Alexa and Copilot: A Tale of Two Assistants

Todericiu Ioana Alexandra[a], Dioşan Laura[b] and Şerban Camelia[c]
Faculty of Mathematics and Computer Science, Babeş-Bolyai University, Cluj-Napoca, Romania
{*ioana.todericiu, laura.diosan, camelia.serban*}*@ubbcluj.ro*

Keywords: Virtual Assistants, Comparative Study, Amazon Alexa, Microsoft Copilot Studio.

Abstract: As virtual assistants (VAs) become essential to contemporary interactions, it is imperative to understand how to evaluate their functionalities. This study offers a comparison framework for assessing the design and execution of Amazon Alexa and Microsoft Copilot Studio, emphasizing their capabilities in question-answering activities. Through the examination of their deterministic and probabilistic approaches, we evaluate response times, precision, flexibility, and linguistic support. We have developed a systematic framework to assess the strengths and shortcomings of each VA, utilizing educational queries as a realistic test case that elucidates the influence of design decisions on performance. Our study lays the groundwork for choosing an appropriate VA according to particular needs, assisting developers and organizations in traversing the varied realm of VA technologies. Regardless of whether precision or adaptability is prioritized, our approach facilitates an educated decision, simplifying the process of aligning the appropriate VA with the corresponding circumstance.

1 INTRODUCTION

The field of technology gained a lot of momentum in the past few years, especially with the grand entrance of ChatGPT (Zarifhonarvar, 2023). As it turned out, it was just the conversation starter. Today, the amount of new innovation based on AI is considerable. Currently, the landscape is abundant with numerous AI-driven ideas, platforms, and tools developed everyday. In this fast-paced technological environment, the notion of agents—particularly intelligent agents—has attracted significant interest.

Intelligent agents are engineered to independently execute tasks for users, utilizing algorithms and machine learning to solve complex tasks. Some argue there is still some way to go until we reach ”independence” for agents, but until we do so, we can regard agents as task-focused tools that can re-engineer the way we used to do certain actions (Xiao et al., 2024). In a similar fashion, just a few years back, the topic of virtual assistance started to emerge. Virtual assistants utilize natural language processing and machine learning to engage users, fostering interactive experiences that improve productivity and facilitate information access (Kusal et al., 2022).

Establishing a link between intelligent agents and virtual assistants reveals a significant opportunity to leverage their functionalities in educational settings (Katsarou et al., 2023). As these technologies become more incorporated into education, they can significantly transform how students learn and engage with knowledge. Virtual assistants can function as customized learning companions, delivering personalized feedback, responding to inquiries in real-time, and granting access to an extensive array of resources that correspond with individual learning trajectories.

Virtual Assistants are utilized throughout various sectors, including healthcare and customer service, where they enhance support operations by addressing routine inquiries (Dojchinovski et al., 2019; Fadhil, 2018; Yadav et al., 2023). In smart home systems, virtual assistants such as Amazon Alexa facilitate effortless management of domestic gadgets, hence augmenting user convenience (Iannizzotto et al., 2018). These many uses highlight the adaptability and promise of VAs to revolutionize interactions across sectors.

The synergy between education and these technologies can foster a dynamic learning environment that empowers students to take control of their educational paths. Integrating intelligent assistants into educational systems can cultivate an engaging and adaptive learning experience that addresses the varied needs of contemporary learners (Bilad et al., 2023; Jayadurga and Rathika, 2023).

[a] https://orcid.org/0000-0002-2469-134X
[b] https://orcid.org/0000-0002-6339-1622
[c] https://orcid.org/0000-0002-5741-2597

Alexandra, T. I., Laura, D. and Camelia, Ş.
Alexa and Copilot: A Tale of Two Assistants.
DOI: 10.5220/0013174900003890

In *Proceedings of the 17th International Conference on Agents and Artificial Intelligence (ICAART 2025) - Volume 1*, pages 445-452
ISBN: 978-989-758-737-5; ISSN: 2184-433X

