# AIlingo: Contextual Language Learning with Artificial Intelligence in Real-World Contexts


Tianqin Lu
Department of Industrial Design, Eindhoven University of Technology
Eindhoven, The Netherlands
t.lu1@tue.nl

Sahil Bhatnagar
Department of Human-Technology Interaction, Eindhoven University of Technology
Eindhoven, The Netherlands
s.r.bhatnagar@student.tue.nl

Nick van Geenen
Department of Industrial Design, Eindhoven University of Technology
Eindhoven, The Netherlands
n.m.v.geenen@student.tue.nl

Yinying Miao
Department of Industrial Design, Eindhoven University of Technology
Eindhoven, The Netherlands
y.miao@student.tue.nl

Zitian Zhang
Department of Industrial Design, Eindhoven University of Technology
Eindhoven, The Netherlands
z.zhang6@student.tue.nl

Aaqib Saeed
Department of Industrial Design, Eindhoven University of Technology
Eindhoven, The Netherlands
a.saeed@tue.nl

Jun Hu
Department of Industrial Design, Eindhoven University of Technology
Eindhoven, The Netherlands
j.hu@tue.nl


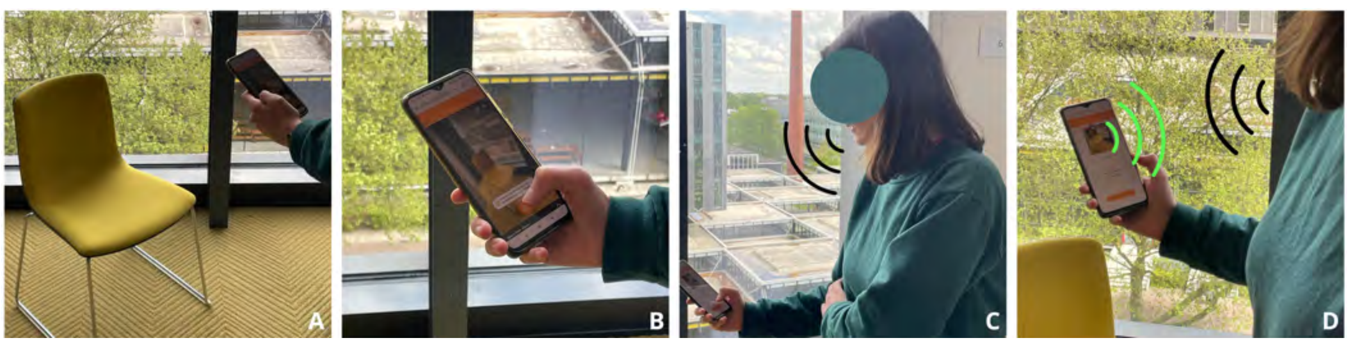

**Figure 1: A user is using AIlingo to learn within their surroundings (A). The user captures a scene using their smartphone (B), describes the scene in their own words (C), receives object captions and scene descriptions in Dutch generated by AIlingo, and checks the verbal description, repeating it (D).**


## ABSTRACT

In today's globalized world, mastering foreign languages is crucial for enhanced cross-cultural understanding and career opportunities. Mobile assisted language learning (MALL) offers a convenient and accessible platform, with recent advancements in AI providing opportunities to enhance contextualized learning experiences. This study introduces AIlingo, an AI-enabled MALL tool designed to provide immersive language learning experiences within learners' real-world contexts. AIlingo employs pre-trained AI models for object detection, image captioning, text generation, and text-to-speech, enabling learners to improve vocabulary and sentence formulation by recognizing everyday objects. The user study evaluates AIlingo through the User Experience Questionnaire (UEQ) and semi-structured interviews, providing insights for further development efforts. This study set the stage for future research exploration of integrating language learning seamlessly into daily life and advancing AI in language education.





*CHCHI 2024, November 22–25, 2024, Shenzhen, China*





ACM ISBN 978-1-4503-XXXX-X/18/06
https://doi.org/XXXXXXX.XXXXXXX


## CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing**; • **Applied computing** → **Education**.

## KEYWORDS

Mobile Assisted Language Learning, Contextual Learning, Artificial Intelligence

**ACM Reference Format:**

Tianqin Lu, Sahil Bhatnagar, Nick van Geenen, Yinying Miao, Zitian Zhang, Aaqib Saeed, and Jun Hu. 2024. AIlingo: Contextual Language Learning with Artificial Intelligence in Real-World Contexts. In *Proceedings of Chinese CHI 2024 (CHCHI 2024).* ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX


## 1 INTRODUCTION

In today's globalized world, mastering foreign languages has become increasingly important. It offers benefits like enhanced cross-cultural understanding and expanded career opportunities [7]. Immersive language learning environments enable learners to acquire language more naturally and provide insights across diverse proficiency levels, whether in-person or computer-mediated [45, 46]. By nature, such learning exposes learners to the target language in various contextual settings [36], allowing them to immerse themselves and practice the target language using tangible objects, interactions with individuals, and contextual situations [46]. However, immersion, like interacting with native speakers or traveling abroad, is often inaccessible. Recent research demonstrated the potential of augmented reality (AR) and virtual reality (VR) technologies to provide immersive language learning experiences [35, 53]. However, barriers such as limited accessibility, physiological concerns, operational complexity, and high costs hinder widespread adoption of the technology and devices [2, 39, 41].

In contrast, mobile assisted language learning (MALL) offers a convenient and accessible platform for learners to access learning material and personalize digital learning content anytime and anywhere [10, 17]. Today's mobile devices are equipped with a variety of powerful features, such as internet connectivity, cameras, and global positioning system, which enable them to gather diverse information about the surrounding physical environment [54]. These capabilities open up new opportunities for situated and contextual learning experiences. By connecting learners with their real environments, MALL facilitates the acquisition and practice of the target language using tangible objects, social interactions, and contextual situations, resulting in an immersive learning experience [46]. Moreover, recent advancements in artificial intelligence (AI), such as natural language processing (NLP) [37], have been used to enhance learning experiences across various environments and contexts. Combining MALL tools with AI technology has produced a variety of effective and motivating educational tools, such as intelligent feedback systems [47], AI-powered chatbots [8], writing performance assessment [16] and image-to-text recognition [48]. However, limited research has been conducted to explore the extent of AI-enabled mobile learning in enhancing immersion through contextually relevant learning materials in real-world environments.

