Digital Assistant in a Point of Sales

Emilia Lesiak[a], Grzegorz Wolny[b], Bartosz Przybył and Michał K. Szczerbak[c]
Orange Innovation Poland, Al. Jerozolimskie 160, Warsaw, Poland
{*emilia.lesiak2, grzegorz.wolny, partosz.przybyl, michal2.szczerbak*}*@orange.com*

Keywords: Digital Assistant, Voice User Interface (VUI), Customer Engagement, Multilingual Support, Experimental Analysis, Technological Adaptability, Customer Service Innovation, Retail Point of Sales.

Abstract: This article investigates the deployment of a Voice User Interface (VUI)-powered digital assistant in a retail setting and assesses its impact on customer engagement and service efficiency. The study explores how digital assistants can enhance user interactions through advanced conversational capabilities with multilingual support. By integrating a digital assistant into a high-traffic retail environment, we evaluate its effectiveness in improving the quality of customer service and operational efficiency. Data collected during the experiment demonstrate varied impacts on customer interaction, revealing insights into the future optimizations of digital assistant technologies in customer-facing roles. This study contributes to the understanding of digital transformation strategies within the customer relations domain emphasizing the need for service flexibility and user-centric design in modern retail stores.

1 INTRODUCTION

The rapid advancements in artificial intelligence (AI) are reshaping various industries, with telco operators being significant beneficiaries. This sector, crucial for global connectivity, faces the dual challenge of escalating customer expectations and the need to remain competitive in a swiftly evolving market. To address these challenges, there is a growing reliance on innovative technologies such as chatbots, voicebots, and videobots. The integration of these digital tools is viewed as a strategic response to enhance customer interactions, streamline operations, and maintain market relevance.

Among these technologies, digital assistants equipped with Voice User Interfaces (VUIs), including those with graphical screen displays, are gaining prominence. These interfaces promise to make customer service interactions more natural and engaging. This study focuses on the deployment of a fully animated digital character, acting as an assistant powered by a voice interface, at a sales point. The aim is to assess its impact on user satisfaction, engagement, and problem-resolution efficacy. Our study was guided by the following research questions:

RQ1. How do digital assistants influence customer engagement and problem-resolution outcomes?

RQ2. What are the challenges and opportunities associated with the use of VUI technologies in practical service environments?

By exploring these questions through a systematic experimental setup, this paper seeks to contribute new insights into digital transformation strategies within customer service. It aims to evaluate the effectiveness of digital assistants and explore their potential in enhancing the quality of service delivery. This introduction sets the stage for detailed discussions on methodology, findings, and implications for future innovations and best practices in customer relations.

The article explores the application of a digital assistant in a retail setting, examining the roles and integration of artificial intelligence and VUIs within digital assistants. It starts by setting the technological context, followed by a detailed description of the experimental design, including the deployment, data collection methodologies, and analysis procedures. The results section discusses the impact and implications of these technologies in service environments, highlighting both challenges and opportunities discovered during the experiment. Finally, the paper concludes by summarizing key findings and offering recommendations for further research and practical application with a focus on improving customer service through digital assistants.

[a] https://orcid.org/0009-0005-8713-2093
[b] https://orcid.org/0009-0006-4590-9236
[c] https://orcid.org/0009-0000-6323-4806

Lesiak, E., Wolny, G., Przybył, B. and Szczerbak, M. K.
Digital Assistant in a Point of Sales.
DOI: 10.5220/0013042500003912

In *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2025) - Volume 1: GRAPP, HUCAPP and IVAPP*, pages 439-451
ISBN: 978-989-758-728-3; ISSN: 2184-4321

2 RELATED WORK

Voice-User Interfaces, which first emerged in the 1990s, represent a significant advancement in human-machine interaction, facilitating direct voice communication between users and systems. The history of VUIs reflects their evolution from simple task automation, like call routing, to complex interaction management through sophisticated natural language processing, thereby enhancing user experience and operational efficiency. Notable early systems include AT&T's voice recognition call processing which adeptly directed calls based on voice commands(Wilpon et al., 1990).

The development of advanced systems such as Spoken Dialogue Systems (SDS) and Embodied Conversational Agents (ECAs) marked a significant expansion in the scope of VUIs. These systems integrate speech with other modalities such as body language and facial expressions to create more engaging and natural interactions. Studies like those by (Cassell et al., 2000) and (Breazeal, 2004) detail how these integrations enhance the immersiveness of human-computer interactions, a crucial aspect for user acceptance.

Artificial Intelligence (AI), Machine Learning (ML), and Natural Language Processing (NLP) have further propelled the capabilities of digital assistants, expanding their functionality to encompass a broad range of tasks including internet searches, schedule management, and smart device control. Digital assistants such as Siri, Google Assistant, Alexa, and Cortana exemplify this technological progression, providing invaluable assistance in daily activities and redefining user interaction with digital platforms (Luger and Sellen, 2016), (McTear et al., 2016), (Amershi et al., 2019).

Research into VUI design emphasizes the necessity for systems to accurately understand and process natural language inputs while providing intuitive, contextually relevant responses. Challenges related to speech recognition accuracy, accommodating diverse user accents, and the naturalness of system-generated responses persist, highlighting the complexity of these systems. (Morgan and Balentine, 1999), (McTear et al., 2016) and (López et al., 2018) provide a comparative analysis of speech recognition technologies, illustrating the technological advancements and remaining hurdles in achieving seamless interaction.

