Prompting an LLM Chatbot to Role Play Conversational Situations for Language Practice

Pablo Gervás[1][a], Carlos León[1][b], Mayuresh Kumar[2][c], Gonzalo Méndez[1][d] and Susana Bautista[3][e]

[1]*Facultad de Informática, Universidad Complutense de Madrid, Madrid, 28040 Spain*

[2]*Aligarh Muslim University, Aligarh, Uttar Pradesh, India*

[3]*Universidad Francisco de Vitoria, Madrid, Spain*

{*pgervas,cleon*}*@ucm.es, 33mayuresh@gmail.com, gmendez@ucm.es, susana.bautista@ufv.es*

Keywords: AI in Education, Conversational AI, ChatGPT, Formal Language Learning, Usability Testing.

Abstract: Chatbots based on Large Language Models (LLMs) have demonstrated a remarkable ability to engage in conversations that are linguistically correct and make sense from a pragmatic point of view. A significant trait of their proven abilities is that, using verbal instructions provided as contributions to an ongoing conversation, they can be configured to provide specific content and/or modify the role that they play in the exchange. The present paper explores the feasibility of developing a framework of prompts designed with such an aim in mind. The prompting should ensure that the chatbot engages a language learner in an interaction where it proposes conversational situations of appropriate complexity, takes part in them playing the role of one of the participants, while monitoring the linguistic correctness of the contributions by the learner and providing feedback on their language performance both proactively and in response to learner requests. The paper reports on an experiment that tested this type of functionality students of Spanish as a second language at Aligarh Muslim University in India.

1 INTRODUCTION

Chatbots based on Large Language Models, by reason of their training, present significant advantages that are relevant for second language learning. First, they have operational competence in holding conversations in a given language. Second, they have knowledge of everyday life situations that might be used as settings for a conversation. Third, they have a certain understanding of the linguistic concepts involved in putting together valid sentences. Fourth, they have the ability to respond to instructions by adapting the general trend of the conversation. Fifth, they can easily switch languages to rephrase or explain problematic points. These advantages suggest they might be a good solution for language students to practice their skills on a one-to-one basis with an accommodating partner.

The present paper explores the practical feasibility of such an approach by designing specific prompts to configure a chatbot to perform in this way.

2 PREVIOUS WORK

A number of topics need to be reviewed to provide context for the work in the present paper: use of AI to support learning, the challenges involved in second language learning and the use of large language models for language learning tasks.

2.1 AI to Support Learning

A significant effort to use AI in support of learning has been invested in the field of Intelligent Tutoring Systems. An Intelligent Tutoring System (ITS) is a computer program that provides instruction adapted to the needs of individual students, doing so mainly by presenting to the student the information to be learned, asking questions, setting tasks, and providing feedback (Paladines and Ramirez, 2020). Existing re-

[a] https://orcid.org/0000-0003-4906-9837
[b] https://orcid.org/0000-0002-6768-1766
[c] https://orcid.org/0000-0002-1728-7349
[d] https://orcid.org/0000-0001-7659-1482
[e] https://orcid.org/0000-0003-1648-0208

Gervás, P., León, C., Kumar, M., Méndez, G. and Bautista, S.
Prompting an LLM Chatbot to Role Play Conversational Situations for Language Practice.
DOI: 10.5220/0013235400003932

In *Proceedings of the 17th International Conference on Computer Supported Education (CSEDU 2025) - Volume 2*, pages 257-264
ISBN: 978-989-758-746-7; ISSN: 2184-5026

search on ITSs has addressed the learning of science-related topics. Such systems were designed to help the student acquire both a set of basic concepts for the target discipline and operational procedures for solving problems associated with them. They relied on complex models of the knowledge to be learned, over which they represented the learning goals, the record of what the student has already learned, and both actual and potential misconceptions by the students. In addition, they often involved a dialogue system – to permit the interaction between the student and the system – and a set of instructional strategies to drive that interaction.