Hence, in this study, we introduce AIlingo, an AI-enabled MALL tool designed to facilitate immersive language learning experiences within learners' real-world contexts (Figure 1). AIlingo focuses on teaching Dutch as a target language but is poised to accommodate other languages in the future. AI is used for four tasks: object detection, image captioning, text generation, and text-to-speech conversion. The process begins with the user capturing an image with their mobile devices, which is then fed into a Convolutional Neural Network (CNN) model, YOLOv8 [38], for object detection. Simultaneously, the same image undergoes processing by Bootstrapping Language Image Pre-training (BLIP), a transformer model, which generates text-based captions describing the visual content [26]. The captions are dynamically rewritten using Large Language Model (LLM), GPT-3.5 Turbo [32], guided by a predefined prompt. The rewritten captions are then synthesized into speech using the TTS-1 model [33], facilitating speaking practice. With these capabilities, AIlingo allows learners to expand their vocabulary and form sentences by recognizing real-world objects in their surroundings. This approach furnishes learners with a platform for independent exploration of real-life scenarios in real-world and ubiquitous learning settings, thereby fostering immersive learning experiences. The user study involved seven participants, in which we evaluated the user experience of AIlingo through the User Experience Questionnaire (UEQ) and semi-structured interviews. The results reveal positive evaluations across all dimensions of AIlingo's usability, particularly highlighting its clarity, efficiency, and dependability. Participants appreciated AIlingo's contextual learning experience but desired more adaptation and personalization features, suggesting opportunities for further development.

The contributions of this paper are (1) the design and implementation of a MALL tool, AIlingo, that leverages AI for contextual language learning. (2) An interactive prototype demonstrating AIlingo's functionalities. (3) A user study that explores user experience and gathers insights for future development.

## 2 RELATED WORK

### 2.1 Mobile Assisted Language Learning

With the continued development and widespread use of mobile devices and the Internet, MALL has emerged as a vital and timely area of research since the early 2000s [49]. MALL is defined as "the use of mobile technologies in language learning, especially in situations where device portability offers specific advantages" [20]. These advantages include but are not limited to flexibility, affordability, portability, and user-friendliness [14]. Consequently, MALL provides learners with plenty of learning opportunities in both formal and informal settings [22]. MALL is widely employed in facilitating language skills and components such as vocabulary [1], grammar [9], speaking [27], and reading [5]. Yet, recent studies indicate that MALL can serve as a medium that provides a flexible, context-aware language learning environment [17, 51]. Such an environment allows learners to expand their exposure to the target language and engage in practice activities across various physical settings and contexts. Moreover, MALL features situated, real-world, and continuous assessment by taking advantage of mobile technologies [19]. Despite these advancements, previous literature has yet to fully explore this potential of MALL. Another notable deficiency in previous MALL research is the lack of emphasis on learners' learning experiences [17]. While language proficiency has been a primary focus of MALL research [13], non-linguistic factors such as self-directedness, autonomy, and immersion have been overlooked.

Therefore, there remains a need for further exploration to better understand and tap the potential of MALL.

## 2.2 Language Learning in Real-World Contexts

The situational learning approach highlights the importance of "context" in language acquisition, as it enhances learners' engagement and effectiveness [12]. Real-world contexts offer an ideal opportunity for learners to interact with physical objects and gain real-world, immersive language learning experiences [15, 34]. This enables learners to acquire knowledge and skills within specific contexts, thereby enabling them to infer word meanings from contextual clues. It is a strategy recognized for its potential to enhance long-term retention [50]. In an early study on vocabulary learning in contexts, Ogata and Yano proposed the use of Tagged Annotations for Learning Objects (TANGO), which allows students to learn vocabulary related to surrounding objects using mobile phones and radio frequency identification technology [31]. Furthermore, a quasi-experimental study, which was conducted to enhance English learners' speaking abilities, implemented a sensor-based, AI-enhanced augmented reality mobile learning environment [29]. The results demonstrated improved speaking skills within the immersive context-aware system.

The current state of connecting learning with its context in foreign language learning calls for enhanced practicality to enable broader use and application in diverse, real-world contexts with different target languages [24]. Developing a more universal and widely applicable system may pose challenges, but addressing this challenge could yield beneficial results. In this study, we focused on developing a MALL tool, which prioritizes contextual information based on the situational learning approach and contextual information [4], such as the learner's real-world surroundings and ubiquitous learning time, to create an immersive, contextual learning experience.

## 2.3 Artificial Intelligence in Language Learning

AI-enabled learning is concerned with integrating AI technologies into learning platforms to deliver tailored content, guidance, pathways, feedback, or interfaces that cater to individual learner needs and preferences [21]. A thorough examination of AI technologies has highlighted that AI-based modelings possess the potential to augment the intelligence and functionality of real-world applications; AI-based solutions can be more extensively utilized in real-world applications for language learning [42].

Recent reviews on AI-based language learning tools have shown that the majority of these tools adopted machine learning (e.g., intelligent tutoring systems) and natural language processing, most of which focus on the cognitive aspects of language acquisition, and some consider affective or psychological aspects [28, 52]. For instance, Shazly integrated AI chatbots into the speaking practices of English learners, with findings indicating promising improvements in linguistic output gains [6]. Chen et al. used AI to develop a personalized, collaborative digital reading annotation to reduce reading anxiety in English learners [3]. In addition to the research field, software applications like Babbel [1] and Duolingo [2] utilize AI technologies to enhance language learning experiences. Babbel provides immediate feedback on learners' inputs, addressing issues like mixed tenses and verb forms [43], while Duolingo employs an AI model called "Birdbrain" to ensure exercises are at the optimal difficulty level based on learners' strengths and weaknesses [11].

However, the potential of AI in language learning extends beyond these examples. AI facilitates immersive language learning by integrating data such as images, voice, and text into learning environments, enabling context understanding and enhancing interaction during the learning process. One innovative example is the AI-based English language learning system (AIELL) developed by Jia et al. [15], which employs image recognition in mobile learning to aid lower-grade learners in improving vocabulary and grammar skills. This demonstrates AI's positive impact on learners' motivation in real and ubiquitous language learning environments.

# 3 DESIGN AND IMPLEMENTATION

## 3.1 Overview

We designed AIlingo, a MALL tool aimed at providing contextual and immersive language learning experiences. Unlike traditional MALL tools that rely on static content, AIlingo utilizes the user's smartphone camera and state-of-the-art AI models to collect contextual information from the users' real environments. Our tool aims to offer a dynamic and versatile learning platform that seamlessly integrates real-world contexts, AI technology, and personalized learning experiences.