In customer service, digital assistants have demonstrated the potential to revolutionize service delivery by ensuring availability around the clock, reducing response times, and personalizing user interactions. Insights from (Xu et al., 2019) show significant impacts on customer satisfaction and operational efficiency. A survey study reported in (Brill et al., 2022) points out, however, that the most important factor for the users is their confirmation of expectations, which the company needs to be properly align with the proposed functionality of their digital assistants.

A more detailed decomposition of digital assistants' aspects and their impact on both general attitude and purchase intentions is studied in (Balakrishnan and Dwivedi, 2024) - the surveys on young population of India show significant correlations with factors like perceived anthropomorphism, usefulness and intelligence. Moreover, the studies by (Xie et al., 2020) underline the importance of transparency, reliability, and security in building user trust and acceptance, pivotal for positive customer experiences.

Emerging trends in digital assistants development, such as the integration of multimodal interfaces and the application of decentralized technologies like blockchain, point towards a future where digital assistants are not only more capable and secure but also better tailored to diverse user needs (Oviatt et al., 2018), (Mik, 2019). (Velkovska, 2019) highlights the potential of emotionally aware characters that recognize and respond to human emotions, enhancing the quality of customer relationships. The latter work describes also the closest experiment to the one described in this document.

Studies could further benefit from a deeper exploration of ethical considerations in development of digital assistants, especially as these systems become more autonomous and integrated into everyday life. (Van Kleef and van Trijp, 2018) discuss the importance of ethical transparency, particularly around nudging techniques in consumer interactions. A more extensive discussion on these ethical issues would provide a comprehensive view of the responsibilities of developers and designers in this field.

In conclusion, this review underscores the vital integration of advanced technology and human-centric design in the development of voice-user interfaces for digital assistants. As the society progresses deeper into the age of artificial intelligence, it is crucial that VUIs not only enhance functionality but also prioritize ethical design, privacy, and emotional intelligence. The challenges identified here highlight the dynamic nature of the field and the opportunities for innovative research. The lack of clear answers to our research questions in the literature highlights the need for our experiment to fill these gaps and to provide results from a real-life VUI-equipped digital assistant usage.

3 EXPERIMENT SETUP

The integration of a digital assistant within a retail environment represents an advancement in merging technologies with voice user interfaces to enhance customer service and increase sales efficiency. This study was conducted in a high-traffic retail store known for its tech-savvy clientele and commitment to sustainability. This setup aimed to investigate the utility of the digital assistant in traditional retail settings, focusing on optimizing customer interactions across various service scenarios, including product inquiries and supporting sales. One might notice, that this setup, as well as the three months period adopted for the experiment have certain similarities to the setting in (Velkovska, 2019), where the researchers observed a robotic arm with a face visualization using also a VUI in a telco shop in Paris.

For our assistant, however, we used a three-dimensional digital character created using Unity 3D software. The character's body was styled as a futuristic robot with mechanical elements and smooth, white surfaces, emphasizing its modern and technological nature. We chose this digital character for our study because it has become the central figure in Orange Poland's advertising campaigns, significantly enhancing the brand's visibility and its connection with customers across various platforms.

The assistant's movements were fluid and realistic, as well as synchronized with speech, making communication with it more natural and engaging. Additional graphics and elements supporting interactions could appear on the screen, such as changing skins and artifacts, which could be adjusted depending on the research context. Utilizing motion sensors, the character could respond to the presence of clients, actively initiating interactions, such as waving in greeting. All these features made the digital character not only effectively support the conduct of research but also serve as an attractive and interactive element designed to increase user engagement.

The digital assistant was tailored to align with the store's operational strategies and customer service needs. The implementation process began with an in-depth analysis of business and customer needs, defining key functionalities such as conversation management, product presentation, and support during visits to the point of sales. Component selection was key to effectively support these features, ensuring the assistant was user-friendly and easily accessible. The main functionalities introduced a set of features aimed at improving customer interactions:

Figure 1: The stand in the point of sales.

Figure 2: Christmas / winter "skin".

- Sales Pokes: Employing both verbal and visual cues, the assistant proactively informs customers about current and seasonal offerings. This feature was designed to engage customers in a two-stage process, initially capturing their attention with prompts and subsequently encouraging further exploration through detailed follow-ups.
- Conversation Engine: At the core of the assistant is a powerful conversational interface, capable of handling a wide array of discussions, from specific sales details to broader customer service inquiries.
- Phone Recommendations: A feature that enhances the shopping experience by offering personalized mobile phone recommendations based on individual user preferences and specific requests.
- Multilingual Capabilities: The assistant's ability to communicate in multiple languages, including English, French, and Ukrainian, in addition to Polish, to enhance accessibility to a diverse demographic.
- Feedback Loop: This functionality collects and analyzes user feedback, playing a crucial role in

the ongoing development cycle of the assistant, facilitating continual enhancements based on user interactions.