Given the wide array of existing virtual assistants (VAs) and the limited availability of systematic methodologies or frameworks for their comparison, we propose a framework for analyzing and comparing the same functional VA implemented differently. The importance of this approach lies in addressing the gap in literature, which typically offers various taxonomies and classifications of VAs but provides limited guidance on how to compare two VAs from their design and implementation stages (Islas-Cota et al., 2022). Existing comparisons often focus on aspects like user experience or the AI component quality, but lack a systematic, quantifiable approach. By developing a structured comparison framework, we aim to provide a more thorough understanding of these differences. This framework is demonstrated through an educational VA example, highlighting the features common to other question-answering systems (QASs) (Biancofiore et al., 2024) and the unique aspects introduced in the Alexa/Copilot implementation. The subsequent research questions address both their present capabilities and their long-term prospects.

RQ1: How do the design and implementation methodologies of Amazon Alexa and Microsoft Copilot differ in handling question-answering tasks in educational settings?

RQ2: How do Amazon Alexa and Microsoft Copilot compare in terms of their effectiveness and efficiency in delivering question-answering capabilities?

2 BACKGROUND

2.1 Review and Taxonomy

Virtual assistants are an essential subset of Question Answering Systems (QASs), evolving from simple information retrieval tools into sophisticated interactive systems (Biancofiore et al., 2024).

The taxonomy of intelligent assistants (IAs) classifies these systems based on their objectives, capabilities, user interactions, and deployment methods (Islas-Cota et al., 2022). This study focuses on VAs like Amazon Alexa and Microsoft Copilot, particularly those supporting question-answering in education. Their objectives center on enhancing learning and offering tailored support to students, aligning with the educational focus of the IA taxonomy.

Both VAs leverage Natural Language Processing (NLP) and personalization, allowing them to interpret user queries and adapt responses. While Microsoft Copilot relies on text-based inputs, making it suitable for written communication, Amazon Alexa uses audio-based inputs, creating a voice-driven interaction. These differences reflect their adaptability to user preferences and contexts.

Powered by AI and Machine Learning (ML), these VAs are deployed on distinct devices: Alexa through smart speakers and Copilot via personal computers or web applications. This influences their user engagement styles, emphasizing their roles as question-answering systems within education. By positioning them within this taxonomy, the study provides a structured comparison, highlighting their unique strengths and adaptability.

2.2 Applications in Different Sectors

Virtual Assistants (VAs) have diverse applications across various sectors, transforming service delivery and user interaction (de Barcelos Silva et al., 2020).

In healthcare, VAs assist patients by managing appointments, providing access to health information, and supporting telemedicine, which became especially crucial during the COVID-19 pandemic (Sezgin et al., 2020). In customer service, VAs like chatbots handle routine queries, improving efficiency and customer satisfaction while reducing costs (Yadav et al., 2023). Additionally, VAs like Amazon Alexa enhance smart home systems, allowing users to control devices with voice commands, which increases convenience and accessibility (Martins et al., 2020).

In education, Virtual Assistants (VAs) have become essential for enhancing digital learning by providing personalized tutoring, managing student inquiries, and assisting with time management. They are used in online learning environments to maintain student engagement and provide interactive experiences during remote classes (Liao and Pan, 2023). VAs have also been implemented in universities to assist students with administrative queries and provide campus information through voice-activated systems like Amazon Alexa (Cernian et al., 2021). Moreover, VAs are being used to enhance learning through interactive quizzes and assessments, providing real-time feedback and making learning more engaging and adaptive to student needs (Ioana-Alexandra et al., 2024). These developments highlight the potential of VAs to transform various sectors by providing tailored support, enhancing user experiences, and improving overall outcomes.

3 RELATED WORK

The field of VAs encompasses a wide range of implementations, each tailored to specific user needs. While existing research often emphasizes the practical applications and user experiences of these systems, there is a lack of structured methodologies for evaluating VAs from the design phase onward. This study seeks to address this gap by focusing on the underlying design choices that shape VA capabilities.

Reyes et al. (Reyes et al., 2019) propose a system for deploying educational virtual assistants via Google Dialogflow, emphasizing organized material delivery to improve student learning experiences. This work is vital for comprehending the systematic design of educational virtual assistants, although it prioritizes reproducibility over the comparative analysis of design decisions. This article contrasts two distinct VAs, Amazon Alexa and Microsoft Copilot, emphasizing how their design philosophies influence their adaptation and efficacy in educational settings.