The design concept behind AIlingo caters to learners at all proficiency levels, based on the principles of Bloom's Taxonomy [18]. It encompasses a blend of memorization, recognition, translation, and interpretation activities. Although the learning experience consists of several iterative steps and spans a longer period, AIlingo focuses on the following key steps: (1) capturing a scene from real-world contexts using the smartphone camera, (2) providing a verbal description of the captured scene generated with a vision AI model [38], (3) receiving a verbal description of the scene from AIlingo and (4) verbally repeating the description provided by AIlingo.

## 3.2 Implementation

The core tasks of AIlingo include object detection, image captioning, caption rewriting, and text-to-speech. Figure 2 illustrates the system pipeline, depicting the flow of data and four AI techniques used within AIlingo.

The system utilizes user-captured images from real-world environments as the primary input. The captured image is first processed by a CNN-based image classification model, YOLOv8. It is the latest iteration of YOLO (You Only Look Once), a highly efficient real-time object detection system [38]. This model was deployed using Supervision [40], an open-source library for computer vision tasks. YOLOv8 acts as a feature extractor, identifying and encoding the prominent visual elements within the image. Simultaneously, the same image is fed into a transformer-based model called BLIP [26]. BLIP excels at image captioning. It analyzes the visual data and generates a textual description of the image content. This study utilizes the BLIP-caption model, accessible through the open-source LAVIS library [25]. BLIP-caption effectively encodes the image and decodes it into a natural language sentence describing the scene.

[1]https://www.babbel.com/
[2]https://www.duolingo.com/

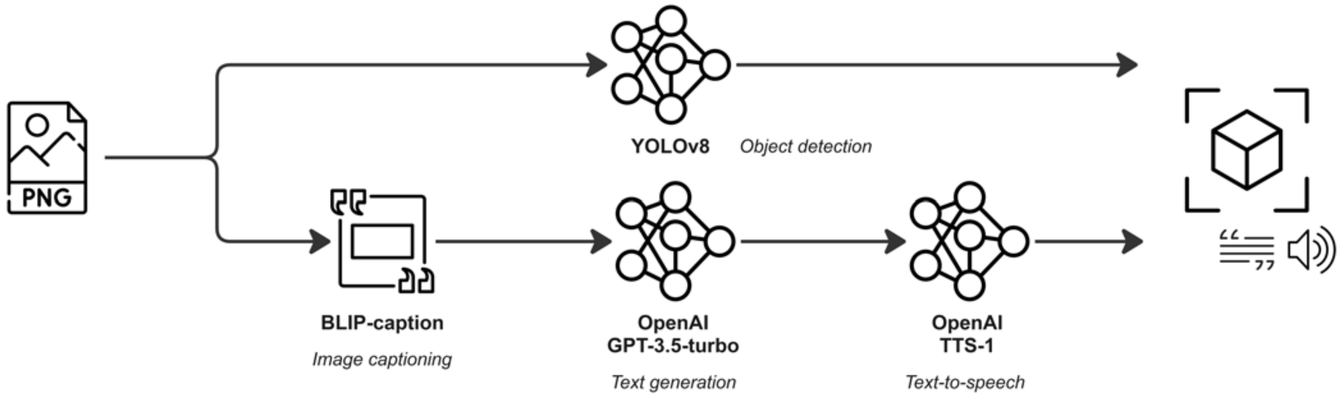


**Figure 2: The overall pipeline of AIlingo.**

While BLIP-caption provides a useful starting point, the captions it generates may not be ideally suited for language learning due to their potentially static nature. As users progress in their language proficiency, captions with increasing complexity would be beneficial. Therefore, the system utilizes OpenAI's GPT-3.5-turbo model [32] via its application programming interface (API). GPT-3.5-turbo is a powerful language model capable of text generation and manipulation, which is employed to dynamically rewrite the captions generated by BLIP-caption. This rewriting process is guided by a specific prompt that instructs GPT-3.5-turbo on how to modify the caption for optimal language learning purposes. An example prompt used in this study is: "Rewrite the following sentence in beginner level dutch: ...". Subsequently, the rewritten caption is inputted into OpenAI's TTS-1 text-to-speech (TTS) model [33] via the same API. TTS-1 converts the textual caption into an audio file. This audio file can then be played back to the user, allowing them to practice their speaking.

### 3.3 Interactive Prototype of AIlingo

An interactive prototype was developed using Figma for experimental purposes. AIlingo is intended to function with any combination of target and source languages, yet this particular prototype teaches Dutch to English speakers.

Figure 3 describes the workflow of the interaction process with AIlingo. It begins when users tap the camera icon, which navigates them to a full-screen camera page, allowing them to capture and upload photos of their immediate surroundings. After capturing an image, users are directed to the selection page, where they can define specific areas of interest within the image for AI recognition by adjusting the corners of a white box, thus filtering out irrelevant objects. Upon selecting the desired area, users simply tap the "select" button to proceed. Users are then prompted to verbally describe the scene while the AI model processes the image to generate a caption. Once the user completes their description and taps the "show results" button, they receive the AI-generated description, accompanied by audio, in the central box. Users also have the option to engage with follow-up exercises by tapping "continue exploring." Alternatively, users can return to the camera page by tapping the camera icon, enabling them to capture new contexts and further their learning.

## 4 USER STUDY

We conducted a user study to evaluate the user experience of AIlingo, which was approved by the university's Ethical Review Board.

### 4.1 Participants

Through purposive sampling, 7 participants (4 males and 3 females, aged between 24 and 26, marked as P1 - P7) who are international students from a university in the Netherlands, where English is the official language, took part in the study. All participants are in the early stages of learning Dutch (aiming to reach the Common European Framework of Reference for Languages level A1: beginners) and have prior experience using other language applications, such as Duolingo. Participation was voluntary.

### 4.2 Materials

We arranged three scenarios; each had a set of physical objects that served as the learning materials for participants, see Figure 4.

These scenarios were designed for participants to interact with the AIlingo prototype. One scenario was developed to accommodate different difficulty levels by varying the number of objects described in a single sentence (Figure 4c). Participants selected between one to three objects, and descriptions were subsequently generated based on their selection (Figure 5).

### 4.3 Measures

The User Experience Questionnaire (UEQ) was used to assess the usability of AIlingo. UEQ facilitates a rapid and direct evaluation of user experience, utilizing 7-point semantic differentials ranging from −3 (indicating complete agreement with negative conditions) to +3 (indicating complete agreement with positive opinions), with a midpoint of 0 representing neutrality [23, 44]. This questionnaire yields outcomes across six dimensions with twenty-six items. These dimensions are attractiveness, perspicuity, efficiency, dependability, stimulation, and novelty.