The rollout timeline for these features spanned from December 1, 2023, to February 2024, structured to progressively introduce and enhance functionalities:

- December 2023: Implementation of multilingual support and the "poke - learn more" feature to improve user engagement and accessibility.
- January 2024: Introduction of the personalized phone recommendation feature and graphical updates to the assistant's avatar, enhancing the visual and interactive appeal.
- February 2024: Deployment of targeted sales pokes and refined conversation prompts aimed at boosting interaction quality, coupled with incentives for deeper exploration of the assistant's capabilities.

The systematic implementation of these features aimed to gradually improve the user experience by increasing usability, accessibility and engagement. The gradual introduction of new elements also contributed to increasing analytical and reasoning capabilities, providing knowledge on the possibilities of improving and optimizing the functionality of the assistant.

4 ASSISTANT'S DESIGN

This section describes the technical architecture of the voice-based assistant and its hardware and software components as used in our study, starting by outlining the user journey during interactions with the assistant.

4.1 User Journey

A person entering the point of sale is detected by the assistant, which either waves or greets them. An intrigued user can approach the stand and change the default language from Polish by touching a corresponding national flag pictogram for French, English, or Ukrainian.

The user presses and holds a push-to-talk button to ask a question, releasing it once done. The question is analyzed, with the assistant's eyes displaying turning gears. After processing, the assistant responds with audio and visual effects, such as animations or offer images, which the user may click for more details.

The assistant primarily provides information, covering telco-related terms, offers, general knowledge, or weather. Requests requiring action or identification (sales, after-sales service) are redirected to human staff. If the button is used incorrectly, the assistant provides instructions.

The conversation ends when the user leaves or says "goodbye". The assistant then invites the user to evaluate the experience on a 1 to 5 scale, either vocally or via the touchscreen showing five stars.

4.2 Architecture

These functionalities required both a front-end physical stand and back-end software services, which communicate over a network.

Input/Output Interface. The assistant is presented via a nearly 2-meter-high stand with a 55 cm diameter touchscreen, as shown in Figure 1. The casing hides the hardware, including the CPU and cables, while the screen displays the assistant's avatar: an animated robot. The model moves and gestures according to its activity, such as waiting, processing requests, or responding.

The stand includes a microphone and speakers for voice interactions. No video from the camera is captured. We used a press-and-hold interface for push-to-talk (P2T), avoiding more complex speech detection methods that could fail in noisy environments.

Additionally, the stand employed infra-red and ultrasonic sensors for detecting presence and measuring distance. These acted as the assistant's "eyes", replacing the use of cameras for legal reasons.

On-site Software. The robot model, animations, and scene were powered by a Unity-based application. Sensors triggered predefined robot actions, like waving when someone passed by or greeting a stationary person. Other software components handled data bridging the user interface with the back-end services.

The physical stand and its software could be accessed remotely, enabling easy monitoring, troubleshooting, and updates to introduce new features in line with the study timeline.

Interactions Data Flow. The back-end processed audio, converting it to text, which was passed to the conversational engine for handling. Textual responses were generated, converted back into audio, and returned to the stand. Audio-text conversions mostly relied on cloud services, except for Polish text-to-speech, which used an open-source solution.

The conversational engine used was Rasa X Community Edition[1], which manages dialogues between

[1] https://rasa.com/rasa-x/

the assistant and users. We defined intents and responses to build the model for handling conversations. Text for possible queries was provided in Polish, while the multilingual BERT model handled language understanding.

We extended the engine to query external services for some intents, such as Wikipedia for general knowledge and a weather forecast provider. The dialogue engine also returned data for actions like displaying images or running animations.

Although the model could manage messages in any language, we supplied translations for predefined replies. On-the-fly replies from external sources required an additional translation service.

Data Storage. The system stored conversation texts, sensor data, system diagnostics, and other events in an InfluxDB instance. This allowed efficient time-series storage and data querying. To optimize service responsiveness, a text-to-speech cache was used to avoid repeated audio generation.

Since conversation data can be sensitive, privacy was a key focus. No cameras were used, limiting the system to motion and distance sensors. User voices were deleted after conversion to text, and any personal data in the conversation text (such as names) was automatically anonymized.

5 EXPERIMENT RESULTS

In this section, we present the results based on the data obtained during the 3 months of the experiment. Indeed, our approach to evaluate the experiment differs from the one adopted by (Velkovska, 2019), where the researchers employed a human observer of the interactions between shop clients and their robot assistant as well as feedback surveys. We, on the other hand, opted for relying on the analyses of the real usage data, while the in-person study was rather a part of the design process preceding the actual experiment.

We begin by summarizing the datasets used in the analysis (5.1). Subsequently, we delve into various aspects grouped into three areas:

- **Initial user engagement** (5.2), presenting metrics concerning the user journey to learn about the digital assistant's presence, ranging from foot traffic at the point of sales to the analysis of interactions in different languages.
- **Interaction experiences** (5.3), providing insights into conversation duration, the flow of the conversation, and technical issues.
- **Business perspective** (5.4), analyzing topics addressed and the impact of attention-grabbing pokes on sales levels.

