts from 2015 to 2023. It has also explored the main challenges, opportunities, and ethical concerns of such tools. NLP-based AI

technologies offer new opportunities for literacy development and assessment since they can be used to enhance literacy instruction, provide personalized feedback, identify areas of weakness, track student progress, and promote accessibility for diverse learners. As NLP techniques continue to improve and become more widely available, they will play an increasingly important role in contemporary and prospective literacy and language education.

The reviewed literature highlights the potential of various software tools to enhance language education. Intelligent Tutoring Systems like iSTART, AutoTutor, and W-Pal offer personalized instruction, (stealth) assessment, and feedback, while game-based systems like iSTART can increase student engagement and scaffold learning of challenging topics. Additionally, tools like Thesis Writer and Open Book demonstrate the potential for supporting student productions and personalizing text difficulty.

However, the findings also highlight crucial challenges and considerations for moving forward. Concerns regarding ethical implications, including academic honesty, limitations of automatic assessment, potential for AI-generated hallucinations, education inequality, and data privacy, call for careful implementation strategies. Additionally, the need for ongoing research that actively involves educators and focuses on the quality and adaptability of AI tools is of paramount importance.

There is no doubt that AI-powered tools will increasingly influence the way students develop their literacy in educational settings and the way educators approach instruction and assessment processes. Therefore, it is relevant to strengthen the merging point between educators and AI developers to better inform these systems from education theory and actual practices. In the same direction, it is increasingly important to emphasize the development of digital literacies among professors and students so that the use of these technologies is as effective, safe, and ethical as possible.

Implications and Future Directions

While the review did not identify statistically significant results for bilingualism and bilingual education (such as the processing of bilingual texts, code-switching, and biliteracy) in AI tools, we believe this is an emerging area that should be a topic of interest for further research at the merging point of AI and language education. Additionally, the data underscores the need for novel theoretical frameworks specifically designed for the educational use of AI-powered literacy tools. Furthermore, developing updated educational and language policies will be crucial to ensure the responsible and effective integration of these technologies into classrooms. The findings of this review can serve as a valuable resource for researchers and educators interested in AI-powered literacy tools. Scholars can use these to make informed decisions about the implementation and usage of tools for their educational needs and specific contexts.

Statement of originality

We declare that this manuscript is the result of original research and that it has not been published before and is not currently under evaluation in another journal.

Conflicts of interest

We do not have any potential conflict of interest to declare

Author Contributions

Daniel Murcia: Conceptualization, Resources, Formal Analysis, Methodology, Visualization, Software, Writing-Reviewing and Editing.

Luis Felipe Jaramillo-Calderón: Conceptualization, Data curation, Methodology, Writing-Original draft preparation.

References

Allen, L. K., Likens, A. D., & McNamara, D. S. (2017). *Recurrence Quantification Analysis: a technique for the dynamical analysis of student writing*. Grantee Submission. http://files.eric.ed.gov/fulltext/ED585783.pdf

Allen, L. K., Snow, E. L., & McNamara, D. S. (2015). Are You Reading My Mind? Modeling Students' Reading Comprehension Skills with Natural Language Processing Techniques. *Grantee Submission*. http://files.eric.ed.gov/fulltext/ED588531.pdf

Alrashidi, H., Almujally, N. A., Kadhum, M., Ullmann, T., & Joy, M. (2022). Evaluating an automated analysis using machine learning and natural language processing approaches to classify computer science students' reflective writing. In *Lecture notes in networks and systems* (pp. 463–477). https://doi.org/10.1007/978-981-19-2840-6_36

Attali, Y., Runge, A., LaFlair, G. T., Yancey, K. P., Goodwin, S., Park, Y., & Von Davier, A. A. (2022). The interactive reading task: Transformer-based automatic item generation. *Frontiers in Artificial Intelligence*, *5*. https://doi.org/10.3389/frai.2022.903077

Barbu, E., Martín–Valdivia, M. T., Martínez-Cámara, E., & Ureña-López, L. A. (2015). Language technologies applied to document simplification for helping autistic people. *Expert Systems with Applications*, *42*(12), 5076–5086. https://doi.org/10.1016/j.eswa.2015.02.044

Bauer, E., Sailer, M., Kiesewetter, J., Fischer, M. R., Gurevych, I., & Fischer, F. (2024). Facilitating justification, disconfirmation, and transparency in diagnostic argumentation. *Zeitschrift Fur Padagogische Psychologie*, *38*(1–2), 49–54. https://doi.org/10.1024/1010-0652/a000363