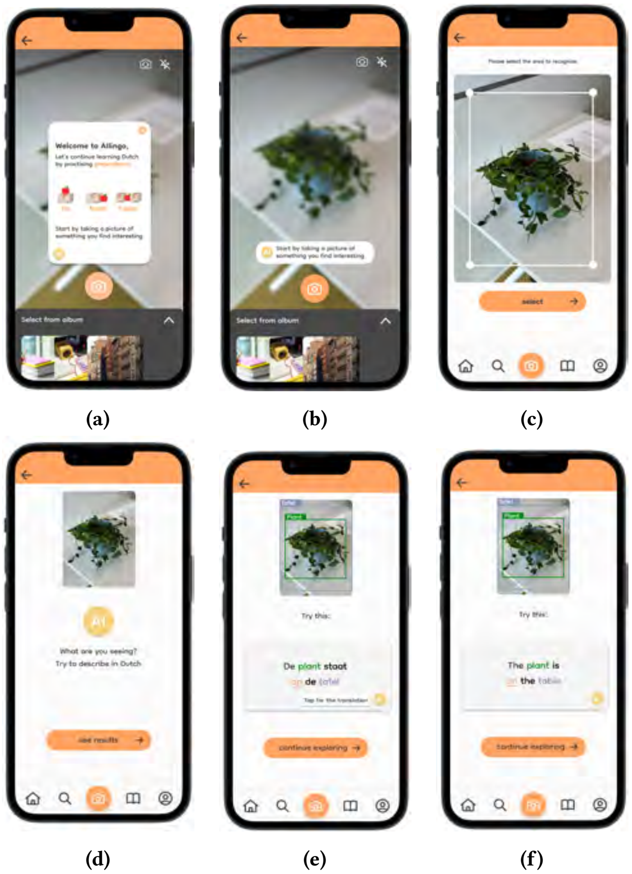


**Figure 3: Screenshots of the interaction process: The user starts at the welcome screen (a), where they can tap the camera icon to capture the scene (b). Next, the user can select an area of interest and tap "select," leading to a prompt for verbal Dutch description (c). Selecting "see results" displays the scene description in Dutch provided by AIlingo (d), while opting for "tap for translation" gives the English translation of the description (f). By tapping "continue exploring," the user can further explore the current scene. The user can start a new turn by tapping the camera icon, returning them to the camera page (b).**

Direct behavior observation was employed to observe participants' real-time interactions with AIlingo. A semi-structured interview was conducted to gain a deeper understanding of participants' experiences with AIlingo and to help interpret the quantitative data collected from UEQ. It included comparisons with the widely used language learning application, Duolingo. We evaluated AIlingo's contextual learning and immersion features, explored customization options, and discussed future implications. Participants were also asked to share their thoughts on memorable features, describe everyday use scenarios, discuss integration into their language learning routines, and raise any concerns.

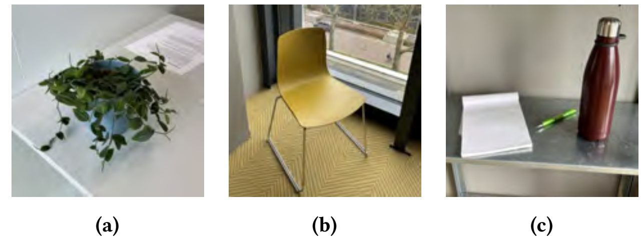


**Figure 4: We organized three scenarios for the user study: (a) a plant on a table, (b) a chair against the window, and (c) a notebook, a pen, and a bottle on a shelf.**

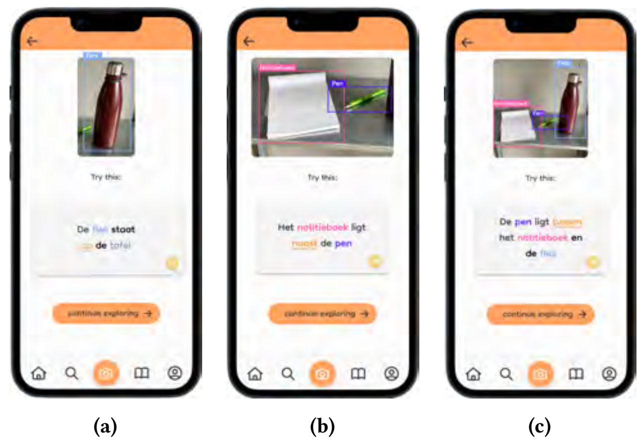


**Figure 5: Descriptions of objects in a single scenario with varying selections. (a) The bottle is standing on the table. (b) The notebook lies next to the pen. (c) The pen lies between the notebook and the bottle.**

## 4.4 Procedure

The study was carried out in a dedicated study room, which was set up with three aforementioned scenarios. The participant was required to sign consent forms before participation. They were briefed on the study's purpose and the data being collected. Prestudy questions were asked to gather demographic information, participants' proficiency in Dutch, and their prior experiences with language learning applications.

The participant was then clearly instructed on how to use the AIlingo prototype, which was prepared and opened on a smartphone provided by the researcher. They were guided to position themselves properly in the study room to ensure that the angles for capturing pictures of the learning materials matched those programmed into the prototype. Subsequently, participants were tasked with using the AIlingo prototype to interact with the three scenarios one by one while thinking aloud, which lasted approximately 10 minutes. One onsite researcher took notes on participants' confusion levels, facial expressions, and interactions with the prototype throughout this process. At the end of the task, participants completed the printed UEQ, followed by a 20-minute individual semi-structured interview.

### 4.5 Data Analysis

The data collected from UEQ, semi-structured interviews, and observation includes quantitative and qualitative data. Quantitative data obtained from UEQ was analyzed using the UEQ Excel Analysis Tool, downloaded from http://www.ueq-online.org (accessed on 5 April 2024).

The qualitative data from the interviews and observations followed the thematic analysis method [30]. All interviews were transcribed using Whisper Web[3], a machine learning-powered speech recognition tool. Three researchers independently reviewed transcripts and observation notes, identifying common topics and generating initial codes. Subsequently, they convened to resolve any discrepancies through discussion until a consensus was achieved on the coding. The coded text segments were then categorized into overarching themes. These initially identified themes were further discussed and refined to establish representative key themes. Subthemes were discerned to offer a deeper understanding. The team systematically reviewed the candidate themes in relation to the dataset, refining them through iterative discussion until consensus was attained.