The work of Todericiu and Serban (Ioana-Alexandra et al., 2021) presents an ontology-based approach to improve accessibility in education through the use of smart speakers like Amazon Alexa. Their study focuses on using structured ontologies to enable effective information retrieval through voice commands, making VAs more accessible for diverse user needs in educational settings. Even though it provides a conceptual formal framework that can be replicated across multiple VAs, it does not compare and contrast the capabilities of different VA systems in varied educational contexts. This contrasts with the current study, which compares Alexa's structured capabilities with the more flexible, adaptive design of Microsoft Copilot, emphasizing how each approach affects user interaction and educational outcomes.

Holstein et al. examine the integration of AI-powered systems inside adaptive learning settings, emphasizing its capacity to deliver real-time interventions depending on student behavior. Their research underscores the capacity of AI systems to function as "learning companions," adapting in real-time to the student's speed and learning preferences (Holstein et al., 2019). This article aligns with their findings by demonstrating that Copilot, via its machine-learning capabilities, provides more personalized and adaptable feedback in contrast to the static, rule-based interactions of Alexa. This versatility is crucial for accommodating diverse learning requirements and improving student engagement.

This paper contributes to the growing body of literature by presenting a structured approach for comparing two VAs—Amazon Alexa and Microsoft Copilot Studio—at the level of design and implementation. By doing so, it provides insights that can guide the development of more effective VAs across various domains, including education.

4 A COMPARATIVE FRAMEWORK FOR VIRTUAL ASSISTANTS

4.1 Developing Platforms

The shared characteristic of effective virtual assistants is the platform that allows creators to innovate and tailor capabilities and skills to fulfill user requirements. Platforms such as Amazon Alexa Developer Console and Microsoft Copilot Studio function as essential tools, enabling developers to create customized educational experiences that improve student engagement and learning results. Although both platforms seek to facilitate a user-friendly and accessible creation of different-scoped interactions, the way in which they achieve this is unique to both.

Amazon Alexa functions inside a deterministic framework, wherein user interactions are characterized by established intents and answers. This design ensures a dependable and uniform user experience, rendering it especially efficient for simple inquiries. Developers can design targeted skills that enable students to obtain information, pose inquiries, and receive prompt feedback in a regulated manner. As the interactive user experience can support so much, for greater customization, different services within Amazon Web Services can be leveraged.

In comparison, Microsoft Copilot Studio employs a probabilistic methodology, utilizing machine learning algorithms to adjust replies according to user interactions. Copilot facilitates a dynamic learning environment by offering real-time support and contextually relevant responses, using retrieval augmented search to retrieve information from different sources. Moreover, its flawless interaction with the Microsoft ecosystem, encompassing programs such as Teams and Word, amplifies collaborative learning prospects, facilitating more effective student collaboration.

Both platforms provide distinct approaches to enhancing the educational experience, and we will explore their functionalities and implementations in greater detail throughout this paper.

4.2 Implementation

In the context of Amazon Alexa, the development of skills starts with the formulation of distinct intentions that align with user inquiries. Developers employ the Alexa Developer Console to clearly declare these intents, guaranteeing that each interaction is predefined and replies are uniform. For instance, when a student requests information regarding their class schedule, developers formulate an intent explicitly for that goal, aligning it with a structured response - please see the following figure. To enhance functionality, developers may incorporate AWS services, such as AWS Lambda, to manage dynamic requests. This connection enables the assistant to access other cloud services, such as a DynamoDB database for real-time information, including the retrieval of individual class schedules based on user input. The deterministic foundation of Alexa guarantees customers receive dependable responses, rendering it especially efficient for simple queries where consistency is essential (Serban and Todericiu, 2020).

On the other hand, the implementation of Microsoft Copilot Studio facilitates a more dynamic and adaptive methodology. A key characteristic of Copilot Studio is its capacity to define intents utilizing natural language. Developers can express their intentions in simple language, and the underlying large language model (LLM) converts these into distinct "topics," similar to intents in Alexa - please see the following figure. This technique incorporates a probabilistic component, enabling the assistant to learn from user interactions and modify its responses accordingly.