Bond, M., Khosravi, H., De Laat, M. et al. (2024). A meta systematic review of artificial intelligence in higher education: A call for increased ethics, collaboration, and rigour. *Int J Educ Technol High Educ, 21*, 4 . https://doi.org/10.1186/s41239-023-00436-z

Bradáč, V., Smolka, P., Kotyrba, M., & Průdek, T. (2022). Design of an intelligent tutoring system to create a personalized study plan using expert systems. *Applied Sciences*, *12*(12), 6236. https://doi.org/10.3390/app12126236

Burstein, J., McCaffrey, D., Klebanov, B. B., & Ling, G. (2017). Exploring Relationships between Writing & Broader Outcomes with Automated Writing Evaluation. *Grantee Submission*. https://files.eric.ed.gov/fulltext/ED598703.pdf

Chan, C., & Colloton, T. (2024). *Generative AI in higher education. The ChatGPT effect*. Routledge.

Cerga-Pashoja, A., Gaete, J., Shishkova, A. M., & Jordanova, V. (2019). Improving Reading in Adolescents and Adults with High-Functioning Autism through an Assistive Technology tool: a Cross-Over Multinational study. *Frontiers in Psychiatry*, *10*. https://doi.org/10.3389/fpsyt.2019.00546

Chalmers, H., Brown, J., & Koryakina, A. (2023). Topics, publication patterns, and reporting quality in systematic reviews in language education. Lessons from the international database of education systematic reviews (IDESR). *Applied Linguistics Review*. https://doi.org/10.1515/applirev-2022-0190

Chen, X., Zou, D., Xie, H., Chen, G., Lin, J., & Cheng, G. (2022). Exploring contributors, collaborations, and research topics in educational technology: A joint analysis of mainstream conferences. *Education and Information Technologies*, *28*(2), 1323–1358. https://doi.org/10.1007/s10639-022-11209-y

Chong, C., Sheikh, U. U., Samah, N. A., & Sha'ameri, A. Z. (2020). Analysis on reflective writing using natural language processing and sentiment analysis. *IOP Conference Series: Materials Science and Engineering*, *884*(1), 012069. https://doi.org/10.1088/1757-899x/884/1/012069

Contreras, J., Hilles, S. M. S., & Abubakar, Z. B. (2019). Automated Essay Scoring using Ontology Generator and Natural Language Processing with Question Generator based on Bloom's Taxonomy's Cognitive Level. *International Journal of Engineering and Advanced Technology*, *9*(1), 2448–2457. https://doi.org/10.35940/ijeat.a9974.109119

Crompton, H., & Burke, D. (2023). Artificial intelligence in higher education: The state of the field. *International Journal of Educational Technology in Higher Education*, *20*(1). https://doi.org/10.1186/s41239-023-00392-8

Demir, Ü. (2019). The effect of using negative knowledge based intelligent tutoring system evaluator software to the academic success in English language education. *Pedagogies: An International Journal*, *15*(4), 245–261. https://doi.org/10.1080/1554480x.2019.1706522

Dergaa, I., Chamari, K., Żmijewski, P., & Saad, H. B. (2023). From human writing to artificial intelligence generated text: examining the prospects and potential threats of ChatGPT in academic writing. *Biology of Sport*, *40*(2), 615–622. https://doi.org/10.5114/biolsport.2023.125623

Eisenstein, J. (2019). *Introduction to natural language processing*. MIT Press. 39. https://doi.org/10.58459/rptel.2023.18002

Feng, H., Sarıcaoğlu, A., & Chukharev-Hudilainen, E. (2016). Automated Error Detection for developing grammar proficiency of ESL learners. *The CALICO Journal*, *33*(1), 49–70. https://doi.org/10.1558/cj.v33i1.26507

Fryer, L. K., Ainley, M., Thompson, A., Gibson, A., & Sherlock, Z. (2017). Stimulating and sustaining interest in a language course: An experimental comparison of Chatbot and Human task partners. *Computers in Human Behavior*, *75*, 461–468. https://doi.org/10.1016/j.chb.2017.05.045

Fu, S., Gu, H., & Yang, B. (2020). The affordances of AI-enabled automatic scoring applications on learners' continuous learning intention: An empirical study in China. *British Journal of Educational Technology*, *51*(5), 1674–1692. https://doi.org/10.1111/bjet.12995

Fung, Y., Kwok, J. C., Lee, L., Chui, K. T., & U, L. H. (2020). Automatic Question Generation System for English Reading Comprehension. In *Communications in Computer and Information Science* (pp. 136–146). https://doi.org/10.1007/978-981-33-4594-2_12

Gao, Y., & Passonneau, R. J. (2021). Automated assessment of quality and coverage of ideas in students' Source-Based writing. In *Lecture Notes in Computer Science* (pp. 465–470). https://doi.org/10.1007/978-3-030-78270-2_82

Gough, D., Oliver, S., & Thomas, J. (2017). *An introduction to systematic reviews* (2nd ed.). SAGE.