One of the few ITSs to focus on language learning was the ITS FeedBook (Rudzewitz et al., 2017), which focused on the teaching of English as a second language. A subsequent empirical evaluation of its pedagogical efficiency (Parrisius et al., 2022) argued that language teachers often face the challenge of not having sufficient time to provide tailored interaction to different students with different needs, and how the use of ITSs might provide an opportunity to address this problem.

The existence of significant surface similarities between the dialogue systems used in ITSs and the new chatbots based on LLMs gave rise to considerations of whether such chatbots might be used directly as tutoring systems (Nye et al., 2023), but the observed strengths – ability to detect misconceptions or generate mostly acceptable content – appear to be outweighed for the time being by the weaknesses – a certain likelihood to generate incorrect content due to hallucination (Ji et al., 2023) and difficulties to observe and apply existing pedagogical strategies.

A recent system that exploits the multi-modal capabilities of LLMs to help children's language learning by guiding children to describe images in a second language (Liu et al., 2024) explores the possibility of getting LLMs to respond using appropriate instructional strategies by injecting instructions in the prompt.

2.2 Challenges of Second Language Learning

When it comes to second or foreign language learning, there are various challenges that learners have to face. One of these challenges is the fear of being evaluated negatively. The students are in constant fear that they might commit a mistake. There are many researchers who have talked about this phenomenon of foreign language anxiety. Foreign language anxiety has been termed a peculiar syndrome that is just like other anxieties linked to first language use (Gregersen and Horwitz, 2002; Horwitz et al., 1986). There are multiple facets to the classroom anxiety when the students have to ask questions or get their doubts cleared in front of their peers. Some of the common fears are the ones of negative evaluation, test anxiety and the apprehension related to communication. It is this anxiety that often leads the students towards getting demotivated. Oxford and Shearin (Oxford and Shearin, 1994) have mentioned that motivation directly influences the strategies used by the students while learning a foreign language.

The challenges might be common while learning a foreign language, but it has to be looked at with an empirical context as well. The majority of the text books that are used in India, are designed keeping the communicative approach in mind (KUMAR, 2018). However, the students have had the habit and practice of learning languages like English traditionally. There are still many places where the archaic grammar-translation method is used to teach languages even today. For instance, the students at Aligarh Muslim University come from all over India and the majority of them is from a humble background. Most of them have not heard about communicative approach to learn a language. They get surprised after looking at a book like 'Aula 1' where there are no grammatical explanations but just conversations and texts.

Apparently, the students find it less challenging to respond or to ask things to a computer. There are some researches like the one by Adair-Hauck et al. (Adair-Hauck et al., 2000), which advocate for technological components to be available for students at any time of the day. There are many challenges while employing technologies also, but the intent should be towards the benefit of the students who are learning the language. The implementation of technological tools is definitely something where teachers' expertise will help a lot, but it should be the needs of the students that drive this process (Yanguas, 2018). Chatbots are the latest in the series of innovative and interactive classroom practices. They might help students learn a foreign language without facing any anxiety or being evaluated by their peers, if they are designed based on the empirical data.

2.3 LLM Chatbots

The advantages of using chatbots for language learning had already been established with solutions based on earlier technologies (Fryer and Carpenter, 2006). The recent revolution of attention-based neural solutions (Vaswani et al., 2017) resulted in the appearance of chatbots (Wu et al., 2023) based on Large Language Models (Hadi et al., 2023) that can engage in

conversations that are linguistically correct and make sense from a pragmatic point of view. A significant trait of their proven abilities is that, given instructions presented verbally as contributions to the conversation, they can adapt both the content and the role in which they participate in the conversation. This has given rise to the discipline of *prompt engineering*, the practice of creating and optimizing prompts for large language models with the intention of guiding toward the correct outputs and preventing hallucination (Liu et al., 2023).