An essential aspect of the implementation entails defining the "knowledge" element of the skill. In Copilot Studio, knowledge refers to the information and context utilized by the assistant to deliver pertinent responses. The knowledge can be wide, from public websites, files and even structured data, such as databases. This knowledge base can be continually expanded and enhanced through user interactions, facilitating a more tailored experience that develops over time.

Additionally, developers can define particular activities within Copilot Studio that the assistant is capable of executing in response to user inquiries. These acts may include offering resources and recommendations as well as enabling collaborative tasks among students. Utilizing the probabilistic characteristics of Copilot, the assistant can modify its activities according on user behavior, thus improving engagement and responsiveness. Actions usually are used for more complex tasks, such as reading from databases, sending an email, and much more.

Both platforms allow for a no-code/ low-code approach when it comes to defining user's requests. Moreover, they also complement this with the possibility of defining more complex actions such as different capabilities in Power Platform, in case of Copilot Studio, or enhance the conversation via code by connection to external hosted functions, such as AWS Lambda, in case of Alexa. The implementation complexity is directly proportional to the complexity of the requirements, both platforms offering flexibility to go from zero to hero.

4.3 Language Support

Language support is a crucial aspect of the usability and effectiveness of virtual assistants, particularly in diverse educational environments. Amazon Alexa skill was designed in English, leading to great performance for people interacting in that language. Nonetheless, a considerable obstacle emerges for non-native English speakers, especially concerning pronunciation. Users may encounter difficulties in having their orders effectively recognized due to accent variances or linguistic subtleties, resulting in misunderstandings or erroneous replies. Although Alexa has broadened its support for multiple languages, its comprehension and contextual awareness can differ markedly among these languages, frequently resulting in mistakes during user interactions in languages other than English (Moussalli and Cardoso, 2020).

In contrast, Microsoft Copilot Studio utilizes the functionalities of large language models (LLMs), like GPT, which have the ability to analyze and produce text in several languages. Copilot Studio accommodates multiple languages, demonstrating a notable capacity to interact with users in diverse linguistic environments. Nonetheless, owing to the versatility of the foundational GPT technology, users can engage in languages that may not be expressly enumerated as supported. This adaptability permits Copilot to expand its linguistic capabilities, allowing it to comprehend and address inquiries in a broader array of languages (Armengol-Estapé et al., 2022).

This capability offers a unique advantage, although it also creates issues with linguistic precision and contextual comprehension. Responses may range in relevancy due to specific phrasing or cultural nuances inherent in other languages. During testing, it was observed that Microsoft Copilot Studio sometimes mixes up languages; if prompted to respond in language X or Y, it may occasionally answer in a different language other than the one used by the user. Additionally, if the dataset provided to Copilot is in a different language from the user's request, the assistant may respond in the language in which the knowledge is presented rather than the language of the inquiry.

In conclusion, although Amazon Alexa excels in its primary language, issues with pronunciation and comprehension may limit its efficacy. In contrast, Microsoft Copilot Studio advantages from its underlying design that accommodates several languages, along with the adaptability provided by GPT, facilitating communication in languages not explicitly enumerated, but its performance is not always excellent. As educational institutions increasingly cater to different populations, the capacity to deliver appropriate language support will be essential for the efficacy of virtual assistants in improving the learning experience.

4.4 Deployment

Upon finalizing the development of a skill for Amazon Alexa, the final product is the Alexa skill, which can be deployed on the Alexa platform. Developers must utilize the Alexa Developer Console to publish a skill, allowing for comprehensive testing to verify that its functionality and user experience adhere to their criteria. Upon completion of testing, developers submit the skill for certification. This certification procedure guarantees that the skill adheres to Amazon's standards for privacy, security, and usability(Chakraborty and Aithal, 2023).