Houston, A. B., & Corrado, E. M. (2023). Embracing ChatGPT: Implications of Emergent Language Models for academia and libraries. *Technical Services Quarterly*, *40*(2), 76–91. https://doi.org/10.1080/07317131.2023.2187110

Huang, X., Zou, D., Cheng, G., Chen, X., & Xie, H. (2023). Trends, Research Issues and Applications of Artificial Intelligence in Language Education. *Educational Technology & Society*, *26*(1), 112–131. https://www.jstor.org/stable/48707971

Jeon, J. H., & Lee, S. (2023). Large language models in education: A focus on the complementary

relationship between human teachers and ChatGPT. *Education and Information Technologies*, *28*(12), 15873–15892. https://doi.org/10.1007/s10639-023-11834-1

Jeon, J. H., Lee, S., & Choi, S. (2023). A systematic review of research on speech-recognition chatbots for language learning: Implications for future directions in the era of large language models. *Interactive Learning Environments*, 1–19. https://doi.org/10.1080/10494820.2023.2204343

Katsarou, E., Wild, F., Sougari, A., & Chatzipanagiotou, P. (2023). A Systematic Review of Voice-based Intelligent Virtual Agents in EFL Education. *International Journal of Emerging Technologies in Learning (Ijet)*, *18*(10), 65–85. https://doi.org/10.3991/ijet.v18i10.37723

Kuhail, M. A., Alturki, N., Alramlawi, S., & Alhejori, K. (2022). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, *28*(1), 973–1018. https://doi.org/10.1007/s10639-022-11177-3

Li, H., & Graesser, A. C. (2021). The impact of conversational agents' language on summary writing. *Journal of Research on Technology in Education*, *53*(1), 44–66. https://doi.org/10.1080/15391523.2020.1826022

Li, H., Gobert, J. D., Dickler, R., & Morad, N. (2018). Students' academic language use when constructing scientific explanations in an intelligent tutoring system. In *Lecture Notes in Computer Science* (pp. 267–281). https://doi.org/10.1007/978-3-319-93843-1_20

Liang, J., Hwang, G., Chen, M. A., & Darmawansah, D. (2021). Roles and research foci of artificial intelligence in language education: an integrated bibliographic analysis and systematic review approach. *Interactive Learning Environments*, *31*(7), 4270–4296. https://doi.org/10.1080/10494820.2021.1958348

Lim, K., Song, J., & Park, J. (2022). Neural automated writing evaluation for Korean L2 writing. *Natural Language Engineering*, *29*(5), 1341–1363. https://doi.org/10.1017/s1351324922000298

Lippert, A., Gatewood, J., Cai, Z., & Graesser, A. C. (2019). Using an Adaptive Intelligent Tutoring System to Promote Learning Affordances for Adults with Low Literacy Skills. In *Lecture Notes in Computer Science* (pp. 327–339). https://doi.org/10.1007/978-3-030-22341-0_26

Liu, M., Yi, L., Xu, W., & Liu, L. (2017). Automated essay feedback Generation and its impact on revision. *IEEE Transactions on Learning Technologies*, *10*(4), 502–513. https://doi.org/10.1109/tlt.2016.2612659

Liu, X., Faisal, M., & Alharbi, A. (2022). A decision support system for assessing the role of the 5G network and AI in situational teaching research in higher education. *Soft Computing*, *26*(20), 10741–10752. https://doi.org/10.1007/s00500-022-06957-5

Maqsood, S., Shahid, A., Afzal, M. T., Roman, M., Khan, Z., Nawaz, Z., & Aziz, M. H. (2022). Assessing English language sentences readability using machine learning models. PeerJ, 7, e818. https://doi.org/10.7717/peerj-cs.818

McNamara, D. S., Arner, T., Butterfuss, R., Ying, F., Watanabe, M., Newton, N., McCarthy, K. S., Allen, L. K., & Roscoe, R. D. (2022). ISTART: Adaptive Comprehension Strategy Training and Stealth Literacy Assessment. *International Journal of Human-Computer Interaction*, *39*(11), 2239–2252. https://doi.org/10.1080/10447318.2022.2114143

Meurers, D. (2012). Natural language processing and language learning. In C. Chapelle (Ed.). *Encyclopedia of Applied Linguistics*. Wiley Blackwell.