## 5 RESULTS

### 5.1 Quantitative data results

According to the standard interpretation of UEQ, the range extends from -0.8 to +0.8, signifying a neutral evaluation of the corresponding scale. Values exceeding +0.8 indicate a positive evaluation, while those below -0.8 indicate a negative evaluation. Figure 6 presents mean scores of the UEQ scales. All results surpass the 0.8 threshold, indicating positive evaluations across all dimensions. Statistical results for mean and variance are outlined in Table 1.

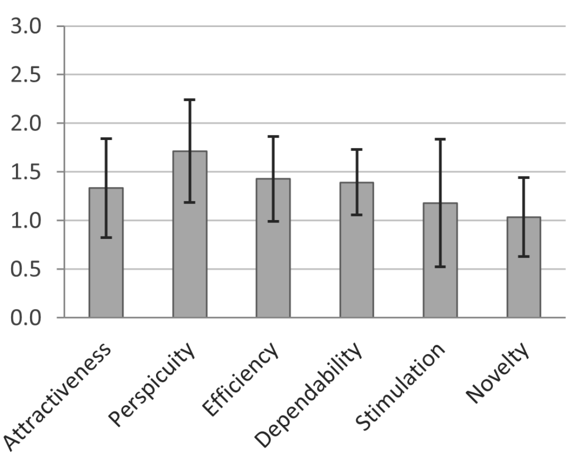


**Figure 6: Mean scores of UEQ with error bars indicating standard errors.**

Participants rated perspicuity the highest (mean = 1.71), reflecting their perception of the interaction process as straightforward and easily comprehensible. Although the variance (0.51) suggests some variability in clarity perception, the overall trend leans toward positive. Notably, the highest rating among items is the clear/confusing item (item 21) in perspicuity, with a mean of 2.0 (variance = 0.7), emphasizing the clarity of the interaction process.

**Table 1: Summary of mean and variance of the UEQ scores**

| Dimensions | Mean | Variance |
|---|---|---|
| Attractiveness | 1.33 | 0.47 |
| Perspicuity | 1.71 | 0.51 |
| Efficiency | 1.43 | 0.35 |
| Dependability | 1.39 | 0.21 |
| Stimulation | 1.18 | 0.79 |
| Novelty | 1.04 | 0.30 |

Pragmatic quality aspects, efficiency, and dependability focused on goal-directed usability received positive evaluations as well (mean = 1.43 and 1.39, respectively), with relatively low variance (0.35 and 0.21). This suggests a consistent consensus among participants regarding AIlingo's effectiveness in facilitating task completion and in fostering trust in its reliability and consistency.

Despite attractiveness receiving a moderately positive mean score (1.33), its variance (0.47) indicates variability in users' perceptions of the prototype's aesthetic appeal. Nonetheless, the design aesthetics of the prototype have an overall positive impact on the user experience.

Stimulation (mean = 1.18, variance = 0.79) and novelty (mean = 1.04, variance = 0.30), categorized as hedonic quality aspects, received lower mean scores. This suggests a range of opinions among users regarding what they found interesting, exciting, and motivating about AIlingo, while perceptions of innovation were consistently not high. This could be attributed to AIlingo's use of AI technologies that were commonly found in various software rather than being perceived as innovative. This is further demonstrated in the lowest mean values (0.4) observed for the Usual/Leading Edge item (item 15) in novelty (variance = 1.0).

### 5.2 Qualitative results

Five themes were identified from the qualitative data.

*5.2.1 Contextual learning experience.* Participants expressed that the contextual learning experience offered by AIlingo positively contributed to their motivation to use this MALL tool. The approach of taking pictures in real-world contexts as input was highlighted as a fun aspect of AIlingo that fosters user engagement. As P6 commented, this experience "is definitely more vivid." The interactive engagement with environmental elements and objects augmented the depth of the learning encounter, making it more immersive, as noted by P4: "To learn in another language then it makes you more aware of your surroundings..." Besides that, participants noted that learning from objects in familiar environments helped them better remember what they learned using AIlingo.

However, some participants did not find the concept of using their real-world surroundings as learning resources relevant to their day-to-day contexts, particularly if they were not actively trying to learn a new language. This indicates that the effectiveness of AIlingo may vary depending on individual goals and circumstances.

[3]https://huggingface.co/spaces/Xenova/whisper-web

*5.2.2 Adaptation and personalization.* Participants offered diverse perspectives on AIlingo's adaptability for users across language levels. P6 proposed its potential for beginner or intermediate learners, citing its focus on object recall as beneficial for vocabulary and foundational language skills. However, participants observed a lack of adaptability to users' proficiency levels, a noted expectation. A primary concern emerged regarding the prototype's limited variation in generated descriptions. Participants stressed the importance of alignment with individual learning goals and interests, advocating for personalization. They also suggested enhancing engagement by providing multiple descriptions for a single input image.

Furthermore, the necessity of basic language proficiency before using AIlingo was noticed, a point emphasized by P1 to facilitate comprehension and recognition of presented language elements.

*5.2.3 Role of AI: now and future.* Some participants raised concerns about the accuracy of image detection by AI in AIlingo, noting instances where detected objects differed from expectations, which may cause frustration. However, they also perceived this potential mistake as beneficial, fostering curiosity and aiding in learning new words. Furthermore, participants highlighted the potential for additional AI features, with P5 suggesting that integrating a chatbot could enhance engagement by allowing users to guide system-generated descriptions.

Additionally, participants expressed a desire for increased customization in AIlingo, advocating for a collaborative approach guided by both human input and AI capabilities. P4 proposed a model where users could provide general learning directions to AI, which would then recommend scenes or pictures for engagement. Moreover, the role of AI was emphasized in achieving adaptation to diverse user levels. Participants recommended using AI to track their strengths and weaknesses, automatically adjusting content complexity akin to AI functions in Duolingo [11], to optimize learning outcomes and maintain an appropriate level of challenge.

*5.2.4 Comparisons and use cases.* Participants offered insightful comparisons between AIlingo and other language learning tools, such as Duolingo. While appreciating AIlingo's contextual focus on everyday situations, participants expressed a desire for gamification elements similar to those found in Duolingo.

Additionally, concerns were raised regarding the practical utility of AIlingo in real-life settings. P1 compared AIlingo with Google Translate's use in the supermarket, questioning AIlingo's efficacy compared to the more stable and reliable translation service offered by Google Translate. P1 mentioned, "I need to give input, which may not make everyone comfortable in public, especially if their language skills are not so good."

