AI-Powered Platforms for Personalized Language Learning

Olha Dolhusheva[1*], Olesia Samokhval[2], Uliana Shostak[3], Viktoriia Perevozniuk[4], Hanna Perederii[5]

[1]Department of Germanic Languages, World Literature and Teaching Methodology, School of Ukrainian Philology, Foreign Languages and Social Communications, Volodymyr Vynnychenko Central Ukrainian State University, Kropyvnytskyi, Ukraine, dolgusheva@ukr.net, https://orcid.org/0000-0001-9558-5846

[2]Department of Foreign Languages and Translation, Faculty of Finance and Accounting, Vinnytsia Institute of Trade and Economics of State University of Trade and Economics, Vinnytsia, Ukraine, samohvallesya@ukr.net, https://orcid.org/0000-0001-7958-945X

[3]Department of Foreign Languages and Translation, Faculty of Finance and Accounting, Vinnytsia Institute of Trade and Economics of State University of Trade and Economics, Vinnytsia, Ukraine, u.shostak@vtei.edu.ua, https://orcid.org/0000-0002-3258-2098

[4]Department of Linguodidactics and Journalism, Faculty of Law, Humanity and Social Sciences, Kremenchuk Mykhailo Ostrohradskyi National University, Kremenchuk, Ukraine, viktoriaperevozniuk@gmail.com, https://orcid.org/0000-0001-9225-3459

[5]Department of English Philology and Translation, Faculty of Philology, Petro Mohyla Black Sea National University, Mykolaiv, Ukraine, perederiy.g@chmnu.edu.ua, https://orcid.org/0009-0005-1392-1241

*Corresponding author

HOW TO CITE:

Hafı Olha Dolhusheva, Olesia Samokhval, Uliana Shostak, Viktoriia Perevozniuk, Hanna Perederiiz Bek, İbrahim Çankaya, Hakan Gülveren (2026). AI-Powered Platforms for Personalized Language Learning.

International Journal of Special Education, 41(7s), 745-761

COPYRIGHT STATEMENT:

Copyright: © 2026 Authors.

Open access publication underthe terms and conditions of the Creative CommonsAttribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

ABSTRACT:

The article is devoted to the analysis of modern platforms for learning foreign languages based on artificial intelligence technologies, with an emphasis on personalizing the learning process and forming individual learning trajectories. The main technological approaches and their corresponding educational platforms are considered, in particular adaptive learning algorithms, speech recognition systems, generative artificial intelligence models that ensure the adaptation of content to the needs of users. Thus, different types of platforms perform different functions within a single AI ecosystem of language learning. The empirical basis of the study was the data of Duolingo Language Reports (2020–2025), which allow you to trace global language preferences in 193 countries of the world and evaluate them in the context of digital learning. English as the first foreign language shows a stable dominant, while the choice of the second language varies significantly, which indicates a significant impact of economic, cultural and geopolitical aspects. On the basis of a two-stage cluster analysis, three main groups of countries with similar linguistic profiles are obtained: Anglo-Spanish, Anglo-globalization and Romano-Francophone cluster. A conceptual model of language clusters in an AI-adaptive learning environment, which is able to interpret It is shown that modern AI platforms are evolving from standard digital courses to intelligent ecosystems capable of predicting users' educational needs, forming personalized learning routes,

and generating content in real time. The risks of implementing such systems are also considered, in particular, the issues of algorithmic bias, data privacy, and reducing the role of live communication.

Keywords: artificial intelligence, foreign language learning, personalized learning, adaptive educational platforms, generative AI

1. Introduction

In today's world, where globalization and international communication are the norm in everyday life, foreign language proficiency has become not an additional useful bonus, but a requirement in careers, education, travel or cultural exchange. Traditional methods of language learning, such as textbooks, group classroom lessons, or even one-on-one lessons, often face significant limitations. They rarely take into account the characteristics of each student (level of initial knowledge, speed of assimilation, learning style, motivation, or even emotional state at the time of the lesson). As a result, most people lose interest, plateau in progress, or leave learning entirely. This is where artificial intelligence (AI) plays an important role, radically redefining the way we learn a language, offering personalized, adaptive, and engaging experiences that work like the best personal teacher available 24/7.

Personalized learning with AI can change the content of the lesson topic in real time according to the needs of a particular user. Algorithms analyze the student's answers in real time, identify weaknesses that additional exercises direct to, and vice versa, speed up the passage of topics that have already been mastered. Here, obviously, we are not just talking about comfort, but about a significant increase in efficiency. According to a number of modern studies, these systems significantly reduce the time it takes to reach a certain level of language proficiency. Also, last but not least, they motivate users through gamification, instant feedback, and simulated live conversation.

The market for digital language learning tools is showing impressive growth. This highlights the true demand of society for these technologies. According to the research report and forecast of Astute Analytica (2025), in 2024 the market value was 26.54 billion USD. USD, while it will grow to 116.88 billion USD. with a CAGR of 17.9% by 2033. The share of AI-based technologies was about 33.5% of the total market, indicating its leading role in the transformation of the industry. It is worth noting that the most dynamic development is observed in the segment of mobile applications and adaptive learning platforms, where Duolingo has successfully proven itself. In 2025, the company reported 50 million users of its services per day and 135 million users per month. Also, due to the integration of generative AI (including GPT-4 to Duolingo Max), the number of users increased by 51% compared to last year, and the projected revenue exceeded 1 billion. USD. (Reuters, 2025; Duolingo Investors, 2025; Chief AI Officer, 2025).

It should be noted that AI does not just replace sets of exercises, it forms completely new ways of interacting. Modern platforms use speech recognition for pronunciation correction (as in ELSA Speak), generative models for creating personalized dialogues and role-playing games (Duolingo Max, Speak, Praktika), and complex adaptive learning algorithms that determine the likelihood of student success and change the learning trajectory (Vo, 2025; Torres & Kahveci, 2025). Such tools are extremely useful for learning rare languages, for people with limited access to native speakers, or for those who learn on their own. In today's conditions in Ukraine, where access to quality language education (especially in rural areas) is limited due to hostilities, AI platforms can become a real bridge to global opportunities.

However, the use of AI in language learning is not only admirable, but also a number of questions. Do these technologies really give a stable result in the long term? To what extent do they reflect the cultural and psychological aspects of learning? Do they contribute to excessive dependence on technology, as a result of which the role of the living teacher is reduced? These questions remain at the center of modern scientific discussions.