The potential application of these LLM-based chatbot technologies to language learning has received attention in terms of general reviews of its affordances for the task (Kohnke et al., 2023), analysis of potential impact from the point of view of theories of learning (Al-Obaydi et al., 2023) and the importance of the teacher's role in supporting student interaction with the technologies (Chiu et al., 2023).

Although there is a pervasive interest in using these technologies to generate material to inform traditional language learning (Baskara et al., 2023; Kim et al., 2023; Yıldız, 2023), some of the reviews on prospective use mention the potential of these technologies to provide a conversational interaction (Kohnke et al., 2023; Yang, 2023; Hong, 2023; Yıldız, 2023). In such interactions, the students can exercise their language skills dynamically, and they can benefit also from the ability of the chatbot to present students with virtual simulations of everyday life situations on which to train (Al-Obaydi et al., 2023). In this paper we choose to focus on this aspect, to allow much-needed exploration of the free interaction between the student and this type of chatbot (Han, 2024).

Report on actual use of this type of chatbot for language learning include: helping students learn English as a second language (Yıldız, 2023; Al-Obaydi et al., 2023; Shaikh et al., 2023). These experiences focused on English at the target language, and often on writing tasks rather than conversational practice. Of particular relevance for the work developed in this paper is (Shaikh et al., 2023), which reports on the use of a post-task questionnaire to gather feedback on the students' experience during their interaction with ChatGPT. Even though this work focused on vocabulary learning for English, we have found it appropriate to apply as existing assessment method for chatbot language learning.

3 LLM CHATBOT AS LANGUAGE PRACTICE PARTNER

At the start of the interactive session, the students are led through basic routine to ensure that they are familiar with the AI chatbot. They are provided with the written script for configuring the chatbot and shown by a tutor how to apply the instructions to set up an interactive session with the chatbot. Each student interacts with the chatbot for a period of time of one hour and a half. Once the allotted time has passed, students are asked to record the conversation, and they are given a survey to complete.

The experiment is carried out for a set of students whose level of Spanish ranges between A1 and B2. To allow each student to customise the chatbot response to their level the following two prompts were suggested.

Beginner students were encouraged to provide the chatbot with the following prompt:

> *I'm a student at university. I'm from India. I'm learning Spanish. I'm practising Spanish. I can speak native Hindi and English. I have A1 Spanish level. The conversation must be in A1 Spanish. Praise my advances. Use Spanish for the conversation. If I make any mistakes please explain them to me in English.*

Advanced students were encouraged to use the following prompt:

> *I'm a student at university. I'm from India. I'm learning Spanish. I'm practicing Spanish. I can speak native Hindi and English. I have a B1 Spanish level. The conversation must be in B1 Spanish. Praise my advances. Let me know when I make a mistake and explain it. Use Spanish for the conversation. Unless I tell you otherwise, use Spanish also for explaining my mistakes and for explaining what you are doing.*

The prompts were designed to ensure that the chatbot proactively engaged in interactions intended to help the students practice their Spanish. These usually involve engaging the student in simple conversations in the language.

In both cases students were asked to converse with the chatbot to get used to the mechanics of the interaction. The students were also encouraged to ask questions about any point they did not understand, and to remember the chatbot is a conversational agent, so they should work with it by asking it questions or giving it instructions.

Although the interactions of the students with the chatbot to this point also count as language practice,

we explicitly wanted to explore the ability of the chatbot to engage in role-playing interactions of the kind usually employed to practice skills in a foreing language. To this end we asked the students to introduce the following prompt:

> *¿Puedes plantearme una situación en la que yo tengo que hablar español, y luego jugar tú el papel de alguno de los personajes, para que yo practique mi conversación en español?*[1]

Initial experiments showed that the chatbot has a tendency in these cases to generate the whole conversation instead of letting the student participate. Students were advised to take the opportunity to review it, then insist using the following prompt:

> *Quiero que escenifiquemos la conversación entre tú y yo. Tienes que decir tú la primera frase y esperar a que yo diga la siguiente antes de continuar*[2].