Overall, participants indicated a tendency to use AIlingo regularly for casual or passive language learning, particularly during leisure time. They appreciated the tool's adaptability and flexibility across different physical contexts, enabling customized learning experiences anywhere. Commonly cited use cases included exploring unfamiliar objects, performing translation tasks, and engaging in leisure activities. Additionally, participants envisioned AIlingo as a supplementary tool alongside traditional language courses as it could improve user motivation and provide a distinctive and immersive learning experience tailored to individual preferences.

*5.2.5 Intuitive user interface.* Though not the primary focus of the study, the user interface (UI) of AIlingo emerged as a crucial component in conveying the concept of the tool and shaping user experiences. Despite occasional confusion due to the limitations of the prototype as a mockup, the clarity and ease of use of the UI were well received by the participants. They noted the UI's contribution to efficient interaction with the system, particularly appreciating features like color-coded bounding boxes and quick switching between translated and non-translated descriptions, enhancing the overall usability and flexibility.

## 6 DISCUSSION

AIlingo offers a unique approach to language learning by integrating real-world surroundings into the learning materials. Overall, the positive evaluations across the UEQ dimensions reflect the design's success in delivering a user-friendly and attractive language learning tool. This aligns with previous research on MALL, as emphasized by Huang [14], which underscores the advantages of MALL in terms of flexibility and user-friendliness. These aspects are corroborated by qualitative feedback, where participants noted the clarity of interactions and the system's effectiveness in supporting task completion. Additionally, users recognized AIlingo as suitable for learning in casual settings. Our study further expands on this by delving into user experience and the role of AI within such environments.

The score for novelty (1.04) in UEQ indicates areas for improvement in terms of making the application innovative. Participants praised AIlingo's fun and engaging nature but also expressed a desire for additional functionalities like integration with other AI technologies and gamification elements, suggesting its potential to integrate with existing language learning platforms rather than a standalone solution. By this, AIlingo can not only enhance the tool's versatility but also ensure that users have access to a comprehensive language learning experience that combines the strengths of multiple resources.

Participants appreciated AIlingo's contextual learning experience, which aligns with the recognized importance of real-world contexts in language acquisition [50]. Unlike translation tools, AIlingo captures real-world objects and transforms them into learning material. However, some found its relevance to daily needs limited, highlighting gaps in research regarding user autonomy and individual learning goals. Some participants also desired adaptation to different proficiency levels and expressed concerns about the limited variation in generated descriptions. Future research could focus on refining AIlingo's functionalities to address these concerns by incorporating varied descriptions, adaptive proficiency levels, and collaborative features.

While prior studies on applying AI in language learning primarily focused on cognitive aspects and language proficiency, our findings showcase AI's potential to enhance user experience through features like object recognition, thus boosting motivation and engagement. Despite concerns about AI's accuracy, participants recognized its value in fostering curiosity and aiding vocabulary acquisition, indicating a shift towards more user-centered approaches where AI adapts to individual needs. The appeal for additional intelligent features to refine personalized learning experiences highlights the imperative for further investigation into AI technologies.

Another interesting finding is the consistently positive evaluations of AIlingo's efficiency and dependability despite some variability in users' perceptions of attractiveness and novelty. This suggests that users prioritize practical aspects, such as task completion and reliability, when assessing language learning tools, which underscores the significance of functionality over aesthetic appeal or novelty. This finding emphasizes the importance of addressing fundamental user needs and expectations in the design and development of language learning tools.

### 6.1 Limitations

While this study provides valuable insights into the usability and potential of AIlingo for language learning, several limitations have to be noted. Firstly, the study's samples consisted primarily of several international students, resulting in a relatively small and homogeneous participant pool, limiting the generalizability and depth of the findings.

Furthermore, during the user study, we used an interactive prototype, a Figma mockup, to simulate AIlingo's functionality within a controlled environment where users interacted with predefined objects. However, participants' experiences and task performance in this simulated environment may not fully mirror real-world usage scenarios. Developing a fully functioning application would provide a more authentic user experience, allowing users to interact with AIlingo without the constraints imposed by our study's settings. Moreover, such an application should support multiple languages and accommodate a wider range of learning objectives. Building upon this functioning application, comparison studies could be conducted to further examine and define AIlingo's effectiveness relative to existing language learning applications.

Meanwhile, we recognize that addressing potential ethical issues within AI-generated content is essential to ensuring user safety, especially for younger audiences. Although this study did not explore these ethical considerations, the risk of exposure to biased or inappropriate content must be mitigated through rigorous content screening and moderation protocols. Future work could include the implementation of advanced AI filters that proactively detect and exclude sensitive or biased information, fostering a safe learning environment.

Last, the evaluation of AIlingo primarily focused on usability within beginner-level language learners, limiting insights into its adaptability across proficiency levels. Future development could incorporate AI to dynamically adjust learning materials, including vocabulary and sentence structures, based on users' proficiency levels, and further test AIlingo across diverse user groups. Additionally, the study did not directly measure actual learning outcomes, as these could be influenced by various factors such as study conditions and participants' prior mastery of Dutch vocabulary. Instead, our evaluation focused on moderating factors like attractiveness and enjoyment through UEQ. Future research should incorporate measures to evaluate factors like vocabulary retention, language proficiency, and comprehension skills to gain a more comprehensive understanding of AIlingo's impact on language learning.

## 7 CONCLUSION

This paper introduces AIlingo, an AI-enabled MALL tool designed to facilitate contextual language learning through real-world context interaction. We developed an interactive prototype using Figma for experimentation. The user study demonstrated its usability and potential for creating an immersive learning experience. By leveraging AI and real-world contextualization, AIlingo exemplifies the potential of such an AI-enabled MALL tool to transform traditional educational paradigms and provide learners with more engaging and effective learning experiences. Further development efforts should prioritize user engagement, personalization features, and refinement of more AI-enabled functionalities to reach its full potential. We hope this work will support future expeditions exploring the seamless integration of language learning into people's everyday lives and the integration of more advanced AI into language education.

## REFERENCES


[1] Abdullah Alhadiah. 2020. EFL learners' experience of a MALL-based vocabulary learning tool. *Indonesian Journal of Applied Linguistics* 10 (10 2020), 283–291. https://doi.org/10.17509/ijal.v10i2.28590

[2] Mustafa Cevikbas, Neslihan Bulut, and Gabriele Kaiser. 2023. Exploring the Benefits and Drawbacks of AR and VR Technologies for Learners of Mathematics: Recent Developments. *Systems* 11, 5 (2023). https://doi.org/10.3390/systems11050244

[3] Yong-Ting Chen Chih-Ming Chen, Jung-Ying Wang and Jhih-Hao Wu. 2016. Forecasting reading anxiety for promoting English-language reading performance based on reading annotation behavior. *Interactive Learning Environments* 24, 4 (2016), 681–705. https://doi.org/10.1080/10494820.2014.917107 arXiv:https://doi.org/10.1080/10494820.2014.917107

[4] Anind Dey. 2000. Providing architectural support for building context-aware applications. (01 2000).