Upon successful certification, the skill becomes available to users on many Alexa-enabled devices, including Echo speakers, smartphones, and tablets. This deployment enables students to engage with the skill through voice commands, facilitating convenient access to information and help. A student can inquire, "Alexa, what classes do I have today?" and obtain prompt, tailored responses derived from the skill's programming and data integration.

The implementation of solutions created in Microsoft Copilot Studio offers a more adaptable methodology. Upon the development of a Copilot assistant, it can be deployed inside the Microsoft ecosystem, facilitating integration with programs such as Microsoft Teams and other Microsoft 365 services. Moreover, Copilot solutions can be implemented on either a demonstration website, provided by Microsoft, or a custom website.

In addition to conventional applications, Copilot Studio facilitates deployment across several platforms, such as Slack, custom mobile apps, Telegram, and Direct Line Speech. This broad array of deployment choices guarantees that the assistant can connect with users through many channels, improving accessibility and user engagement. A student may utilize the Copilot assistant in Microsoft Teams to obtain study tips or access collaborative resources pertinent to their academic endeavors, or through a mobile application for convenient support.

4.5 Quality Assessment

4.5.1 Response Speed

The velocity of response is a critical determinant of user satisfaction and engagement, especially in the context of VAs, where prompt answers is essential.

Response speed was assessed by tests utilizing a defined set of 100 inquiries for each platform. These queries were created to encompass a variety of both simple and more complex topics related to the educational context. The average response times measured the duration from when a user submitted a question to when the assistant provided a response.

Amazon Alexa generally attains an average response time of 2.5 seconds for simple queries that roughly correspond with established intentions. When users conform to the prescribed language, the skill can provide results very instantaneously. Nevertheless, if the inquiry diverges from the specified intents—such as when a student inquires, "Can you provide my schedule for today?"—the response time may extend to 3–4 seconds while the system tries to digest the input and align it with the closest intent, query the appropriate dataset and formulate a response.

A significant factor in Alexa's performance is its reliance on AWS Lambda for managing dynamic requests. Upon initial invocation or following a period of inactivity, a serverless function may undergo a "cold start" (Vahidinia et al., 2020), leading to extended reaction times. During a cold start, AWS Lambda must initialize the execution environment, potentially increasing the response time by 1 to 3 seconds, contingent upon the skill's complexity and the resources needed. The cold start phenomena presents a possible latency challenge in serverless architectures, where ensuring an ideal user experience is essential, especially in dynamic educational settings. As an alternative, a dedicated server function mitigates cold start delays but entails fixed maintenance expenses, necessitating consideration of budget and performance needs.

In contrast, Microsoft Copilot Studio typically attains an average response time of 2–3 seconds for basic orders, which is comparable to the performance of Amazon Alexa. For intricate inquiries necessitating comprehensive analysis or contextual comprehension, response times may extend to 4-5 seconds. The diversity in response time is mostly due to the utilization of GPT (Generative Pre-trained Transformer) technology, enabling Copilot to comprehend a wider array of user inputs. Although GPT improves the assistant's capacity to produce nuanced and contextually appropriate responses, this probabilistic characteristic may result in extended processing durations as the model evaluates and adjusts to user behavior (Rogora et al., 2020). Therefore, the abundance of interactions provided by Copilot results in occasional delays, especially when addressing complex or multi-faceted inquiries.

4.5.2 Correctness Rate

The correctness rate is an essential indicator for evaluating the accuracy of responses given by virtual assistants. It directly influences user trust and the overall efficacy of the assistant in educational settings. To assess accuracy, 5 questions asked 50 times each were presented to each assistant. Responses were classified as "relevant," "somewhat relevant," or "incorrect," with accuracy rates determined by the proportion of correct responses relative to the total number of questions posed.

When users stick to with established intents, Alexa attains a 100% accuracy rate for inquiries. This deterministic approach ensures dependable responses, as the system is engineered to provide accurate answers when user commands correspond with the designated intents. Nevertheless, if the inquiries are reformulated in a way that deviates from the established expressions for each intent—specific terms that the assistant is trained to identify—the accuracy rate may diminish. This shortcoming underscores a critical facet of Alexa's functionality: although it thrives in situations with explicitly specified intents, its efficacy may decline when confronted with diverse linguistic expressions. Consequently, developers must ensure that the skills contain a diverse array of expressions to accommodate various ways users may articulate their inquiry, thereby reducing potential misunderstandings.