These instructions were intended to allow students unfamiliar with the chatbot to experience the type of situations of conversational interaction that we wanted to explore.

4 EXPERIMENTS AND RESULTS

There are a number of aspects of the interaction that need to be analysed in detail.

Language students involved in this type of interaction are likely to make mistakes in their use of the language they are attempting to learn. It is important that we evaluate the extent to which the chatbot is capable of identifying such mistakes, whether it can point them out and whether it can correct them appropriately.

On the other hand, the performance of the chatbot itself needs to be evaluated, both at the level of linguistic correctness of its contributions and at the level of any contributions that can be associated to the specific pedagogical strategies that it has been prompted to apply.

In addition, we also evaluate the perception that the students have of how usable the chatbot is in this context.

[1]Can you set up a situation where I have to speak Spanish, and then you play the role of one of the characters, so that I can practice my Spanish conversation?

[2]I want us to stage the conversation between you and me. You have to say the first sentence and wait for me to say the next one before continuing.

Table 1: Averages for levels of Spanish language skills (1–5) for participating students (as self-declared in post-experiment questionnaire).

Listening	Writing	Speaking	Reading
4.40	4.35	4.50	4.35

The sessions described here were carried out at Aligarh Muslim University earlier this year. They involved $N = 20$ students, with 12 of them being male (60%) and 8 of them being female (40%). The students in this cohort had different levels of skill in Spanish language. The average values for this, as captured in the survey they completed after the experiment, are reported in Table 1.

The responses on Spanish language skill level are generally consistent across the four skill areas, and roughly break down into 4 students at level 3, 4 students at level 4 and 12 students at level 5.

4.1 Assessing Linguistic Correctness of the Interactions

At the end of the interactive session, the students were asked to share the link to their conversation with ChatGPT and email it to the researchers. The conversations recorded during the session were reviewed by an experienced teacher of Spanish.

When a student mentions an interest in language learning, the chatbot offers help as a practice partner. If the student accepts, the chatbot engages the student in a simple conversation, and gently corrects any obvious mistakes made. The proposed prompts help focus on particular levels of language difficulty. Overall, three different modes of interaction are observed: if conversation is requested, the chatbot offers engaging conversation with the student; if exercises are requested, the chatbot proposes language exercises to solve; if clarifications are requested, the chatbot provides explanations of related linguistic concepts. In conversation mode, the chatbot tends to forego the correction of mistakes. If the request occurs in a context where exercises have been mentioned before, the chatbot tends to provide the transcript of a complete conversation rather than engage the student in an interactive manner. Some of the prompts proposed for the students to use were intended to help them break out of such situations. In exercise mode, the chatbot usually provides exercises that require the student to fill in words missing from a set of given sentences. When the student solves the proposed exercises, the chatbot tends to provide the list of solutions, indicating which ones were correct and incorrect in the students' response, and often justifying why. In explanation mode, the chatbot tends to provide a list of bullet

Table 2: Metrics on the recorded interactions, grouped by overall details, student performance and chatbot performance.

Aspect	Average	Min	Max	SD
Overall				
Total length	24.95	5	56	14.57
Role-play span	7.45	0	37	13.05
Student				
Ask to Clarify	0.40	0	2	0.38
Errors	8.60	0	39	9.58
Chatbot				
Set Exercise	1.15	0	5	1.38
Praise	3.15	0	13	3.79
Encouragement	0.75	0	3	0.83

points that cover the main concepts of the linguistic feature being considered.

Over a total of 20 recorded interactions, 3 students used the prompt to set the level to A1 and 3 used the prompt to set the level to B1.

Table 2 reports on some essential metrics computed over the recorded interactions. These metrics capture both the size and the different nature of the interactions and the use of the desired pedagogical strategies by the chatbot.