Accordingly, the purpose of this article is a comprehensive analysis of AI-powered platforms for personalized language learning, with a special emphasis on technological aspects (adaptive algorithms, language recognition, generative AI). This study examines the evolution of these tools, their effectiveness based on empirical data, the main challenges and prospects for development.

2. Literature Review

Over the past four years, scientific research has been actively conducted in the context of the use of artificial intelligence in the study of foreign languages. If earlier researchers talked mainly about simple adaptive programs, then after the appearance of ChatGPT in 2022, the emphasis shifted to truly personalized systems. They are able not just to check answers, but to imitate live communication, understand the emotional state of the student, and build an individual learning path (Vo, 2025; Liu & Zhao, 2025).

One of the most significant studies in this area is the meta-analysis by Torres and Kahveci (2025). The authors carefully analyzed 46 empirical papers and 117 effect-sizes. The results were encouraging: artificial intelligence shows a medium-large positive effect on language skills in general (Hedges' g = 0.74). It has been proven that AI helps develop vocabulary and reading the most, and speaking and listening are slightly weaker. It is important to note that the best results are obtained when AI is used not instead of the teacher, but together with him. The authors warn that the technology is good for increasing knowledge and motivation in the short term, but so far does not sufficiently develop students' independence and deep self-regulation (Torres & Kahveci, 2025).

No less valuable is the study by Liu and Zhao (2025). The authors reviewed 65 studies of informal language learning using AI. The study found that students who regularly communicate with AI chats and avatars not only improve speech, but are also much less afraid of mistakes, become more confident, and learn to better control their own learning process. Most of these studies were conducted in East Asian countries, so it is important to develop similar work in other regions.

A systematic review by Du and Daniel (2024) focuses specifically on AI chatbots for speaking practice. The conclusion is clear: students who regularly talk to chatbots experience less anxiety, are more willing to speak, and achieve better results. Many of them note that AI creates a safe environment where you can make mistakes without shame. This is what is often lacking in traditional lessons.

In the context of platform-specific analysis, the situation is also positive. After the introduction of generative AI, Duolingo Max users began to feel significantly more confident in speaking practices (Kittredge et al., 2025). ELSA Speak effectively helps to improve pronunciation and intonation (Abdelrady et al., 2026; Sa'diyah, 2025). Research by Chen (2025) and Nair (2025) confirms that personalized platforms powered by generative AI increase not only academic performance but also overall learning satisfaction.

At the same time, researchers do not idealize the technology. Many researchers (Vo, 2025; Liu & Zhao, 2025; Qiao et al., 2025) draw attention to the current serious challenges: possible bias of algorithms for different accents; data privacy issues; the risk that students will get too used to the machine and lose motivation to communicate with real people.

Thus, the modern scientific literature is generally optimistic about the potential of artificial intelligence in language education, but emphasizes that the best result is achieved when

AI is intelligently combined with human teaching, and not completely replaces it.

3. Materials and Methods

As part of the study, the mixed-method was used, which combines quantitative analysis of large amounts of empirical data and typological analysis of AI platforms for language learning, with a qualitative interpretation of current trends in the development of AI-oriented language education. This approach allows not only to identify statistical patterns, but also to explain them in the context of the transformation of digital educational platforms.

The empirical basis is the open data of Duolingo Language Reports for 2020-2025, which covers 193 countries around the world. This data is considered as a reflection of user behavior within modern AI-oriented language platforms that form personalized learning paths. Quantitative analysis included methods of descriptive statistics, frequency analysis and comparison of language distributions by year. To assess the relationship between the first and second languages, crosstabulation tables and contingency coefficient were used, which made it possible to check the statistical dependence between language choices. A key analytical tool was TwoStep Cluster Analysis, which is optimal for working with large samples of categorical data. Its advantage is the automatic determination of the number of clusters and the formation of internally homogeneous groups of countries according to the similarity of language profiles.

An important methodological component of the study is the developed conceptual model of language clusters in the AI-adaptive learning environment based on the Duolingo digital language platform. Within this model, language preferences are interpreted not as isolated choices, but as elements of a structured space of educational trajectories.

Thus, the Duolingo dataset is not only used to describe language trends but also as an empirical basis for building predictive AI models of personalized learning.

In addition, the study emphasizes that similar cluster and forecast models can be adapted to other educational tasks – from assessing the level of knowledge to predicting educational motivation, which confirms the universality of the proposed conceptual framework.

4. Results

AI-Powered Language Learning Platforms as a New Educational Ecosystem

Today's digital language learning market is demonstrating a fundamental transformation from classic mobile applications to complex intelligent ecosystems in which learning is implemented through a combination of adaptive algorithms, generative artificial intelligence, language recognition, and predictive analytics of user behavior. In this context, artificial intelligence is no longer just a tool for automating exercises and is gradually turning into a cognitive partner that simulates learning situations, adjusts learning trajectories, and provides personalized learning experiences in real time (Vo, 2025; Qiao et al., 2025).

The most ambitious example of such a transformation is the Duolingo platform, which has evolved from a gamified application to a full-fledged AI-driven educational system. Its modern version of Duolingo Max integrates large language models (GPT-4) that provide role-based dialogues, error explanation, and contextual learning functions. A study by Kittredge et al. (2025) showed that the use of generative AI in Duolingo increases students' self-assessment of language competence and students' level of confidence in communication. Additionally, empirical results confirm that AI functions (Explain My Answer, Roleplay) significantly increase user learning engagement and motivation, especially in the short term. It is Duolingo that is currently considered as the largest adaptive learning model in the global EdTech sector (Duolingo Investors, 2025; Reuters, 2025).

The second key segment is platforms focused on the development of pronunciation and phonetic

accuracy, with ELSA Speak (English Language Speech Assistant) dominating. This system uses speech recognition and phoneme-level analysis technologies to accurately identify errors in pronunciation, intonation, and speech rhythm. As noted by Abdelrady et al. (2026), ELSA Speak demonstrates a statistically significant improvement in oral competence among EFL students, as it provides instant and detailed feedback. Other studies also confirm that the system allows for autonomous correction of articulatory errors, which is a critical factor in the formation of a correct phonetic base (Sa'diyah et al., 2025).

A separate category is made up of conversational AI platforms of the new generation, such as Speak, Praktika, TalkPal and other systems based on generative models. Their main innovation is to create a fully interactive language environment where the user can conduct a dialogue with AI agents who adapt the speech style, vocabulary complexity, and pace of the conversation. According to Du and Daniel (2024), such systems significantly reduce the level of speech anxiety and increase readiness for spontaneous speech, as the user is in a psychologically safe environment without social pressure. Modern reviews also confirm that AI chatbots are one of the most effective tools for developing speaking fluency in informal learning.