5.1 Considered Data for Evaluation

To evaluate the effectiveness of the virtual assistant, we utilized a diverse array of datasets, as outlined below:

- **Event-based interaction logs:** Time-series records capturing customer interactions, such as dialogue exchanges, button presses, screen touches, notification displays, and survey responses. These datasets provided a detailed view of user activity and engagement with the assistant.
- **Distance sensor data:** Continuous measurements captured every 0.5 seconds by the assistant's distance sensor, reflecting customer proximity and physical engagement with the assistant.
- **Conversational system logs:** Data from the Rasa framework, detailing conversational flows, user intents, and system responses, enabling an evaluation of dialogue quality and effectiveness.
- **Foot traffic metrics:** External data capturing visitation patterns within the store and other locations, allowing for an analysis of customer behavior changes following the assistant's deployment.
- **Sales performance data:** Metrics disaggregated by product categories, used to correlate assistant interactions with purchasing behavior and variations in sales volumes.

By integrating these datasets, our analysis provides a holistic view of the virtual assistant's impact on customer experience and sales outcomes. Note that some visualizations, such as those related to foot traffic and sales levels, have been stripped from absolute values to protect sensitive business information.

5.2 User Engagement

Our analysis delves into several facets of the background and initial stages of user engagement with the virtual assistant.

Traffic Level. The analysis of daily visitor traffic aims in bringing a perspective into the impact of our experiment on the point-of-sale routines and the typical patterns of customer visits observed so far. Figure 3, illustrates the average daily footfall in the store from November 2023 onwards, segmented on a weekly basis. It is evident that the level of foot traffic has remained relatively stable before and after the

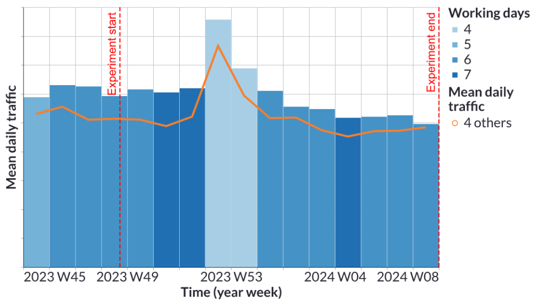

Figure 3: Traffic level in point of sale.

commencement of the VUI based digital assistant experiment in December 2023 and that its introduction did not significantly alter the store's visitation patterns.

Comparing the foot traffic trends of the analyzed store with the average of four other similar stores, represented by the orange line, reveals marginally higher values in the subject store. However, the overall behavior of the analyzed metric, characterized by fluctuations and trends, remains highly analogous across all stores. Even the peak towards the end of the year, driven by the holiday season and fewer working days in the two peak weeks, occurred in both the analyzed case and the others. This suggests that factors influencing the dynamics of the traffic level are likely consistent and irrespective of the presence of the digital assistant.

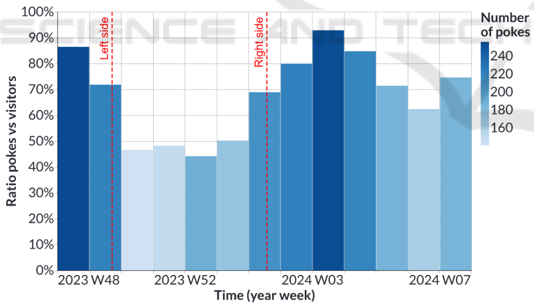

Figure 4: Level of pokes.

Level of Pokes. Figure 4 presents the number of vocally and visually attention-grabbing elements, referred to as *pokes* activated upon user motion detection. An important factor here is that the virtual assistant changed its location twice during the observation period. Initially positioned on the right side of the store, close to the mall corridor, it was then relocated to the left side, farther away from the main aisle, before returning to the right side but positioned deeper within the store.

The graph illustrates a notable variation in the percentage of displayed pokes relative to the number of visitors, depending on the virtual assistant's placement. In the locations on the right-hand side, this percentage fluctuated between 60% to over 90%, with an average of around 77%. Conversely, when situated on the left, the percentage did not exceed 50%, with an average of approximately 48%.

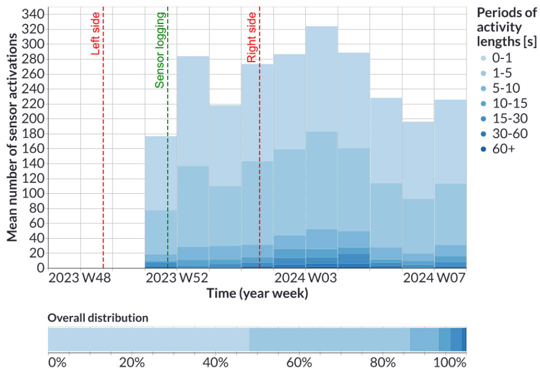

Figure 5: Activity time from sensors.

Time Spent in Front of Assistant. In order to better examine the whole journey of the client entering a shop, passing before the digital assistant and eventually engaging in an interaction or not, we focus now on the motion sensor data that started to be collected after approximately three weeks into the experiment. These data were gathered at half-second intervals, providing an approximation of the distance from the sensor and enabling an examination of the duration of time individuals spent in proximity to the virtual assistant, and depicted in Figure 5.

Our analysis considered intervals shorter than 2 seconds, which stand for 4 measurements, within a continuous sequence of measurements to be part of the same interaction. Results indicate that nearly half of the activations were momentary, that is below 1 second, indicating instances where individuals were passing by the virtual assistant without pausing. Approximately 40% of activations fell within the range of 1 to 5 seconds, suggesting momentary halts in front of the virtual assistant and potential engagement with it. Longer pauses exceeding 5 seconds occurred at a rate of approximately 25-30 per day when the virtual assistant was positioned on the left side of the store, increasing to over 40 per day in peak period when it was placed on the right side.