[5] Wanda Dian and Made Santosa. 2022. A Systematic Review: The Implementation of MALL toward Students' Reading Skill. 10 (12 2022). https://doi.org/10.23887/jpbi.v10i2

[6] Reham El Shazly. 2021. Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems* 38 (01 2021). https://doi.org/10.1111/exsy.12667

[7] Rebecca Fox, Olga Corretjer, and Kelley Webb. 2019. Benefits of foreign language learning and bilingualism: An analysis of published empirical research 2012–2019. *Foreign Language Annals* 52, 4 (2019), 699–726. https://doi.org/10.1111/flan.12424 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/flan.12424

[8] Luke Fryer, Kaori Nakao, and Andrew Thompson. 2019. Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior* 93 (04 2019), 279–289. https://doi.org/10.1016/j.chb.2018.12.023

[9] Negin Ghorbani and Saman Ebadi. 2020. Exploring learners' grammatical development in mobile assisted language learning. *Cogent Education* 7, 1 (2020), 1704599. https://doi.org/10.1080/2331186X.2019.1704599

[10] Sufyan Habib, Ahsanuddin Haider, Samar Sufyan Mohammad Suleman, Syed Akmal, and Mohammed Arshad Khan. 2022. Mobile Assisted Language Learning: Evaluation of Accessibility, Adoption, and Perceived Outcome among Students of Higher Education. *Electronics* 11, 7 (2022). https://doi.org/10.3390/electronics11071113

[11] Parker Henry. 2023. *How Duolingo uses AI to create lessons faster.* Duolingo Blog. https://blog.duolingo.com/large-language-model-duolingo-lessons/ Accessed: April 13, 2024.

[12] A. S. Hornby. 1950. The Situational Approach in Language Teaching (I). *ELT Journal* IV, 4 (01 1950), 98–103. https://doi.org/10.1093/elt/IV.4.98 arXiv:https://academic.oup.com/eltj/article-pdf/IV/4/98/9791428/98.pdf

[13] Zhuohan Hou and Vahid Aryadoust. 2021. A review of the methodological quality of quantitative mobile-assisted language learning research. *System* 100 (2021), 102568. https://doi.org/10.1016/j.system.2021.102568

[14] Yueh-Min Huang, Shu-Hsien Huang, and Yen-Ting Lin. 2012. A ubiquitous English vocabulary learning system: Evidence of active/passive attitudes vs. usefulness/ease-of-use. *Computers & Education* 58 (01 2012), 273–282. https://doi.org/10.1016/j.compedu.2011.08.008

[15] Fenglin Jia, Daner Sun, Qing Ma, and Chee-Kit Looi. 2022. Developing an AI-Based Learning System for L2 Learners' Authentic and Ubiquitous Learning in English Language. *Sustainability* 14, 23 (2022). https://doi.org/10.3390/su142315527

[16] Zilu Jiang, Zexin Xu, Zilong Pan, Jingwen He, and Kui Xie. 2023. Exploring the Role of Artificial Intelligence in Facilitating Assessment of Writing Performance in Second Language Learning. *Languages* 8, 4 (2023). https://doi.org/10.3390/languages8040247

[17] Kadir Karakaya and Aras Bozkurt. 2022. Mobile-assisted language learning (MALL) research trends and patterns through bibliometric analysis: Empowering language learners through ubiquitous educational technologies. *System* 110 (2022), 102925. https://doi.org/10.1016/j.system.2022.102925

[18] David R Krathwohl. 2002. A revision of Bloom's taxonomy: An overview. *Theory into practice* 41, 4 (2002), 212–218.

[19] Agnes Kukulska-Hulme. 2009. Will mobile learning change language learning? *ReCALL* 21 (2009), 157 – 165. https://api.semanticscholar.org/CorpusID:215986228

[20] Agnes Kukulska-Hulme. 2020. *Mobile-Assisted Language Learning*. John Wiley & Sons, Ltd, 1–9. https://doi.org/10.1002/9781405198431.wbeal0768.pub2 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781405198431.wbeal0768.pub2

[21] Agnes Kukulska-Hulme. 2021. Conclusions: A lifelong perspective on mobile language learning. In *Mobile assisted language learning across educational contexts*. Routledge, 122–133.

[22] Karey Lan, Yao-Ting Sung, and Kuo Chang. 2013. From Particular to Popular: Facilitating EFL Mobile-Supported Cooperative Reading. *Language Learning and Technology* 17 (10 2013), 23–38.

[23] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. In *HCI and Usability for Education and Work*, Andreas Holzinger (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 63–76.

[24] Sangmin-Michelle Lee. 2022. A systematic review of context-aware technology use in foreign language learning. *Computer Assisted Language Learning* 35, 3 (2022), 294–318. https://doi.org/10.1080/09588221.2019.1688836 arXiv:https://doi.org/10.1080/09588221.2019.1688836

[25] Dongxu Li, Junnan Li, Hung Le, Guangsen Wang, Silvio Savarese, and Steven C.H. Hoi. 2023. LAVIS: A One-stop Library for Language-Vision Intelligence. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*, Danushka Bollegala, Ruihong Huang, and Alan Ritter (Eds.). Association for Computational Linguistics, Toronto, Canada, 31–41. https://doi.org/10.18653/v1/2023.acl-demo.3

[26] Junnan Li, Dongxu Li, Caiming Xiong, and Steven C. H. Hoi. 2022. BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. In *International Conference on Machine Learning*. https://api.semanticscholar.org/CorpusID:246411402

[27] Rui Li. 2023. Effects of Mobile-Assisted Language Learning on EFL Learners' Listening Skill Development. *Educational Technology & Society* 26 (09 2023), 36–49. https://doi.org/10.30191/ETS.202304_26(2).0003

[28] Jia-Cing Liang, Gwo-Jen Hwang, Mei-Rong Alice Chen, and Darmawansah Darmawansah. 2023. Roles and research foci of artificial intelligence in language education: an integrated bibliographic analysis and systematic review approach. *Interactive Learning Environments* 31, 7 (2023), 4270–4296.

[29] Tsung-Yu Liu. 2009. A context-aware ubiquitous learning environment for language listening and speaking. *Journal of Computer Assisted Learning* 25 (09 2009), 515 – 527. https://doi.org/10.1111/j.1365-2729.2009.00329.x

[30] Moira Maguire and Brid Delahunt. 2017. Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars. *All Ireland journal of higher education* 9, 3 (2017).