In addition, an important trend is the emergence of professionally oriented AI training systems that adapt language content to a specific field of user activity. For example, modern AI platforms can generate business negotiation scenarios for business users, medical dialogues for medical students, or technical simulations for IT specialists. Studies by Chen (2025) and Nair (2025) demonstrate that such personalization not only improves academic outcomes, but also significantly increases the level of satisfaction with the learning process, as knowledge acquires practical relevance.

Summing up, modern AI platforms can be classified according to four dominant learning models. The first is adaptive learning systems, presented by Duolingo, where learning is optimized through algorithmic modeling of progress. The second is speech intelligence systems, such as ELSA Speak, specializing in phonetic accuracy. The third is conversational AI systems, including Speak and Praktika, which simulate real language interactions. The fourth is context-aware professional learning systems, which are being formed as a new direction of EdTech and focus on the professional relevance of content.

Table 1 summarizes the functional features of the leading AI platforms.

Table 1. Comparative analysis of AI-based language learning platforms

Platform	Core AI technology	Key functionality	Type of personalization	Main Focus of Training
Duolingo (Duolingo Max)	Large language models based on the Generative Pre-trained Transformer architecture, a recommendation system based on machine learning	Adaptive exercises, role-playing games with AI, explanation of mistakes	Adaptive and Predictive Learning	Vocabulary, reading, speaking practice
ELSA Speak	AI-based speech recognition system	Pronunciation correction, phonetic feedback	Micro-level pronunciation personalization	Pronunciation, accuracy of speech
Speak	Generative AI engine for dialogues	Real-time simulation of real conversations	Contextual adaptation of dialogues	Fluency, confidence in conversation

Praktika	Learning with AI avatars	Role-playing scenarios with virtual interlocutors	Scenario-oriented personalization	Speaking practice in real situations
Babbel Speak	Natural Language Processing + Adaptive Training Courses	Structured adaptive lessons	Curriculum adaptation	Grammar, structured learning

Source: data synthesized from Duolingo Investors (2025), Reuters (2025), Kittredge et al. (2025), Abdelrady et al. (2026), Du and Daniel (2024), Qiao et al. (2025), Chen (2025), Nair (2025), Vo (2025), and official platform documentation (Duolingo, ELSA Speak, Speak, Praktika, Babbel Speak)

Thus, the modern AI language learning ecosystem ceases to be a set of separate applications and forms a holistic cognitive infrastructure in which different platforms perform complementary functions in the process of user language development. It is important that these systems not only adapt educational content, but also begin to predict educational needs, forming the basis for the transition from reactive to proactive learning.

Statistical Modeling of Language Choices in the Context of AI Education Platforms

One of the most valuable and accessible sources for understanding real trends in foreign language learning is the statistics of the Duolingo platform (n.d.). They are based on data from millions of active users of the platform around the world and allow you to track how people's language preferences have changed over the past six years. This period covers the COVID-19 pandemic, geopolitical upheaval, and the rapid development of artificial intelligence in education.

Based on the analysis of open statistics from Duolingo Language Reports for 2020–2025, our study assessed the dynamics of changes in users' language preferences in different countries and regions of the world. Particular attention is paid to transformations in the rankings of the most popular languages to learn, changes in regional educational priorities, as well as the impact of global socio-economic, cultural, and geopolitical processes on the choice of a foreign language. This approach allows not only to record the current state of digital language learning, but also to track long-term trends, identify languages that show a steady increase or, conversely, a decrease in interest, and understand how personalized AI platforms can respond to these changes.

The study focuses on two key indicators:

– pop1 – most popular language studied in the country. Thus, pop1 shows the real educational priority of the country's population at a certain stage;

– pop2 – second most popular language studied in the country. Comparing pop1 and pop2 allows you to better understand the structure of language interests in a country, identify dominant trends, and assess how concentrated or diverse the demand for learning foreign languages is.

Note that the study contains data from 193 countries of the world. Figure 1 shows summary data on the pop1 indicator for the three most popular languages.

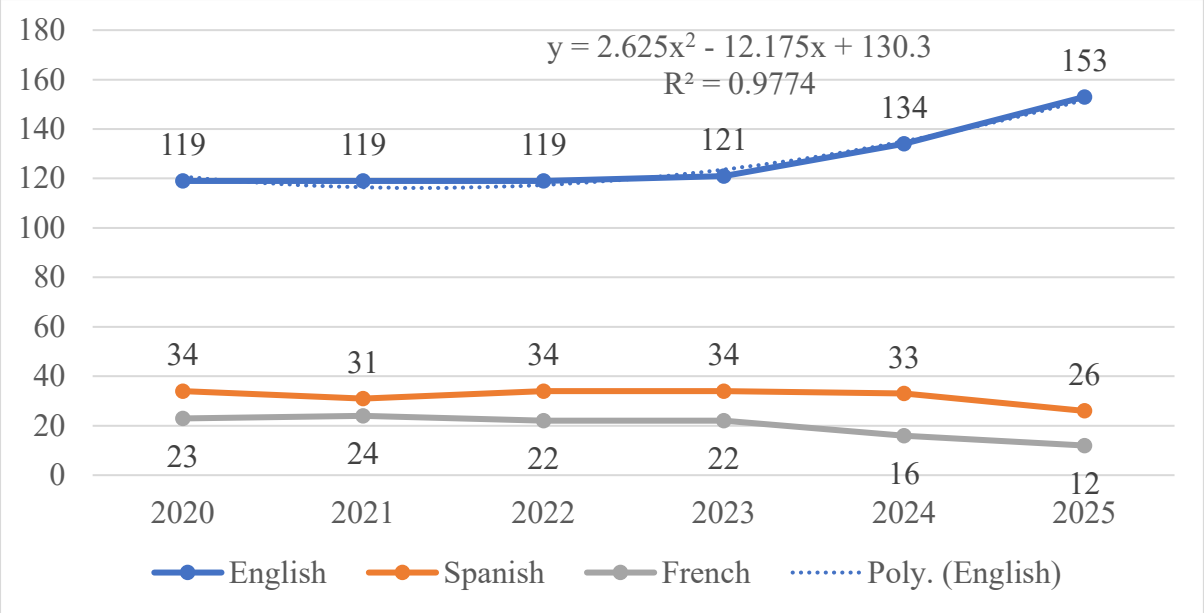

Figure 1. Number of countries where English, Spanish, and French are the most popular languages among Duolingo users (pop1 indicator), 2020–2025
Source: compiled by the authors based on open data from Duolingo (n.d.)