The accuracy of Microsoft Copilot Studio is assessed by its capacity to deliver pertinent responses across diverse instructions. These questions were identical for both MS Copilot Studio and Amazon Alexa. During the evaluation, five particular questions were presented to the assistant, with each question repeated 50 times to measure the consistency and precision of the responses. The findings are encapsulated in Table 1, which classifies the responses as "relevant", "somewhat relevant", and "incorrect". The evaluation procedure aimed to replicate authentic questions that students may raise, concentrating different day-to-day university related inquiries.

The performance assessment of Microsoft Copilot Studio indicates both strengths and challenges in its capacity to deliver pertinent and precise answers to user questions. The findings indicate differing levels of success across various inquiry kinds, underscoring the significance of context and the inherent difficulties of natural language processing.

The inquiry regarding recent university events was derived from a collection of pages from the university website, encompassing both the homepage and the events page. The assistant discovered 22 pertinent responses; nevertheless, a significant portion of the "somewhat relevant" entries (14) related to outdated events rather than the most recent occurrences. The misclassification probably occurred because the system identified related terms like "event" and "symposium," resulting in the inclusion of these outdated records as pertinent, while ignoring the rest of the announcements that weren't so obviously labeled as "events". Moreover, occurrences of hallucination were observed when the assistant delivered generic responses, such as enumerating general university events, unrelated to the data set (McIntosh et al., 2024).

When users requested, "I need a mentor," the assistant excelled due to the structured approach built into the query handling process. This inquiry was structured with distinct steps, instructing the assistant to initially inquire about the mentorship topic, subsequently solicit the student's email, and ultimately establish communication with the professor. The findings revealed that 47 responses were pertinent, indicating that the procedure typically operated efficiently. Nevertheless, there were instances in which the assistant defaulted to offering generic guidance on locating tutors, resulting in hallucinations that presented irrelevant information rather than fostering specific mentorship ties. The performance was efficient due to its rather deterministic and structured approach, that left little to interpretations.

The inquiry into class timetables presented further difficulties. The knowledge set used for the inquiry "What classes do I have on Monday?" was a CSV file comprising students' schedules. Despite the assistant providing 19 pertinent responses, it occasionally presented only a partial list of classes rather than the entire timetable. Hallucinations were noted, with the assistant recommending users to "consult your schedule" instead of offering definitive responses. This suggests that although the system can get structured data, its capacity to deliver thorough and precise information may be impeded by the methods of data querying and interpretation.

When it came to internship announcements, the behaviour was similar with the latest university news. Some answers were relevant, others outdated, and while some were simply untrue. This highlights the necessity for enhanced contextual filtering to ensure that only the latest and most relevant internship opportunities are presented.

Table 1: Correctness of Responses (Effectiveness) of Microsoft Copilot Studio.

Question	**Relevant Responses**	**Somewhat Relevant**	**Incorrect Responses**	**Total Queries**
Latest university events.	22	14	14	50
I need a mentor.	46	4	0	50
What classes I have on Monday?	19	24	7	50
What are the latest internship announcements?	28	16	6	50
Tell me about Erasmus opportunities.	33	13	4	50

Ultimately, the inquiry regarding Erasmus prospects produced a mix of answers. The assistant generated predominantly relevant comments, with 33 entries categorized as pertinent. The precision of the phrase "Erasmus" seems to assist in pinpointing pertinent material. Nevertheless, there were occasions when it merely redirected users to the generic Erasmus website or lacked meaningful information, which doesn't necessarily point towards the way the topic was created, but also on the importance of prompting.

Overall, the assessment of Microsoft Copilot Studio reveals its versatility and promise in educational environments, while also highlighting key areas for improvement, especially in hallucinations, precision and contextual comprehension. Confronting these problems will be crucial for optimizing the assistant's efficacy and improving the educational experience for students.

4.6 Comparison Conclusions

In this sub-section, we will revise the findings of previous sub-sections and see how they address the research questions.