Started Interactions. Another step in the engagement process is the physical interaction with the robot assistant, starting from the moment a conversation is initiated by pressing the "Start conversation" button or selecting a language. The average daily count of interactions for most analyzed weeks ranged between

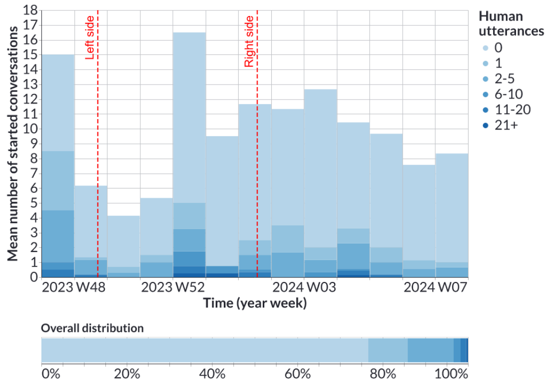

Figure 6: Level of interactions.

7 and 12 but could drop to 4 during the weakest performing week and rise to 16 during the week between Christmas and New Year's Day - see Figure 6.

Shades of blue color denote the number of utterances from the human side. As evident from the graph and the distribution of interactions with different utterance counts, approximately 76% of cases recorded no utterances from the person. Interactions with at least one utterance, referred to as *dialogues/conversations* henceforth, accounted for about 24% of cases, thus averaging 2 to 4 conversations per day. These observations further corroborate the findings from the previous section.

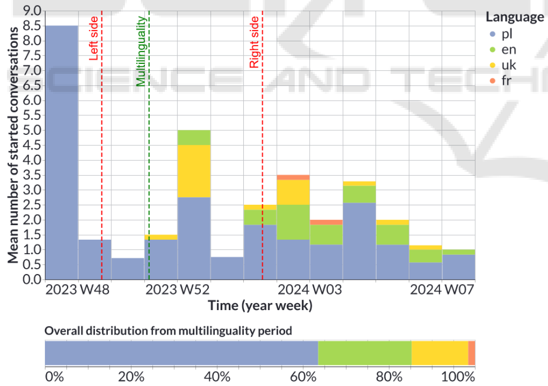

Figure 7: Conversations by languages.

Multilinguality. The multilingual support was introduced in the fourth week of the experiment and was another planned mean of encouraging users to engage in conversations in other languages than Polish. This option was presented in the graphical interface using flag icons. The chart in Figure 7 illustrates the average daily number of conversations categorized by language. Additional languages included English, Ukrainian, and French. Analysis of the data reveals that approximately 36% of conversations took place in a foreign language following the introduction of multilingual support, particularly in English and Ukrainian.

5.3 Interaction Experience

This subsection focuses on analyzing user interactions with the assistant in the context of various factors. The first paragraph presents the analysis of conversation lengths. It then discusses the frequency of interruptions of assistant utterances by users, providing insights into the dynamics of the dialogue. The subsequent paragraph analyzes the effectiveness of intent detection in user utterances, identifying areas for improvement. Furthermore, technical issues encountered during the experiment are discussed, as well as the assessment of user satisfaction with the assistant interaction.

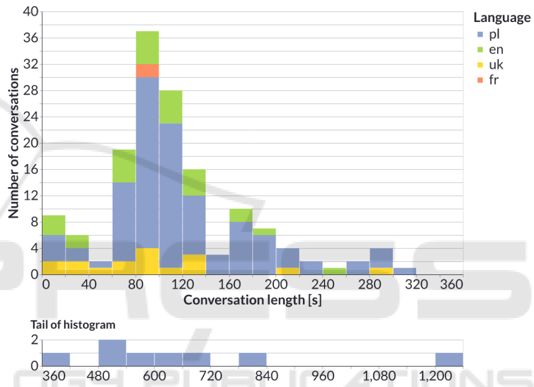

Figure 8: Conversations lengths.

Conversations Lengths. One straightforward way of quantifying the experience is related to the duration of conversations. Figure 8 presents a histogram depicting this metric, with different colors standing for the four languages introduced. Analysis of the chart indicates that the majority of conversations lasted between 60 and 140 seconds. Only a few conversations, all of which were in Polish, extended beyond 5 minutes. Within the timeframe of less than 5 minutes, there were no significant differences in conversation lengths among different languages.

Interruptions of Assistant Utterances. In order to better understand the way the people conversate with the digital assistant and the flow of the conversations, we took into consideration the timelines and the moments when people start to speak. As an interesting insight from this study, we present a chart in Figure 9 which shows how often the assistant's utterances are interrupted (by pressing the P2T button) depending on their length measured by the number of characters.

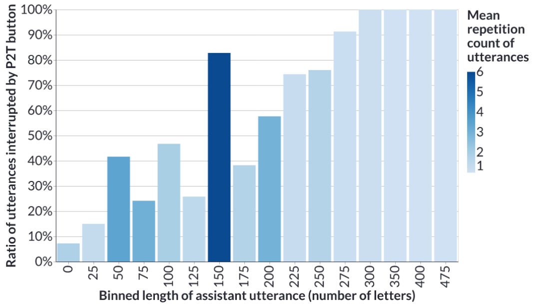

Figure 9: Interruptions of assistant utterances.