The data presented in Figure 1 vividly illustrate how global language priorities have changed significantly over the past six years. English continues to dominate confidently. The polynomial trend line for the English language is characterized by a high level of reliability of approximation ($R^2 = 0.9774$), which indicates a steady upward dynamics and an acceleration of the growth in the popularity of the English language in the study period.

Instead, Spanish and French are gradually losing their positions. Spanish, which has long been a fairly strong "second number", lost its primacy in 7-8 countries in 2025 compared to previous years. French shows an even more noticeable drop - from 23-24 countries in 2020-2021 to only 12 countries in 2025.

Why did such changes occur? This is not an accident, but a reflection of profound transformations in the modern world. English has finally established itself as the main "language of opportunities" – it is studied for careers, higher education, remote work, and international communication. The active introduction of artificial intelligence in Duolingo Max also played a significant role, which made the process of learning English much more efficient and enjoyable.

Such changes indicate that the traditional appeal of Romance languages is gradually weakening. Their position as first choice languages has noticeably weakened during 2020–2025.

For a deeper understanding of global language preferences, we also analyzed the pop2 indicator (languages that are the second most popular to learn in different countries of the world). For the analysis of pop2, we considered languages that were at least once included in the top 2 in 10 or more countries.

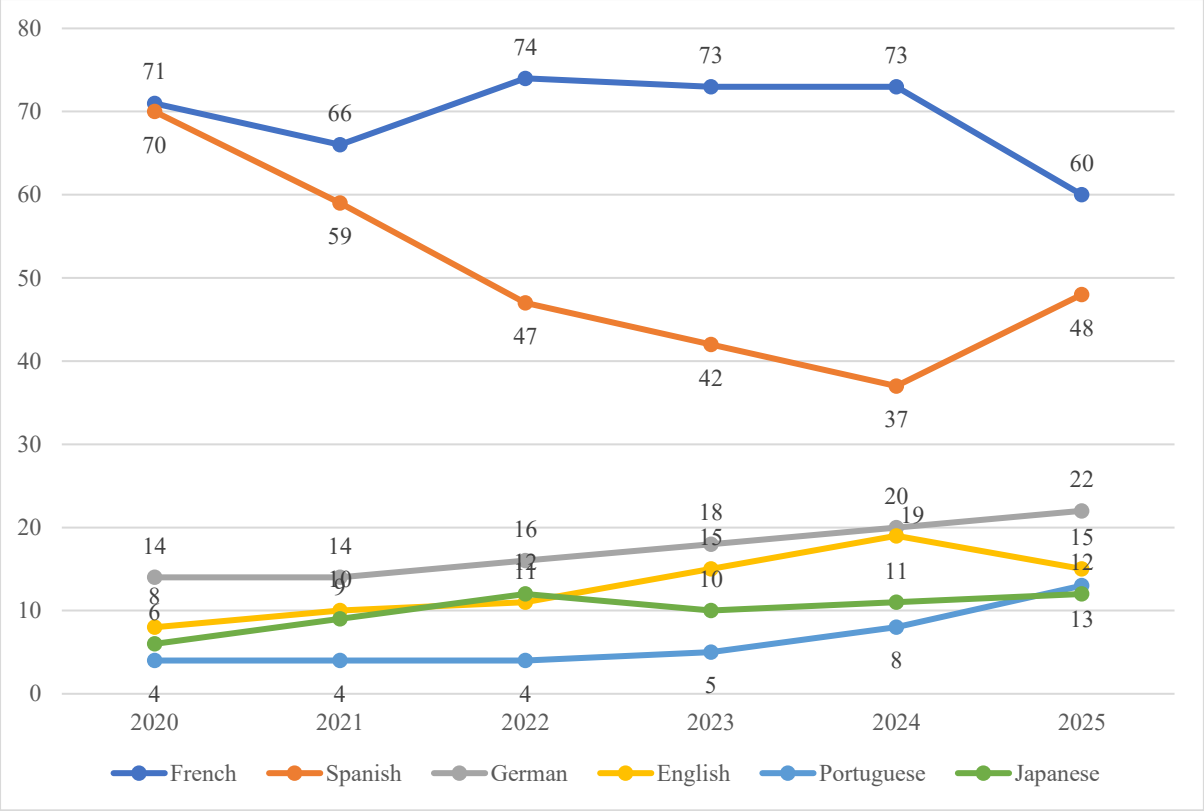

Figure 2. Number of countries in which these languages are the second most popular to learn (pop2) among Duolingo users, 2020–2025
Source: compiled by the authors based on open data from Duolingo (n.d.)

The data obtained show an interesting and rather ambiguous picture of the second choice of languages in the world.

In contrast to the absolute dominance of English in the first choice (pop1), the situation with the second place looks much more diverse. French and Spanish remain the main "second" languages to learn throughout the period. French is especially strong in second place – in 2022-2024 it was chosen as pop2 in 73-74 countries. However, in 2025, it still lost some of its positions, dropping to 60 countries.

Spanish has also been very strong as a second language for a long time (70 countries in 2020), but since 2022 there has been a steady decline in its position. In 2024, it reached a low (37 countries) and in 2025 it recovered slightly to 48 countries.

It is interesting to watch the growth of other languages. German is showing stable and steady growth: from 14 countries in 2020 to 22 countries in 2025. English as a second language is also gaining popularity, especially after 2022, when it began to be chosen more often as an additional language. Japanese and Portuguese are also gradually strengthening their positions. Since 2022, Japanese has been steadily staying at the level of 10-12 countries, and Portuguese in 2025 reached a significant figure of 13 countries for the first time.

The results suggest that a new structure of language preferences is taking shape in the world. If English becomes an almost universal first choice, then the second place remains more open to competition. Traditional European languages (French and Spanish) still dominate, but their predominance is slowly decreasing. At the same time, there is a growing interest in German, Japanese and Portuguese, reflecting the new economic, cultural and career priorities of users.

It should be noted that during the entire period under consideration, the first foreign language for Ukraine was English, and the second was German.