The metrics show a wide discrepancy in the values reported for the length of the interactions. Shorter interactions may be associated with students that have difficulty carrying out a conversation, whether because they are unfamiliar with the chatbot mechanisms or they have poorer language skills.

In spite of our insistence that the exercise focus on role-playing situations, on average only about a third of the total length of the interactions achieved this goal. In specific terms, the observed performance is that only 23.34 % of the recorded interactions (in terms of number of turns over complete number of turns of each conversation) actually involves a conversation where the chatbot and the student are each role playing a different character in a situation. This is undesirably low. Improvements may be achieved by informed refinement of the prompt suggestions provided by the students.

A number of the interactions involved the chatbot proposing language exercises to the students. This is spontaneous behaviour of the chatbot on learning that the user is interested in language learning. Revisions of the prompts may be required to block this behaviour if role playing conversation is preferred.

The recorded interactions show instances of the chatbot providing both praise and encouragement to the student. However, both strategies are applied when the chatbot is operating in exercise mode and not applied when the chatbot is operating in role playing conversation mode. This is probably related to the chatbot having been trained specifically during fine-tuning that explicitly included these strategies.

Table 3: Metrics on linguistic correctness.

Av. % student errors identified	47.48
Av. % appropriate corrections (over identified errors)	64.33
Av. % inappropriate corrections (over identified errors)	35.67

Regarding linguistic correctness, the chatbot has not been found to make any linguistic mistakes, but it often fails to identify mistakes committed by the student. Corrections volunteered by the chatbot are mostly correct. The actual statistics on this evaluation are presented in Table 3.

It is important to note that detection of errors was carried out successfully when the chatbot was operating in exercise mode, but was much less in evidence when the chatbot was role-playing a conversation. If detection and correction of errors is deemed to be positive for learning, the proposed prompts may be revised to ensure error detection is also applied in that mode. The chatbot has certainly shown itself capable of the task.

We did observe cases where a student challenged corrections made by the chatbot and the chatbot backed down and gave the student's contribution as valid even though it was incorrect. This problem is related to the chatbot's default behaviour of backpedalling with no hesitation if challenged by the user. Although this may be a useful strategy to limit the impact of conceptual mistakes inherent to the technology, it may have a negative impact on its applicability as a teaching aid.

Only two instances were recorded where the chatbot refused a request by the student: both were situations where the student asked the chatbot for advice leading to illegal activities.

Finally, on the issue of which language to employ, we observed that, given the proposed prompts, the chatbot will generally respond in the selected target language (Spanish). If the student issues instructions in a different language (English), the chatbot will generally answer in Spanish. Overall, we observed that 97.68 % of the dialog turns in the full set of recorded interactions have been in Spanish.

4.2 Assessing Learner Satisfaction

The students were asked to complete an on-line survey that included the questions used in (Shaikh et al., 2023) to evaluate a similar experiment of using Chat-

Table 4: Values for aspects of usability covered by SUS and USE questionnaires: present experiment and (Shaikh et al., 2023).

	Max. score	Mean	SD	Min	Max	Skewness	Kurtosis
Our Experiment							
SUS	50	40.45	8.91	20	50	-1.22	0.19
Usefulness	40	36.30	5.21	23	40	-1.27	0.87
Ease of Use	55	49.25	8.01	28	55	-1.97	4.19
Ease of Learning	20	19.02	2.09	12	20	-2.58	5.82
Satisfaction	35	32.95	4.36	20	35	-2.15	3.84
Sheik et al., (2023)							
SUS	35	29.90	3.00	26	35	0.15	-0.76
Usefulness	40	30.80	5.35	19	38	-1.12	1.85
Ease of Use	55	43.30	7.60	31	54	-0.52	-0.91
Ease of Learning	20	17.50	2.36	13	20	-1.00	0.15
Satisfaction	35	29.70	3.88	21	35	-1.07	2.16

GPT to help Norwegian students learn vocabulary for English.