RQ1: How do the design and implementation methodologies of Amazon Alexa and Microsoft Copilot differ in handling question-answering tasks in educational settings?

The analysis in Section 3.1 shows that Amazon Alexa uses a deterministic approach, relying on predefined intents configured through the Alexa Developer Console. This ensures consistent responses, making it effective for straightforward queries like retrieving schedules. In contrast, Microsoft Copilot uses a probabilistic approach, leveraging large language models (LLMs) and natural language understanding (NLU) for more dynamic responses. This allows Copilot to adapt to varied user queries, making it suitable for complex interactions, though it can result in variability in accuracy.

RQ2: How do Amazon Alexa and Microsoft Copilot compare in terms of their effectiveness and efficiency in delivering question-answering capabilities?

As discussed in Section 3.4, Alexa generally offers faster responses for predefined queries but may experience delays due to cold starts when using AWS Lambda. Copilot's response times are similar for simple queries but increase for complex ones due to LLM processing demands. Correctness testing shows Alexa's high accuracy with defined intents, but a decline when queries deviate. Copilot, while more adaptable, can suffer from occasional hallucinations in responses. In terms of language support, Section 3.2 highlights Alexa's strength in English and its list of supported languages and Copilot's versatility with multiple languages, though the latter may sometimes mix languages.

Overall, Section 3 indicates that Alexa is best for consistent, simple queries, while Copilot excels in handling varied, dynamic interactions. Both platforms offer distinct advantages depending on the educational context and user needs.

5 CONCLUSION

In conclusion, through the course of this paper, we observe how Amazon Alexa and Microsoft Copilot Studio demonstrate that both systems possess remarkable functionalities, although they also present distinct quirks and challenges. Alexa excels in its systematic methodology, providing dependable results when user orders are clear, yet falls short when confronted with the unexpected turns of natural language and varied pronunciations. On the other hand, Copilot Studio advances the frontier with its probabilistic model and adaptability, demonstrating the promising potential of LLMs-driven interactions. Nonetheless, it is not without its challenges—mixing languages and sporadically deviating from the intended course serve as a reminder that even state-of-the-art technology had areas for enhancement.

As we approach a new era in educational technology, the potential of these tools is substantial. They hold the promise of transforming how students engage with information and interact with their educational environments. However, let us not deceive ourselves; this is merely the beginning. The pursuit of developing genuinely intuitive and encouraging learning companions is in its early stages, propelled by continuous innovations and refinements. The future of education is poised to become far more intelligent.

REFERENCES

Armengol-Estapé, J., de Gibert Bonet, O., and Melero, M. (2022). On the multilingual capabilities of very large-scale English language models. *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3056–3068.

Biancofiore, G., Deldjoo, Y., Di Noia, T., Di Sciascio, E., and Narducci, F. (2024). Interactive question answering systems: Literature review. *ACM Computing Surveys*, 56.

Bilad, M., Yaqin, L. N., and Zubaidah, S. (2023). Recent progress in the use of artificial intelligence tools in education. *Jurnal Penelitian dan Pengkajian Ilmu Pendidikan: e-Saintika*.

Cernian, A., Tiganoaia, B., and Orbisor, B. (2021). An exploratory research on the impact of digital assistants in education. *eLearning and Software for Education*.

Chakraborty, S. and Aithal, S. (2023). Let us create an alexa skill for our iot device inside the aws cloud. *International Journal of Case Studies in Business, IT, and Education*, pages 214–225.

de Barcelos Silva, A., Gomes, M. M., da Costa, C. A., da Rosa Righi, R., Barbosa, J. L. V., Pessin, G., De Doncker, G., and Federizzi, G. (2020). Intelligent personal assistants: A systematic literature review. *Expert Systems with Applications*, 147:113193.

Dojchinovski, D., Ilievski, A., and Gusev, M. (2019). Interactive home healthcare system with integrated voice assistant. *2019 42nd International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*, pages 284–288.

Fadhil, A. (2018). Beyond patient monitoring: Conversational agents role in telemedicine & healthcare support for home-living elderly individuals. *ArXiv*, abs/1803.06000.