To better understand the structure of language preferences in the modern world, an analysis of the crosstabulation table between the indicators pop1_2025 (the most popular language) and pop2_2025 (the second most popular language) was carried out for 193 countries (Table 1). These and subsequent calculations were performed in the IBM SPSS Statistics 27 statistical package.

Table 2. Distribution of countries by combinations of pop1 and pop2 in 2025

pop2/ pop1	English	French	German	Spanish	Total
Arabic	1	0	0	0	1
Chinese	9	0	0	0	9
English	0	5	1	9	15
French	46	0	0	14	60
German	22	0	0	0	22
Hindi	1	0	0	0	1
Irish	0	0	0	1	1
Italian	3	0	0	0	3
Japanese	10	0	0	2	12
Korean	3	0	0	0	3
Portuguese	13	0	0	0	13
Russian	4	0	0	0	4
Spanish	40	7	1	0	48
Turkish	1	0	0	0	1
Total	153	12	2	26	193

Source: compiled by the authors based on open data from Duolingo (n.d.)

Analysis of the table reveals clear and interesting patterns. The most revealing is the behavior of countries in which French ranks first (pop1). In all 12 such countries, the second most popular language (pop2) is represented exclusively in English or Spanish. No other language appears as pop2 in these countries. This indicates that in countries with a strong French cultural and historical influence (mainly West and Central Africa, some Caribbean countries), French is studied as an important regional language, which is almost always supplemented by global English or Spanish.

A similar, albeit somewhat milder, situation is observed with regard to Spanish. In the 26 countries where Spanish is the most popular language to learn (pop1), the second language also has a rather limited range. Most often, these are French (14 countries) and English (9 countries). Only in rare cases does Japanese (2 countries) appear. That is, in this case, too, Spanish, acting as the first language, is almost always complemented by French or global English. This indicates regional specifics - in particular in Latin America and some European countries, where Spanish is strong, but needs to

be supplemented by international languages for wider opportunities.

In countries where English (153 countries) takes the first place, the second language is much more diverse. French (46 countries) and Spanish (40 countries) are most often chosen here, which confirms their status as traditional "second" languages. At the same time, interest in German (22 countries), Portuguese (13 countries) and Japanese (10 countries) is growing markedly.

A statistical assessment of the relationship shows that the choice of the first and second languages is not random. The Contingency Coefficient is 0.588 with a significance level of $p < 0.001$. This indicates a medium in strength, but statistically highly significant relationship between pop1 and pop2. In other words, the second language that people choose is systematically dependent on which language is their first.

To identify groups of countries with a similar language structure of the popularity of the first and second foreign languages in 2025, a two-stage cluster analysis was carried out according to the most popular platform Duolingo (n.d.) (IBM Corporation, 2004, 2026). The choice of this method is reasonable, since it is most suitable for working with categorical (nominal) variables, allows you to automatically determine the optimal number of clusters and form statistically homogeneous groups of observations. Two variables were used in the model: pop1_2025 (the most popular foreign language) and pop2_2025 (the second most popular foreign language). According to the results of the automatic clustering procedure, three stable clusters of countries were obtained (Figure 3).

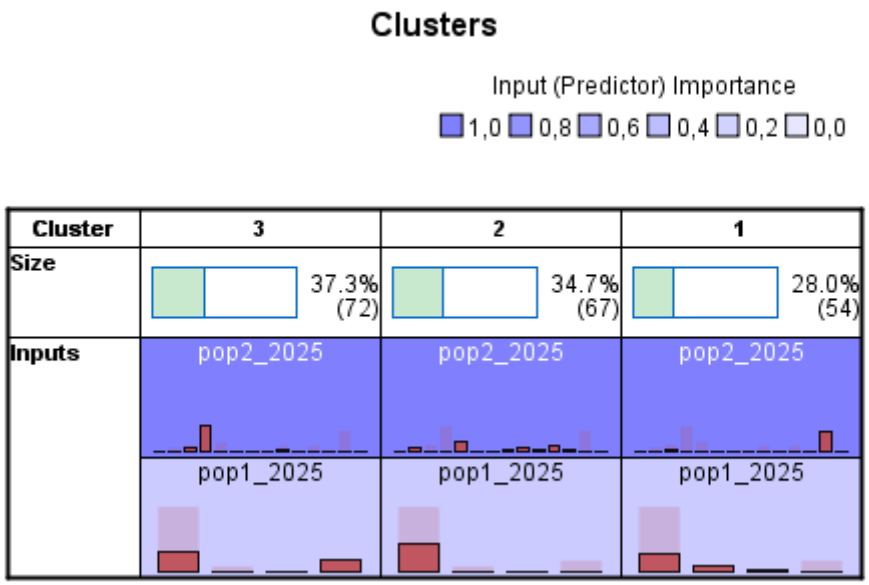

Figure 3. Visualization of the results of a two-stage cluster analysis of countries by the structure of popularity of the first (pop1_2025) and second (pop2_2025) foreign languages in 2025 according to the Duolingo platform

Source: calculated by authors based on open data from Duolingo (n.d.)

The first cluster can be described as "Anglo-Spanish". It includes 54 countries (28.0% of the sample). Its main feature is the dominance of English as the first most popular language, and it also covers all countries with the first French and German (Table 2). At the same time, Spanish almost completely prevails as a second foreign language here: 48 out of 54 countries in this

cluster have it (Table 3). Only isolated cases are presented in English as a second language. Thus, this cluster characterizes countries where English has already become the basic language of international communication, and Spanish is the next most important area of linguistic interest. This cluster includes mainly countries in Europe, Latin America, parts of Africa and the Middle East, which are characterized by a combination of English as the first foreign language and Spanish as the second (in particular, France, Spain, Brazil, Poland, the Netherlands, Portugal, Greece, Canada, Oman, Qatar).

Table 3. Distribution of the most popular foreign language (pop1_2025) by clusters of countries formed based on the results of a two-stage cluster analysis, 2025

Cluster	English	French	German	Spanish
1	40	12	2	0
2	67	0	0	0
3	46	0	0	26
Combined	153	12	2	26

Source: calculated by authors based on open data from Duolingo (n.d.)

The second cluster can be defined as "Anglo-globalization". These are 67 countries, or 34.7% of the population, which makes it the second largest in number. All countries of this cluster, without exception, have English as the most popular foreign language (Table 2). Instead, the structure of the second language is the most diverse: German, Portuguese, Japanese, Chinese, Russian, Italian, Korean and other languages are represented here (Table 3). The largest shares are German, Portuguese and Japanese. This indicates that English is already unquestionably dominant in these countries, and the second foreign language is chosen depending on the economic, regional or cultural ties of a particular state. This cluster includes the countries of Central and Eastern Europe, Asia and partly South America, where English confidently dominates as the first foreign language, and the second language is diverse and depends on regional ties (for example, Germany, Italy, Ukraine, Czech Republic, Japan, South Korea, China, Kazakhstan, Argentina, Peru).