[31] Hiroaki Ogata and Yoneo Yano. 2004. Context-aware support for computer-supported ubiquitous learning. In *Proceddings - 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education (Proceedings - 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education)*, J. Roschelle, J. Roschelle, T.W. Chan, and S.J.H. Yang (Eds.). 27–34. https://doi.org/10.1109/WMTE.2004.1281330 Proceedings - 2nd IEEE International Workshop on Wireless and Mobile Technologies in Education ; Conference date: 23-03-2004 Through 25-03-2004.

[32] OpenAI. 2023. GPT-3.5 Turbo. https://platform.openai.com/docs/models/gpt-3-5 Accessed: January 23, 2024.

[33] OpenAI. 2023. TTS. https://platform.openai.com/docs/models/tts Accessed: January 23, 2024.

[34] Ildeniz Ozverir, Jan Herrington, and Ulker Vanci Osam. 2016. Design principles for authentic learning of English as a foreign language. *Br. J. Educ. Technol.* 47 (2016), 484–493. https://api.semanticscholar.org/CorpusID:26536143

[35] Antigoni Parmaxi. 2023. Virtual reality in language learning: a systematic review and implications for research and practice. *Interactive Learning Environments* 31, 1 (2023), 172–184. https://doi.org/10.1080/10494820.2020.1765392

[36] Samuel Porter and Mariana Sofia Castillo. 2023. The Effectiveness of Immersive Language Learning: An Investigation into English Language Acquisition in Immersion Environments versus Traditional Classroom Settings. *Research Studies in English Language Teaching and Learning* 1, 3 (2023), 155–165.

[37] Pascual Pérez-Paredes, Carlos Ordoñana Guillamón, Julie Van de Vyver, Alice Meurice, Pilar Aguado Jiménez, Grainne Conole, and Purificación Sánchez Hernández. 2019. Mobile data-driven language learning: Affordances and learners' perception. *System* 84 (2019), 145–159. https://doi.org/10.1016/j.system.2019.06.009

[38] Joseph Redmon and Ali Farhadi. 2018. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767* (2018).

[39] Guido Makransky Richard E. Mayer and Jocelyn Parong. 2023. The Promise and Pitfalls of Learning in Immersive Virtual Reality. *International Journal of Human–Computer Interaction* 39, 11 (2023), 2229–2238. https://doi.org/10.1080/10447318.2022.2108563

[40] Roboflow. 2024. *Supervision*. https://supervision.roboflow.com/latest/" Accessed: January 23, 2024.

[41] Nor Saidin, Noor Abd halim, and Noraffandy Yahaya. 2015. A Review of Research on Augmented Reality in Education: Advantages and Applications. *International Education Studies* 8 (06 2015). https://doi.org/10.5539/ies.v8n13p1

[42] Iqbal Sarker. 2022. AI-Based Modeling: Techniques, Applications and Research Issues Towards Automation, Intelligent and Smart Systems. *SN Computer Science* 3 (02 2022). https://doi.org/10.1007/s42979-022-01043-x

[43] Torben Schmidt and Thomas Strasser. 2022. Artificial Intelligence in Foreign Language Learning and Teaching: A CALL for Intelligent Practice. 33 (05 2022), 165–184.

[44] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2014. Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios. In *Design, User Experience, and Usability. Theories, Methods, and Tools for Designing the User Experience*, Aaron Marcus (Ed.). Springer International Publishing, Cham, 383–392.

[45] Olivia Seow. 2023. LingoLand: An AI-Assisted Immersive Game for Language Learning. In *Adjunct Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (, San Francisco, CA, USA,) *(UIST '23 Adjunct)*. Association for Computing Machinery, New York, NY, USA, Article 120, 3 pages. https://doi.org/10.1145/3586182.3625117

[46] Rustam Shadiev, Liu Taoying, and Wu-Yuin Hwang. 2020. Review of research on mobile-assisted language learning in familiar, authentic environments. *British Journal of Educational Technology* 51 (04 2020), 709–720. https://doi.org/10.1111/bjet.12839

[47] Rustam Shadiev, Xun Wang, Yuliya Halubitskaya, and Yueh-Min Huang. 2022. Enhancing Foreign Language Learning Outcomes and Mitigating Cultural Attributes Inherent in Asian Culture in a Mobile-Assisted Language Learning Environment. *Sustainability* 14, 14 (2022). https://doi.org/10.3390/su14148428

[48] Rustam Shadiev, Ting-Ting Wu, and Yueh-Min Huang. 2020. Using image-to-text recognition technology to facilitate vocabulary acquisition in authentic contexts. *ReCALL* 32 (02 2020), 1–18. https://doi.org/10.1017/S0958344020000038

[49] Glenn Stockwell and Philip Hubbard. 2013. Some emerging principles for mobile-assisted language learning. *The International Research Foundation for English Language Education* 2013 (2013), 1–15. https://api.semanticscholar.org/CorpusID:59898496

[50] Gesa van den Broek, Atsuko Takashima, Eliane Segers, and Ludo Verhoeven. 2018. Contextual Richness and Word Learning: Context Enhances Comprehension but Retrieval Enhances Retention. *Language Learning* (03 2018). https://doi.org/10.1111/lang.12285

[51] Hongyou Wang, Vivien Lin, Gwo-Jen Hwang, and Gi-Zen Liu. 2019. Context-aware language-learning application in the green technology building: Which group can benefit the most? *J. Comp. Assisted Learning* 35, 3 (2019), 359–377. https://doi.org/10.1111/jcal.12336

[52] Jin Ha Woo and Heeyoul Choi. 2021. Systematic Review for AI-based Language Learning Tools. arXiv:2111.04455 [cs.CY]

[53] Xing yue Qiu, Chuang-Kai Chiu, Lu-Lu Zhao, Cai-Feng Sun, and Shu jie Chen. 2023. Trends in VR/AR technology-supporting language learning from 2008 to 2019: a research perspective. *Interactive Learning Environments* 31, 4 (2023), 2090–2113. https://doi.org/10.1080/10494820.2021.1874999

[54] Ozgur Yurur, Chi Harold Liu, Zhengguo Sheng, Wilfrido Moreno, and Kin Leung. 2014. Context-Awareness for Mobile Sensing: A Survey and Future Directions. *IEEE Communications Surveys & Tutorials* 99 (01 2014), 1–1. https://doi.org/10.1109/COMST.2014.2381246