The questionnaire includes five different sections: (1) Demographics information, (2) Previous Spanish language knowledge level, (3) Feedback using "Usefulness, Satisfaction, and Ease of Use (USE)" Questionnaire, (4) Feedback using "System Usability Scale (SUS)" Questionnaire, (5) Perception relative to aspects being considered.

The responses for the first two sections have been reported in the section above.

The USE questionnaire (Gao et al., 2018) is used to evaluate the usability of a technical system. Users are expected to score 30 items divided into 4 subgroups (usefulness, ease of use, ease of learning, satisfaction) on a Likert scale from 1 to 5 by showing their disagreement or agreement. A score of 1 is strong disagreement and 5 is strong agreement on the scale. Following (Shaikh et al., 2023) for comparability, for each aspect of the USE questionnaire we computed a score as the ratio of the sum of values provided by the user for the set of questions on that aspect. Similarly, the SUS questionnaire (Brooke et al., 1996) measures the perceived usability of a software system. It also uses a Likert scale from 1 to 5 to score 10 items. Table 4 reports the values obtained for our experiment side by side with those reported by (Shaikh et al., 2023).

Comparing our results with those reported in (Shaikh et al., 2023) we can see that the results of our experiment are slightly higher in all categories (SUS: $40.45 > 29.90$, Usefulness: $36.30 > 30.80$, Ease of Use: $49.25 > 43.30$, Ease of Learning: $19.02 > 17.50$, Satisfaction: $32.95 > 29.70$). The negative values for skewness show the asymmetrical behaviour of the responses with respect to a normal distribution, with the values concentrated on the right of the distribution; the relatively high values for kurtosis in some of the items respond to the fact that most users scored those items with a 5, whereas very few users ranked them with lower values.

Table 5 summarizes student responses to the part of the questionnaire that focused on the specific task they had been asked to carryout. The answers tend towards the positive side of the spectrum (for all the questions the mode value was 5). The questionnaire also included questions aimed to gather qualitative comments. These comments were all were positive or very positive, and no negative comment was collected.

The relatively lower value (4.20) for the question on clarifications is supported by the very low numbers for clarifications requested in Table 2. The same goes for the response to question on explanations in English (3.65), which the analysis above shows to be very scarce. The questions about mistakes may be negatively affected by the students' inability to perceive how many mistakes they were making. This needs to be revised on future versions of the survey. The student's perception of the frequency of praise may be related to differences between students receiving more role playing and students receiving more exercises.

5 CONCLUSIONS

The experiment reported in this paper indicates that even when LLM-based chatbots of this kind are not trained or fine-tuned specifically for the task, they can come up with descriptions of everyday situations suitable for exercising language skills in a role playing mode, (if asked) they can explain the relevant linguistic concepts involved, and they can identify any linguistic mistakes made by the students and correct them. The chatbot has also been seen to retain and

Table 5: Reported scores on relevant aspects.

	Mean	SD
Did you feel that the chatbot proposed conversation topics that allowed you to practice the targeted feature of the language?	4.60	0.68
Did you request clarifications?	4.20	1.15
Did you need to ask for explanation in your native language?	3.65	1.42
If you asked for clarification in English, did the chatbot remember to return to Spanish after explaining?	4.60	0.60
Did the chatbot remember to correct your mistakes?	4.30	0.92
Did you have to remind it to correct your mistakes?	4.05	1.32
Did you at any point receive praise from the chatbot?	4.25	1.12
Do you think you have learnt new languages skills?	4.65	0.59
Do you think the session helped you practice your Spanish?	4.80	0.52

obey the guideline to switch back to the target language after explanations. The observed interactions show instances of use of pedagogical strategies such as praise and encouragement, though these seem related to particular modes of interaction – exercise solving – that may be the result of specific fine-tuning. In general terms, the chatbot responds to instructions by adapting the general trend of the conversation to particular topics requested by the students.

Other features desirable for language learning still need to be tested empirically such as the ability to switch languages on demand or the ability to adjust the complexity of explanations to different levels of expertise.