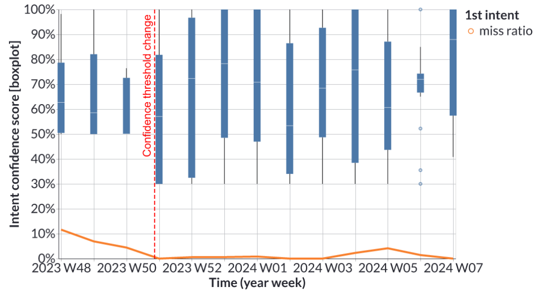

Figure 10: Intent detection confidence.

From the chart, one can see that utterances longer than 225 characters are interrupted in as much as 75% of cases, with those longer than 300 characters never being listened to fully.

The intensity of the blue color is correlated with the average number of utterance repetitions of the same length within the same conversation and provides an additional dimension to the chart. It shows that in the case of utterances repeated multiple times, the likelihood of interruption also increases. This is most evident, for example, in the case of an utterance serving as an instruction for using the P2T button, which is 155 characters long and corresponds to the dark blue bar on the chart. Its presence clearly indicates that a message repeatedly uttered during a single conversation was almost always interrupted by the user.

Human Utterances. We have also analyzed other available data points from logs generated by the assistant and subjectively evaluated the contents of the conversations themselves. Each conversation was assessed based on several criteria, including whether the utterances related to Orange-related topics, whether intents were correctly assigned, whether the speech-to-text module correctly recognized the text and whether the conversation could be considered satisfactory from both the user and the assistant perspectives.

Over the course of three months, 762 conversations were registered, of which 172 (22.6%) contained at least one non-empty user utterance, totaling 637 utterances. Topics related to offers and the store were discussed in 87 (50.6%) conversations and in 150 (23.5%) utterances. For the Polish language, the percentage of conversations related to Orange was 44.4%, while for English and Ukrainian, it was over 60% (65.7% and 61.1%, respectively), highlighting the importance of support for non-native speakers, which should be developed with particular attention.

Figure 10 depicts the confidence levels during the detection of intents in human utterances. Following three weeks of experimentation, the threshold level responsible for accepting the intent categorization was adjusted from 0.5 to 0.3. This adjustment resulted in a decrease in the miss ratio of the best-intent choice, as evaluated manually. Overall, less than 8% of utterances were annotated as misses by the conversational engine. Furthermore, in nearly half of the cases where a miss occurred, the second choice was deemed a better option. These errors often arose in situations where multiple intents related to similar topics were present in the intent database. In such instances, a more effective approach would involve querying for additional details — a functionality that warrants consideration in future iterations.

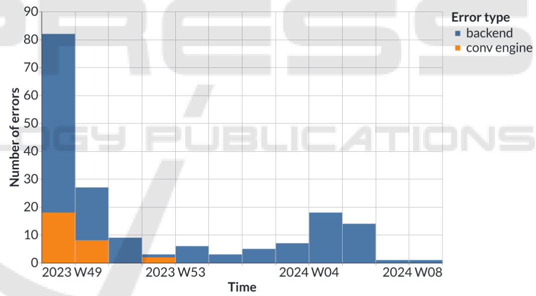

Figure 11: Number of errors in time.

Technical Problems. Throughout the duration of the experiment, we monitored errors occurring both on the conversational engine side and in the operation of the assistant itself, which might have also hindered user experience. The number of detected issues one can consult in the chart from Figure 11. Even though the first week, the number of errors was quite high, it was quickly reduced significantly to below 10 errors per week for most of the time, and in the case of the conversational engine, even eliminated entirely. Additionally, we manually annotated suspicious cases of text recognition by the speech-to-text module. Less than 4% of utterances were marked as incorrectly recognized, which is a very good result considering the assistant's environment, which can generate a lot of additional noise.

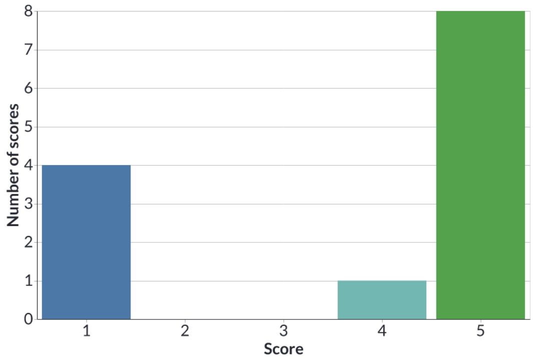

Figure 12: Survey scores.

User Satisfaction. Without access to video and without analyzing audio sound for detected emotions, assessing user satisfaction during interactions with the assistant was a challenging task. One of the evaluation mechanisms was a feedback survey displayed when the assistant detected the end of the conversation or when the user explicitly expressed a desire to provide feedback. The collected ratings, ranging from 1 to 5, are presented in the chart in Figure 12. Unfortunately, there are very few of them, only 13, with an average score of 3.6, but one can already observe a known tendency in declarative ratings hinting to more engagement in either very positive or very negative experiences (Park et al., 2018). The survey was also displayed 6 additional times with no score given.