Table 4. Distribution of the second most popular foreign language (pop2_2025) by clusters of countries formed based on the results of a two-stage cluster analysis, 2025

Cluster	Arabic	Chinese	English	French	German	Hindi	Irish	Italian	Japanese	Korean	Portuguese	Russian	Spanish	Turkish
1	0	0	6	0	0	0	0	0	0	0	0	0	48	0
2	1	9	0	0	22	1	0	3	10	3	13	4	0	1
3	0	0	9	60	0	0	1	0	2	0	0	0	0	0
Combined	1	9	15	60	22	1	1	3	12	3	13	4	48	1

Source: calculated by authors based on open data from Duolingo (n.d.)

The third cluster can logically be called "Romano-Francophone". It includes 72 countries, or 37.3% of the sample, which makes it the largest cluster. Within this segment, the first foreign language is represented mainly in English, but it is here that all the countries where Spanish is the leader (26 cases) are concentrated (Table 2). The second foreign language has a distinct French profile: in 60 countries of the cluster, it is French that ranks second (Table 3). English and Japanese are also present to a small extent. This structure indicates the special role of French as a traditional second foreign language in many countries, especially where historical, educational or cultural French-speaking ties remain strong. This cluster includes numerous countries in Africa, the Caribbean, Oceania and some European countries in which French occupies a significant position as a second foreign language, and English or Spanish is the first (in particular, Algeria, Morocco, Cameroon, Senegal, Mexico, Great Britain, New Zealand, Jamaica, Trinidad and Tobago, Switzerland).

Thus, the results of the cluster analysis showed that the countries of the world in 2025 are not grouped randomly, but according to stable language models. One block focuses on the combination of English and Spanish, the second demonstrates globally diversified demand after the dominance of English, and the third combines English-speaking or Spanish-speaking leadership with a strong position of French as a second foreign language. This allows us to talk about the existence of several geolinguistic spaces of the modern world, formed under the influence of history, economics and international communication.

In the context of the results obtained, it is advisable to propose ***a conceptual model of language clusters*** in an AI-adaptive learning environment, which allows interpreting global language preferences not as a set of isolated choices, but as a structured system of interconnected language trajectories. This model is based on the assumption that the choice of foreign languages by users of digital platforms (in particular Duolingo) is not formed by chance, but in accordance with stable clusters, which are determined by the combination of the first (pop1) and second (pop2) languages. Each of which reflects different scenarios of language development – from basic global integration to regionally specific cultural and economic orientations. In the context of personalized language learning, these clusters can be used as a basis for building AI-recommendation systems that not only adapt educational content to the user's level, but also predict the most appropriate next language to learn. Thus, the model of language clusters turns the results of statistical analysis of global language data into a practical tool for the formation of individual educational trajectories in digital AI platforms, which directly reinforces the concept of personalized and predictive learning considered in this study.

We will provide a formalized description of the proposed model based on our dataset.

The Duolingo Language Reports dataset for the period 2020-2025 is formally set as a panel dataset:

$$D = \{(c_i, t, pop1_{it}, pop2_{it}) | c_i \in C, t \in T\},$$

where is the set of countries, is the time interval of observations, and , are the categorical variables reflecting, respectively, the first and second most popular language for the country per year.$CT = \{2020, 2021, 2022, 2023, 2024, 2025\} pop1_{it}\ pop2_{it} c_i\ t$

Each observation is a discrete language profile of a country that belongs to a finite set of languages. Thus, the data can be interpreted as the implementation of a stochastic process:$(c_i, t) L$

$$(c_i, t) \rightarrow (L_1, L_2),$$

where $L_1, L_2 \sim P(L_1, L_2 | c_i, t).$

Based on this, the matrix of transitions of linguistic similarities is formalized:

$$M_t = \left\{m_{ab}^{(t)}\right\}, m_{ab}^{(t)} = P(pop2_{it} = b | pop1_{it} = a, t),$$

which reflects the conditional probability of choosing a second language under the condition of the dominant first language in a particular year $.bat$

Additionally, an aggregated feature space is used to build language clusters:

$$X_{i,t} = [1(pop1_{it} = l_a), 1(pop2_{it} = l_b)],$$

where indicator variables form a binary representation of the country's language profile.

Thus, clustering is carried out not at the level of individual countries, but at the level of distribution of language profiles in the panel data space:

$$D \to \{K_1, K_1, \dots, K_k\},$$

where each cluster is a set of countries with similar probabilistic structures for language selection over time.K_j

In the context of AI-adaptive learning, this dataset is interpreted as a training sample for the recommendation function:

$$f: (c_i, t, pop1, X) \to L_{next},$$

where these are the external characteristics of the user (educational goals, professional field, digital behavior), and this is the optimal next language to learn.XL_{next}

Thus, Duolingo dataset acts as an empirical basis for evaluating not only statistical language trends, but also for building a predictive model of language trajectories in artificial intelligence systems.

Within the framework of the proposed approach, it is important to emphasize that the model of language clusters is not an isolated tool, but can be generalized to a class of similar models that are implemented within AI-oriented educational platforms. In particular, on the basis of the same data formalization, alternative analytical models can be built that describe other aspects of language learning, such as models of transition between levels of language proficiency, models of cognitive load during the acquisition of grammatical structures, or models for predicting the user's language motivation. All these approaches can be presented in the form of a generalized functional scheme:$D = \{(c_i, t, pop1_{it}, pop2_{it})\}$

$$f_k: D \to Y_k,$$

where f_k is a separate AI model, and the corresponding target variable (for example, language level, probability of completion of the course, or optimal complexity of educational content). Thus, the model of language clusters proposed in the paper is only one of the elements of a broader ecosystem of AI-adaptive learning, in which different models can work in parallel and complement each other, providing a comprehensive personalization of the user's educational trajectory within modern digital language platforms.Y_k

5. Discussion

Implications of the obtained trends for the development of AI platforms for personalized language learning

The results of the analysis of global statistics by Duolingo (n.d.) demonstrate that the modern foreign language learning market is becoming more and more structured. As demonstrated above, English has finally established itself as a universal basic language of international mobility, education and digital communication. On the other hand, the choice of a second foreign language turns out to be much more variable and depends on regional, economic, cultural and professional factors. For platforms supported by artificial intelligence, this means a transition from a universal learning model to a multi-level personalization, which takes into account not only the level of knowledge of the user, but also their language trajectory, goals, behavior, and global trends. In this context, AI platform architectures play a key role, including large language models and recommendation systems that form customized learning routes.