With respect to the challenges specific to second language learning as described in Section 2.2, the chatbot shows a commendable flexibility to depart from the grammar / translation approach to language teaching by entering into conversations of specific interest to the student. In the students' conversation logs we have seen topics related to the Golden Age in Spain, the conquest of America, serial killers, shopping, or how to get a scholarship to study in Spain.

The chatbot's default behaviour of not picking out students' mistakes during conversation may have helped to reduce the students' fear of being evaluated negatively, which may lead to longer lasting engagement in conversation. There is however a risk that overlooking serious mistakes may have a negative impact on the student's learning progress.

An often voiced concern is that the use of ChatGPT weakens real-life communication and interaction between people (Al-Obaydi et al., 2023; Li, 2024; Chen, 2024) and critical thinking (Li, 2024). Whereas this may be the case in situations where students access such resources remotely from their home, during the experiments reported in this paper, which involved students accessing the chatbot from rows of computers placed in a laboratory, we observed significant interactions between the students to comment on the replies. These comments often involved disparaging remarks on particular responses by the chatbot. Overall, we believe that the risk of such negative impacts of chatbot technologies may not necessarily be greater than for books or videos, in the sense that these more traditional technologies also provide opportunities for students to isolate themselves, without the added opportunity of having a dynamic interaction. It is also important to note that traditional technologies are also subject to occasional errors in accuracy or bias.

ACKNOWLEDGEMENTS

This publication is part of the R&D&I projects DARK NITE, PID2023-146308OB-I00, funded by MICIU/AEI/10.13039/501100011033/, ADARVE (SUBV20/2021), funded by the Spanish Council of Nuclear Security; CANTOR (PID2019-108927RB-I00), funded by the Spanish Ministry of Science and Innovation; and EA-DIGIFOLK (101086338), funded by the European Commission and the SPARC project Developing applications for learning Spanish in India using articial intelligence and digital media, grant No. P2557 (Indian Ministry of Human Resource Development).

REFERENCES

Adair-Hauck, B., Willingham-McLain, L., and Youngs, B. E. (2000). Evaluating the integration of technology and second language learning. *CALICO journal*, pages 269–306.

Al-Obaydi, L. H., Pikhart, M., and Klimova, B. (2023). ChatGPT and the General Concepts of Education: Can Artificial Intelligence-Driven Chatbots Support

the Process of Language Learning? *International Journal of Emerging Technologies in Learning (iJET)*, 18(21):39–50.

Baskara, R. et al. (2023). Exploring the implications of ChatGPT for language learning in higher education. *Indonesian Journal of English Language Teaching and Applied Linguistics*, 7(2):343–358.

Brooke, J. et al. (1996). Sus-a quick and dirty usability scale. *Usability evaluation in industry*, 189(194):4–7.

Chen, X. (2024). The Application of ChatGPT in Second Language (L2) Learning Classrooms: Opportunities and Challenges. *Transactions on Social Science, Education and Humanities Research*, 5:132–137.

Chiu, T. K., Moorhouse, B. L., Chai, C. S., and Ismailov, M. (2023). Teacher support and student motivation to learn with Artificial Intelligence (AI) based chatbot. *Interactive Learning Environments*, pages 1–17.

Fryer, L. and Carpenter, R. (2006). Bots as language learning tools. *Language Learning & Technology*, 10(3).

Gao, M., Kortum, P., and Oswald, F. (2018). Psychometric evaluation of the USE (usefulness, satisfaction, and ease of use) questionnaire for reliability and validity. *Proceedings of the human factors and ergonomics society annual meeting*, 62(1):1414–1418.

Gregersen, T. and Horwitz, E. K. (2002). Language learning and perfectionism: Anxious and non-anxious language learners' reactions to their own oral performance. *The Modern Language Journal*, 86(4):562–570.