Additionally, we conducted manual annotation of conversations, assigning ratings of 0, 0.5, and 1 based on subjective assessments of user satisfaction and the assistant's performance in adhering to project guidelines. The evaluation process focused on the conversation flow, including whether the user received substantive answers to their questions, displayed signs of satisfaction or dissatisfaction, or repeated similar questions multiple times. While repeated questions could indicate user dissatisfaction, in cases where the assistant's responses adhered strictly to the predefined conversational framework, such interactions were rated positively from the assistant's perspective. This approach ensured that all responses provided according to the designed operational schema were consistently recognized as meeting project expectations, regardless of potential user dissatisfaction. Summarizing this subjective annotation, we estimate that approximately 65% of conversations can be considered satisfactory from a human perspective, while for the assistant ratings, this coefficient was 87%.

5.4 Business Perspective

Digital assistants are increasingly being integrated into various business environments to enhance customer interactions and streamline operations. In this section, we delve into the business implications of deploying a virtual assistant in a retail setting, focusing on key insights derived from conversation topics and sales data analysis.

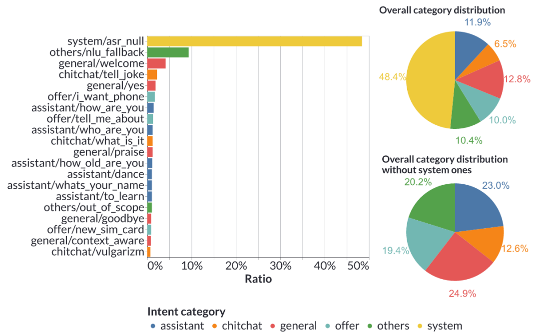

Figure 13: Most popular intents.

Most Popular Intents. Based on meaningful two-sided conversations, we analyze the main intents of user utterances from our conversational engine. Each user utterance was assigned an intent from a previously prepared set. For the purpose of analysis, intents were grouped into the following categories:

- *offer* - utterances related to the company's offers,
- *general* - general utterances such as *"yes"*, *"no"*, *"good morning"*,
- *chitchat* - utterances related to queries such as weather, Wikipedia knowledge, time, jokes,
- *assistant* - utterances related to the assistant itself,
- *others* - utterances for which it was difficult to provide an answer without access to external knowledge or due to difficulties in assigning intents.
- *system* - internal events generated by the assistant in response to user behavior, triggering specific actions within the conversational engine.

Charts in Figure 13 show statistics related to the detected intents.

The analysis of these charts reveals that nearly half of the analyzed utterances were empty. This is a significant issue stemming from the communication solution used with the robot, which required users to press and hold the button while speaking to the assistant. This solution proved to be unintuitive and difficult for shop visitors to use, even despite the assistant's messages informing about the button usage upon detecting improper usage.

A notable portion of utterances - category ”others” could have been handled if the assistant could leverage generative AI and the power of LLMs. The decision not to utilize LLMs was dictated by the design choice of complete accountability to clients for the assistant’s utterances in a point of sales setting, which, if generated by LLMs, could result in hallucinations, which were a quite common technology drawback at the time of the experiment. The remaining categories were roughly evenly distributed. It is worth noting the high percentage of utterances related to the assistant itself - users were interested in learning about it and its capabilities.

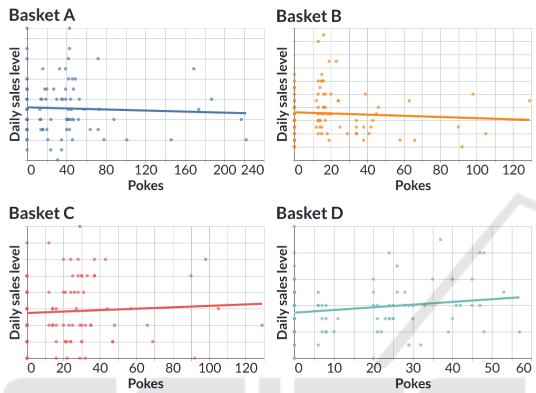

Figure 14: Impact of pokes on sales level.

Impact of Pokes on Sales Level. One of the assistant’s functionalities involved displaying promotional *pokes*. Each of these was assigned to a specific product group offered in the store. Based on information about the number of pokes and sales levels in each basket, we analyzed whether the presentation of graphical incentives accompanied by voice messages had an impact on sales. The data is visualized in Figure 14. Each point on the graph represents a single day and expresses the level of sales on the Y-axis and the number of displayed incentives on the X-axis, divided into four baskets for which we gathered the most information. Based on the analysis of these graphs, it is difficult to conclude that the information presented by the assistant influenced sales levels in any significant way.

An additional observation regarding the impact of pokes on users is the fact that the implemented ”find out more” button functionality, an invitation to follow up on the promotion that accompanied each *poke*, was not very popular, and it was used only 49 times during the whole experiment.

6 RESULTS DISCUSSION

The experiment and the analysis of the collected data yield numerous insights, which we group into three complementary perspectives, namely: technical, user and business.

6.1 Technical Improvements

Regarding the experiment described in (Velkovska, 2019) the authors point out technical difficulties as one of the main pain points preventing the users to actually benefit from the assistant’s help. Even though the technology, including AI models enhancing the audio processing in the VUIs, has significantly progressed since then, the technical context of the VUIs and conversational engines has still some margin of progress ahead.