Miaschi, A., Brunato, D., & Dell'Orletta, F. (2021). A NLP-based stylometric approach for tracking the evolution of L1 written language competence. *Journal of Writing Research*, *13*(vol. 13 issue 1), 71–105. https://doi.org/10.17239/jowr-2021.13.01.03

Minoofam, S. a. H., Bastanfard, A., & Keyvanpour, M. R. (2022). RALF: an adaptive reinforcement

learning framework for teaching dyslexic students. *Multimedia Tools and Applications*, *81*(5), 6389–6412. https://doi.org/10.1007/s11042-021-11806-y

Miranda, J. P., (2020). Assisting vocabulary acquisition and script writing skills using a mobile-based Kapampangan intelligent tutoring system. So, H. J. et al. (Eds.) Proceedings of the 28th International Conference on Computers in Education. Asia-Pacific Society for Computers in Education.

Komeili, M., Shuster, K., & Weston, J. (2022). Internet-augmented dialogue generation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics, 1, 8460-8478*. https://doi.org/10.48550/arXiv.2107.07566

Murcia, D., Jaime-Osorio, M. F., Jaramillo-Calderon, L. F., & Hoyos-Pipicano, Y. A. (2025). GenAI in Language Teaching, Learning, and Assessment: Stakeholders Insights from Two Undergraduate Language Programs. *Lenguaje*, 53(1S), e20314387. https://doi.org/10.25100/lenguaje.v53i1S.14387

Nehyba, J., & Štefánik, M. (2022). Applications of deep language models for reflective writings. *Education and Information Technologies*, *28*(3), 2961–2999. https://doi.org/10.1007/s10639-022-11254-7

New London Group. (1996). A pedagogy of multiliteracies: Designing social futures. *Harvard Educational Review, 66*, 60-92.

Nguyen, L., Yuan, Z., & Seed, G. (2022). Building Educational Technologies for Code-Switching: Current Practices, Difficulties and Future Directions. *Languages, 7*(3). https://doi.org/10.3390/languages7030220

Núñez, J. A., & Da Cunha, I. (2022). El impacto del uso de herramientas automáticas de ayuda a la redacción en el proceso de escritura de estudiantes universitarios. *Circulo De Linguistica Aplicada a La Comunicacion*, *89*, 131–144. https://doi.org/10.5209/clac.73906

Olave, G., Rojas, I., & Cisneros, M. (2013). Deserción universitaria y alfabetización académica. *Educación y Educadores, 16*(3), 455-471.

Öncel, P., Flynn, L. E., Sonia, A. N., Barker, K. E., Lindsay, G. C., McClure, C. M., McNamara, D. S., & Allen, L. K. (2021). Automatic Student Writing Evaluation: Investigating the Impact of Individual Differences on Source-Based Writing. *LAK21: 11th International Learning Analytics and Knowledge Conference*. https://doi.org/10.1145/3448139.3448207

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ (Clinical Research Ed.),* 372, n71. https://doi.org/10.1136/bmj.n71

Pegrum, M. (2019). *Mobile lenses on learning: Languages and literacies on the move*. Springer.

Pengel, N., Martin, A., Meissner, R., Arndt, T., Neumann, A. T., De Lange, P. J., & Wollersheim, H. (2021). TecCoBot: Technology-aided support for self-regulated learning. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2111.11881

Pokrivčáková, S. (2019). Preparing teachers for the application of AI-powered technologies in foreign language education. *Journal of Language and Cultural Education*, *7*(3), 135–153. https://doi.org/10.2478/jolace-2019-0025

Popenici, S. A. D., & Kerr, S. (2017). Exploring the impact of artificial intelligence on teaching and learning in higher education. *Research and Practice in Technology Enhanced Learning*. https://doi.org/10.1186/s41039-017-0062-8

Rapp, C., & Kauf, P. (2018). Scaling Academic Writing instruction: Evaluation of a scaffolding tool (Thesis writer). *International Journal of Artificial Intelligence in Education*, *28*(4), 590–615. https://doi.org/10.1007/s40593-017-0162-z

Reyes, R. V., Garza, D., Garrido, L., De La Cueva, V., & Ramírez, J. (2019). Methodology for the implementation of virtual assistants for education using Google Dialogflow. In *Lecture Notes in Computer Science* (pp. 440–451). https://doi.org/10.1007/978-3-030-33749-0_35

Rus, V., & Ştefănescu, D. (2016). Non-intrusive assessment of learners' prior knowledge in dialogue-based intelligent tutoring systems. *Smart Learning Environments*, *3*(1). https://doi.org/10.1186/s40561-016-0025-3