Hadi, M. U., Qureshi, R., Shah, A., Irfan, M., Zafar, A., Shaikh, M. B., Akhtar, N., Wu, J., Mirjalili, S., et al. (2023). A survey on large language models: Applications, challenges, limitations, and practical usage. *Authorea Preprints*.

Han, Z. (2024). ChatGPT in and for second language acquisition: a call for systematic research. *Studies in Second Language Acquisition*, 46(2):301–306.

Hong, W. C. H. (2023). The impact of ChatGPT on foreign language teaching and learning: Opportunities in education and research. *Journal of Educational Technology and Innovation*, 5(1).

Horwitz, E. K., Horwitz, M. B., and Cope, J. (1986). Foreign language classroom anxiety. *The Modern language journal*, 70(2):125–132.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y. J., Madotto, A., and Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12):1–38.

Kim, S., Shim, J., Shim, J., et al. (2023). A study on the utilization of OpenAI ChatGPT as a second language learning tool. *Journal of Multimedia Information System*, 10(1):79–88.

Kohnke, L., Moorhouse, B. L., and Zou, D. (2023). ChatGPT for language teaching and learning. *Relc Journal*, 54(2):537–550.

KUMAR, M. (2018). Peculiaridades de la enseñanza del español en la india y el uso del hindi en las clases. In *Monográficos SinoELE, 17. IX Congreso Internacional de la Asociación Asiática de Hispanistas*, pages 286–291.

Li, R. (2024). Current Survey of ChatGPT Empowering Foreign Language Education. *International Journal of Educational Research and Practice*, 12(1):11–17.

Liu, P., Yuan, W., Fu, J., Jiang, Z., Hayashi, H., and Neubig, G. (2023). Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9):1–35.

Liu, Z., Yin, S. X., Lee, C., and Chen, N. F. (2024). Scaffolding language learning via multi-modal tutoring systems with pedagogical instructions. *arXiv preprint arXiv:2404.03429*.

Nye, B., Mee, D., and Core, M. G. (2023). Generative large language models for dialog-based tutoring: An early consideration of opportunities and concerns. In *AIED Workshops*.

Oxford, R. and Shearin, J. (1994). Language learning motivation: Expanding the theoretical framework. *The modern language journal*, 78(1):12–28.

Paladines, J. and Ramirez, J. (2020). A systematic literature review of intelligent tutoring systems with dialogue in natural language. *IEEE Access*, 8:164246–164267.

Parrisius, C., Wendebourg, K., Rieger, S., Loll, I., Pili-Moss, D., Colling, L., Blume, C., Pieronczyk, I., Holz, H., Bodnar, S., et al. (2022). Effective features of feedback in an intelligent tutoring system-a randomized controlled field trial (pre-registration).

Rudzewitz, B., Ziai, R., De Kuthy, K., and Meurers, D. (2017). Developing a web-based workbook for english supporting the interaction of students and teachers. In *Proceedings of the joint workshop on NLP for Computer Assisted Language Learning and NLP for Language Acquisition*, pages 36–46.

Shaikh, S., Yayilgan, S. Y., Klimova, B., and Pikhart, M. (2023). Assessing the usability of ChatGPT for formal English language learning. *European Journal of Investigation in Health, Psychology and Education*, 13(9):1937–1960.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L., and Tang, Y. (2023). A brief overview of chatgpt: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, 10(5):1122–1136.

Yang, X. (2023). ChatGPT Empowers the Informatization of Foreign Language Education in Colleges and Universities. In *2023 International Conference on Educational Knowledge and Informatization (EKI)*, pages 39–42. IEEE Computer Society.

Yanguas, Í. (2018). Technology in the Spanish Heritage Language Classroom: Can Computer-Assisted Language Learning Help? *Hispania*, 101(2):224–236.

Yıldız, T. A. (2023). The Impact of ChatGPT on Language Learners' Motivation. *Journal of Teacher Education and Lifelong Learning*, 5(2):582–597.