Transition from static to predictive personalization

Traditional digital language courses mostly offer the same sequence of topics for all users. This process is implemented through machine learning systems and recommendation systems that analyze user behavior, the level of progress, and the context of learning. If the system sees that a user from Ukraine is learning English and is interested, for example, in the IT field, it can

recommend German or Polish as the languages of labor mobility in Europe as the next stage. If the user, for example, demonstrates interest in Asian culture, the algorithm can suggest Japanese or Korean.

Thus, personalization moves from the reactive learning model to predictive learning. That is, the prediction of the future needs of the student is performed. It is this approach that is considered as one of the key areas of development of AI education (Vo, 2025).

Intelligent recommendation systems of language pairs and trajectories

The obtained statistical patterns between the indicators of pop1 and pop2 indicate that the choice of a second language is not random. This opens up the prospect of creating language recommendation systems similar to those used by Netflix or Spotify.

Such systems are based on collaborative filtering algorithms and content-oriented models, which are already widely used in digital recommendation services.

For example:

– if a user of a certain platform chooses French as the main language, the system can recommend English as a functional addition;

– if the main language is English, then depending on the profile, German (for career development), Japanese (for development in the direction of technology), Spanish (for strengthening global communication), Portuguese (for future business development in Latin America) can be recommended;

– For medical students or scientists, the system may offer academic English and German;

– for users of creative industries – Japanese or Korean.

Such recommendations increase motivation, since learning will be perceived as part of a personal development strategy, and not as an abstract course.

Generative AI as a tool for contextual learning content

The emergence of generative artificial intelligence (GPT models, multimodal LLMs, conversational AI) is radically changing the way content is created for language learning. If earlier platforms were limited to pre-prepared exercises, today the system can create in real time: personalized dialogues; role-playing games; adaptive texts for reading; vocabulary exercises according to the user's profession; explanation of grammar in the student's native language; modeling of interviews, travel, business negotiations.

For example, a user learning English for IT can receive dialogues about Scrum meetings, technical presentations, or code reviews. For a user learning German for university studies, the system can generate exercises about academic writing or communication with a professor. It is this adaptation of content that significantly increases the relevance of learning and intrinsic motivation (Kittredge et al., 2025; Chen, 2025).

Development of oral speech and pronunciation technologies

One of the most difficult aspects of learning a foreign language has traditionally been speaking. Many students have sufficient knowledge of grammar, but are afraid to speak because of shame, low self-esteem, or lack of an interlocutor. AI platforms are changing this situation through a combination of: automatic speech recognition; text-to-speech of the new generation; analysis of phonetic errors; conversational agents; emotionally neutral, safe practice environment.

Platforms like ELSA Speak, Speak, Praktika are already demonstrating high efficiency in correcting pronunciation and developing fluency. AI is able not only to point out an error, but also to explain which sound was pronounced incorrectly, how articulation works, and how to train intonation (Abdelrady et al., 2026; Sa'diyah et al., 2025). In the future, it is expected that systems will analyze not only phonetics, but also pragmatics of speech (politeness, naturalness, appropriateness of phrases, emotional tone).

Multimodal learning and digital avatars

A new generation of language platforms is on the way from simple text-based exercises to

multimodal interaction. This means using many of the following tools at the same time, namely text, voice, video, images, gestures, and virtual characters. In this case, AI avatars can act in a variety of roles and positions, including as interlocutors, teachers, customer simulators, examiners, cultural guides, and more.

This point is extremely important for introverted learners or learners who are afraid of live speaking practice. The avatar will provide a psychologically comfortable space for the learner where they can make mistakes without social pressure. Recent research shows that it is the reduction of language anxiety that is one of the strongest effects of AI chatbots in language education (Du & Daniel, 2024).

Localization and support of language communities with limited access

For Ukraine and many other countries, an important consequence of the development of AI platforms is the democratization of access to quality language education. In regions where there is a shortage of qualified teachers or there are restrictions due to war, economic instability or migration, AI systems can provide: continuous training 24/7; affordable price; personal digital tutor; learning from a smartphone; quick adaptation to an individual pace. For Ukraine under martial law, this is especially true for internally displaced persons, students, military, migrant workers, and schoolchildren in remote communities.

From assessment of knowledge to assessment of competencies

Traditional language tests measure mainly knowledge of the rules. AI allows you to move on to the assessment of already real communicative competence. For example, the system can analyze the following characteristics in the process of learning a foreign language: reaction speed in dialogue; vocabulary; naturalness of speech; grammatical accuracy in real communication; stability of long-term progress; readiness for work or academic scenarios. This means that future language certifications may become much more practical and personalized.

Main risks and limitations

Despite the significant potential, the development of AI platforms may be accompanied by the following challenges: bias of algorithms for accents; preference for English-language data in training models; privacy of voice and behavioral data; risk of excessive dependence on the app; reduced contact with real people; superficial learning through gamification without deep language practice. That is why most modern researchers support the AI + Teacher model, where artificial intelligence strengthens the teacher rather than replaces him (Torres & Kahveci, 2025; Qiao et al., 2025).

The obtained trends show that the future of language education lies not just in the digitalization of courses, but in the creation of intelligent ecosystems of personalized learning that will be able to perform the following functions: predict user needs; recommend optimal language routes; generate relevant content; provide conversational practice; assess real competencies; maintain motivation in the long term.

Thus, platforms supported by artificial intelligence become not an auxiliary tool, but a central element of a new model of learning foreign languages, where learning is adaptive, continuous and as close as possible to the real needs of a person.

6. Conclusions

The study confirms that modern language education is in a phase of structural transformation, in which artificial intelligence technologies and big data analytics play a key role. The transition from traditional educational models to adaptive digital systems is accompanied by the formation of a new learning logic focused on individual user trajectories. The study looked at the leading AI platforms for language learning (including Duolingo, ELSA Speak, Speak, Praktika and Babbel Speak), as well as their main technological mechanisms, including generative language models, speech

recognition systems, and adaptive learning algorithms.