Salas-Pilco, S. Z., & Yang, Y. (2022). Artificial intelligence applications in Latin American higher education: a systematic review. *International Journal of Educational Technology in Higher Education*, *19*(1). https://doi.org/10.1186/s41239-022-00326-w

Shehab, A., Elhoseny, M., & Hassanien, A. E. (2016). A hybrid scheme for Automated Essay Grading based on LVQ and NLP techniques. *2016 12th International Computer Engineering Conference (ICENCO)*. https://doi.org/10.1109/icenco.2016.7856447

Steuer, T., Filighera, A., Tregel, T., & Miede, A. (2022). Educational Automatic Question Generation Improves reading comprehension in non-native speakers: a Learner-Centric Case Study. *Frontiers in Artificial Intelligence*, *5*. https://doi.org/10.3389/frai.2022.900304

Strobl, C., Ailhaud, É., Benetos, K., Devitt, A., Kruse, O., Proske, A., & Rapp, C. (2019). Digital support for academic writing: A review of technologies and pedagogies. *Computers & Education*, *131*, 33–48. https://doi.org/10.1016/j.compedu.2018.12.005

Taele, P., Koh, J. I., & Hammond, T. (2020). Kanji Workbook: A Writing-Based Intelligent Tutoring System for Learning Proper Japanese Kanji Writing Technique with Instructor-Emulated Assessment. *Proceedings of the . . . AAAI Conference on Artificial Intelligence*, *34*(08), 13382–13389. https://doi.org/10.1609/aaai.v34i08.7053

Tyen, G., Brenchley, M., Caines, A., & Buttery, P. (2022). Towards an open-domain chatbot for language practice. *Proceedings of the 17th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2022)*. https://doi.org/10.18653/v1/2022.bea-1.28

UNESCO. (2023). Guidance for generative AI in education and research. UNESCO. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems.

Venegas, R. (2021). Aplicaciones de inteligencia artificial para la clasificación automatizada de propósitos comunicativos en informes de ingeniería. *Revista Signos*, *54*(107), 942–970. https://doi.org/10.4067/s0718-09342021000300942

Vitartas, P., Midford, S., & Kanjere, A. (2019). Supporting student writing with an intelligent tutoring system for assignment checking. *ASCILITE Publications*, 588–592. https://doi.org/10.14742/apubs.2019.336

Wambsganss, T., Janson, A., & Leimeister, J. M. (2022). Enhancing argumentative writing with automated feedback and social comparison nudging. *Computers & Education*, *191*, 104644. https://doi.org/10.1016/j.compedu.2022.104644

Whitelock, D., & Bektik, D. (2018). Progress and challenges for Automated Scoring and feedback Systems for Large-Scale Assessments. In *Springer International Handbooks of Education* (pp. 617–634). https://doi.org/10.1007/978-3-319-71054-9_39

Yan, D. (2023). Impact of ChatGPT on learners in a L2 writing practicum: An exploratory investigation.

Education and Information Technologies, *28*(11), 13943–13967. https://doi.org/10.1007/s10639-023-11742-4

Yang, S. and Stansfield, K. (2022) AI chatbot for Educational Service Improvement in the post-pandemic ERA: A case study prototype for supporting Digital Reading List, *13th International Conference on E-Education, E-Business, E-Management, and E-Learning (IC4E)*. https://doi.org/10.1145/3514262.3514289

Zhang, K., & Aslan, A. B. (2021). AI technologies for education: Recent research & future directions. Computers and Education: Artificial intelligence, 2, 100025

Zhang, R., Zou, D., & Cheng, G. (2023). A review of chatbot-assisted learning: pedagogical approaches, implementations, factors leading to effectiveness, theories, and future directions. *Interactive Learning Environments*, 1–29. https://doi.org/10.1080/10494820.2023.2202704

Zhao, X., Wu, W., Xu, C., Tao, C., Zhao, D., & Yan, R. (2020). Knowledge-grounded dialogue generation with pre-trained language models. *In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP),* 3377–3390. https://doi.org/10.18653/v1/2020.emnlp-main.272

Ziegler, N., Meurers, D., Rebuschat, P., Ruíz, S., Moreno-Vega, J. L., Chinkina, M., Li, W., & Grey, S. (2017). Interdisciplinary research at the intersection of CALL, NLP, and SLA: Methodological Implications from an input Enhancement Project. *Language Learning*, *67*(S1), 209–231. https://doi.org/10.1111/lang.12227