Holstein, K., Mclaren, B., and Aleven, V. (2019). *Designing for Complementarity: Teacher and Student Needs for Orchestration Support in AI-Enhanced Classrooms*, pages 157–171.

Iannizzotto, G., Bello, L. L., Nucita, A., and Grasso, G. (2018). A vision and speech enabled, customizable, virtual assistant for smart environments. *2018 11th International Conference on Human System Interaction (HSI)*, pages 50–56.

Ioana-Alexandra, T., Pop, M., Serban, C., and Diosan, L. (2024). Quiz-ifying education: Exploring the power of virtual assistants. pages 589–596.

Ioana-Alexandra, T., Serban, C., and Laura, D. (2021). Towards accessibility in education through smart speakers. an ontology based approach. *Procedia Computer Science*, 192:883–892.

Islas-Cota, E., Gutierrez-Garcia, J. O., Acosta, C. O., and Rodríguez, L.-F. (2022). A systematic review of intelligent assistants. *Future Generation Computer Systems*, 128:45–62.

Jayadurga, D. R. and Rathika, M. S. (2023). Significance and impact of artificial intelligence and immersive technologies in the field of education. *International Journal of Recent Technology and Engineering (IJRTE)*.

Katsarou, E., Wild, F., Sougari, A.-M., and Chatzipanagiotou, P. (2023). A systematic review of voice-based intelligent virtual agents in efl education. *International Journal of Emerging Technologies in Learning (iJET)*.

Kusal, S., Patil, S., Choudrie, J., Kotecha, K., Mishra, S., and Abraham, A. (2022). Ai-based conversational agents: A scoping review from technologies to future directions. *IEEE Access*, 10:92337–92356.

Liao, X.-P. and Pan, X. (2023). Application of virtual assistants in education: A bibliometric analysis in wos using citespace. *Proceedings of the 2023 14th International Conference on E-Education, E-Business, E-Management and E-Learning*.

Martins, D., Parreira, B., Santos, P., and Figueiredo, S. (2020). Netbutler: Voice-based edge/cloud virtual assistant for home network management. pages 228–245.

McIntosh, T. R., Liu, T., Susnjak, T., Watters, P., Ng, A., and Halgamuge, M. N. (2024). A culturally sensitive test to evaluate nuanced gpt hallucination. *IEEE Transactions on Artificial Intelligence*, 5(6):2739–2751.

Moussalli, S. and Cardoso, W. (2020). Intelligent personal assistants: can they understand and be understood by accented l2 learners? *Computer Assisted Language Learning*, 33:865 – 890.

Reyes, R., Garza, D., Garrido, L., de la Cueva, V., and Ramírez, J. (2019). Methodology for the implementation of virtual assistants for education using google dialogflow. pages 440–451.

Rogora, D., Carzaniga, A., Diwan, A., Hauswirth, M., and Soulé, R. (2020). Analyzing system performance with probabilistic performance annotations. pages 1–14.

Serban, C. and Todericiu, I.-A. (2020). Alexa, what classes do i have today? the use of artificial intelligence via smart speakers in education. *Procedia Computer Science*, 176:2849 – 2857.

Sezgin, E., Huang, Y., Ramtekkar, U., and Lin, S. M. (2020). Readiness for voice assistants to support healthcare delivery during a health crisis and pandemic. *NPJ Digital Medicine*, 3.

Vahidinia, P., Farahani, B., and Shams Aliee, F. (2020). Cold start in serverless computing: Current trends and mitigation strategies. pages 1–7.

Xiao, Y., Cheng, Y., Fu, J., Wang, J., Li, W., and Liu, P. (2024). How far are llms from believable ai? a benchmark for evaluating the believability of human behavior simulation.

Yadav, R. R., Sawarkar, D., Dhurwade, A., Kawtikwar, P., and Pansare, D. (2023). Implementing intelligent virtual assistant. *International Journal of Advanced Research in Science, Communication and Technology*.

Zarifhonarvar, A. (2023). Economics of chatgpt: a labor market view on the occupational impact of artificial intelligence. *Journal of Electronic Business & Digital Economics*, 3.