Empirical analysis of Duolingo data has shown the stable dominance of English as a global educational standard, while the choice of a second foreign language remains more variable and depends on regional and socioeconomic factors. This confirms the existence of structured language patterns that can be formalized through a cluster approach.

The results of a two-stage cluster analysis made it possible to identify three main groups of countries with different language strategies: Anglo-Spanish, Anglo-globalization and Romano-Francophone clusters. This indicates the existence of stable geolinguistic models that are formed under the influence of historical, cultural and economic factors.

The key theoretical contribution of the study is the proposed conceptual model of language clusters in the AI-adaptive learning environment, which interprets language choices as elements of structured educational trajectories. Within this model, clustering acts not only as an analytical tool, but also as a basis for building predictive recommendation systems capable of determining the optimal next language for learning.

Thus, the model allows you to move from descriptive analysis of language data to their applied use in AI systems of personalized learning, where the user's educational trajectory is formed based on the behavioral patterns of similar groups.

The practical results of the study show that modern AI platforms, considered in the paper, already go beyond traditional learning, forming interactive environments with personalized content, adaptive routes, and predictive learning logic. At the same time, the hybrid model remains the most effective, in which artificial intelligence complements but does not replace the role of the teacher.

As a result, it can be argued that the proposed conceptual model of language clusters is not only a tool for analyzing global language trends, but also a potential basis for creating a new generation of AI education systems focused on personalization, forecasting and continuous development of language competencies. In this context, the platforms considered in the study (in particular, Duolingo, ELSA Speak, Speak, Praktika and others) act as a practical embodiment of the relevant mechanisms, namely adaptive learning, generative interaction, speech recognition and scenario personalization, which actually implement the logic of the proposed cluster model in real educational environments.

References

Abdelrady, A. H., Akram, H., Mohamed, A. T. A., Abaalkhail, G. F., Osman, E. H. E., Mohamed, G. T. A. H., & Suliman, W. M. M. (2026). Harnessing AI for pronunciation development: Investigating the impact of Elsa Speak on EFL learners' oral proficiency. *World Journal of English Language, 16*(4), 311. https://doi.org/10.5430/wjel.v16n4p311

Astute Analytica. (2025). *Digital Language Learning Market Set to Exceed US$ 116.88 Billion by 2033 as Record User Engagement Accelerates the Global Shift Toward Habitual Learning Says Astute Analytica*. GlobeNewswire. https://www.globenewswire.com/news-release/2025/12/08/3201751/0/en/Digital-Language-Learning-Market-Set-to-Exceed-US-116-88-Billion-by-2033-as-Record-User-Engagement-Accelerates-the-Global-Shift-Toward-Habitual-Learning-Says-Astute-Analytica.html

Chen, Y. (2025). Evaluation of the impact of AI-driven personalized learning platform on medical students' learning performance. *Frontiers in Medicine, 12*, 1610012. https://doi.org/10.3389/fmed.2025.1610012

Chief AI Officer. (2025). *How Duolingo's AI-first strategy drove 51% user growth and $1 billion revenue forecast*. https://chiefaiofficer.com/how-duolingos-ai-first-strategy-drove-51-user-growth-and-1-billion-revenue-forecast/

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

Duolingo Investors. (2025). *Duolingo surpasses 50 million daily active users, grows DAU 36% and revenue 41% in third quarter 2025*. http://investors.duolingo.com/news-releases/news-release-details/duolingo-surpasses-50-million-daily-active-users-grows-dau-36

Duolingo. (n.d.). *Duolingo language report [2020–2025]: Public data* [Google Sheets spreadsheet]. Google Sheets. https://docs.google.com/spreadsheets/d/1CndYC5ZovYfmPuMN9T9Jxfa4CQXOzZrfQ2kAUaWG1ZU/edit?ref=blog.duolingo.com&gid=532174835#gid=532174835

Kittredge, A. K., Hopman, E. W. M., Reuveni, B., Dionne, D., Freeman, C., & Jiang, X. (2025). Mobile Language App Learners' Self-Efficacy Increases after Using Generative Ai. *Frontiers in Education, 10*, 1499497. https://doi.org/10.3389/feduc.2025.1499497

Liu, G. L., & Zhao, X. (2025). A scoping review of AI-mediated informal language learning: Mapping out the terrain and identifying future directions. *ReCALL, 38*(1), 111–130. https://doi.org/10.1017/S0958344025100359

Nair, M. (2025). *The impact of AI-driven personalized learning platforms on student outcomes: A comprehensive review* (Master's thesis, Iowa State University). https://dr.lib.iastate.edu/server/api/core/bitstreams/ae014a30-35d7-4dc2-93e5-6ed8ff84b132/content

Qiao, S., Gu, M. M., & Lu, C. (2025). Artificial Intelligence for Language Learning: A Systematic Review of its Design, Theoretical Foundations, Implementation, and Impact. *International Journal of Applied Linguistics*. https://doi.org/10.1111/ijal.70034

Reuters. (2025). *Duolingo raises 2025 revenue forecast as AI tools boost user engagement*. https://www.reuters.com/business/duolingo-raises-2025-revenue-forecast-ai-tools-boost-user-engagement-2025-08-06/

Sa'diyah, L., Bahtiar, Y., Nur, L. C. N., Afidah, N., & Ma'arif, I. B. (2025). The effectiveness of Elsa Speak to improve speaking skill in extra class MA Al-Bairuny. *APPLICATION: Applied Science in Learning Research, 5*(1), 27–35. https://ejournal.unwaha.ac.id/index.php/application/article/download/5805/2643

Torres, P. J., & Kahveci, Y. E. (2025). Effectiveness of artificial intelligence (AI) in language teaching. *Computers and Education: Artificial Intelligence, 9*, 100522. https://doi.org/10.1016/j.caeai.2025.100522

Vo, T. K. A. (2025). Transforming language learning with AI: Adaptive systems, engagement, and global impact. *Engineering Proceedings, 107*(1), 7. https://doi.org/10.3390/engproc2025107007

IBM Corporation. (2004). *TwoStep cluster analysis* (SPSS Statistics algorithms 14.0). IBM. https://public.dhe.ibm.com/software/analytics/spss/support/Stats/Docs/Statistics/Algorithms/14.0/twostep_cluster.pdf

IBM Corporation. (2026). *TwoStep cluster node*. IBM Cloud Pak for Data as a Service. https://dataplatform.cloud.ibm.com/docs/content/wsd/nodes/twostep.html