Undoubtedly, the push-to-talk button proved to be a significant hurdle for many users. Despite graphical information (text on the button) and voice instructions provided by the assistant, the frequency of issuing these instructions indicates that such a mode of interaction was non-intuitive and challenging to master. It is difficult to imagine continuous background sound analysis as constant eavesdropping raises significant legal concerns and high costs for such a solution or triggering audio capture based on sound intensity making it difficult to filter in a noisy store environment. However, pressing the button at the beginning of a statement with automatic recognition of its conclusion and possibly enriching the graphical interface with an animation explaining how to use the voice interface could help mitigate or at least reduce the issue with P2T.

Regarding the development of the conversational engine, the analyses suggest that integration with generative AI language models could aid both in handling topics beyond the virtual assistant’s domain and in diversifying the conversation. Data indicates that repetitive topics within the same conversation were often interrupted by the speakers. Analyses also revealed that people tended to interrupt long statements during the experiment, so this aspect should also be considered when designing the assistant’s manner of speaking. Perhaps it is worth considering it as a factor that could be subject to personalization depending on the interlocutor’s expectations. Ensuring conversational context is crucial. People naturally assume they can refer back to previous statements, which is not always easy to handle with intent-based conversational solutions. Here, Large Language Models capable of handling long contexts could come to the rescue. It is also important to adapt the technology used in the speech-to-text module so that it can contextually correct errors from the assistant business domain in translated statements.

The operation of the conversational engine based on the detection of each statement's intent revealed some problems. Firstly, the system had difficulties in categorizing intentions that were thematically close to each other. There were cases where two intents received a high confidence score, but neither exceeded the required threshold, and the assistant reacted as if it did not understand. To improve this aspect in the future, attention should be paid to intents that may be confused with each other or introduce functionality to inquire about details in case of uncertainty regarding categorization. Such grounding aspect could increase the naturalness of the assistant's conversation.

Additionally, it's not uncommon for user utterances to contain multiple intents, indicating a complexity that the assistant must navigate through effectively. The assistant was not prepared for this and reacted in its statement only to one of them, leaving the other unhandled or forcing repetition from the speaker. Improving keywords and named entities recognition is crucial, as evidenced by instances where the speech recognition module was finding the closest Polish words to company names of English origin. Enhanced accuracy in understanding user inputs is essential for a seamless conversational experience. Moreover, topics related to the context in which the digital assistant is working that are not yet explicitly addressed should be quickly identified based on conversations analysis, maybe even automatically, and where possible, handling of relevant threads should be continuously added. Such a proactive approach can contribute to enhancing user satisfaction and efficiency.

Furthermore, people tend to address digital assistants differently than human-locutors and frequently resort to using abbreviated forms rather than full phrases, opting for expressions like "*Orange offer*", "*Registration*", or "*My Orange application*" instead of "*could you tell me about ...*". This suggests a tendency towards brevity and efficiency in communication. Technical inquiries also feature prominently in user utterances, with individuals seeking details such as microphone positioning or the purpose of the P2T button. This underscores the importance of providing clear explanations and transparent functionality. Volume control emerges as a recurring topic of interest, reflecting users' desire for customization and control over their interaction experience.

6.2 Value for the User

The virtual assistant was designed to offer several valuable benefits to users, in order to enhance their overall experience and interaction satisfaction. One notable advantage is its multilingual capability, enabling users to conduct conversations and resolve issues in their preferred language. Analysis of user interactions revealed that individuals who spoke languages other than Polish more often raised inquiries related to Orange services, highlighting the importance of multilingual support in increasing service accessibility and inclusivity.

Furthermore, the virtual assistant extends beyond business domain matters to address broader topics and provide entertainment as a factor complementary to the straight-to-the-business approach. Users frequently engage in casual conversations, requesting jokes or even asking the assistant to dance. This additional functionality serves as a source of amusement, particularly for younger users, and helps alleviate waiting times in queues, enhancing the overall customer experience. The entertainment-enjoyment dimension plays a significant role here, emphasizing the importance of incorporating engaging and enjoyable elements into the assistant's repertoire, provided they align with its context and operational role in a given physical space. However, this entertainment factor has also been reported as a potential distractor in previous experiments ((Velkovska, 2019)) and should therefore be dosed with caution.

Moreover, the potential for personalization represents a significant value proposition for users. As the technology evolves, users may have the opportunity to customize the assistant's speech patterns or even enable automatic adaptation to their continuously detected needs. This level of personalization fosters a sense of tailored service and enhances user satisfaction, very critical to feel special rather than just as another customer a company puts an insensitive bot in front of. This is one of the aspects to experiment with in the future.

Additionally, the virtual assistant has potential to facilitate user feedback, allowing clients to express their opinions on the interaction experience but also on the point of sales service as a whole. This feedback loop not only enables users to voice their concerns but also can also provide valuable insights for service improvement efforts, ultimately leading to enhanced service quality and user engagement.

Another key benefit is the graphical presentation of conversation aspects, enabling users to better understand and visualize certain details discussed during the interaction, such as the option to view and compare available colors of a specific model of phone casing. This visual representation empowers users to make more informed decisions, enhancing their overall comprehension and decision-making process.

6.3 Business Perspective

The deployment of virtual assistants in retail environments holds significant implications for business operations and customer service strategies. Despite their potential, several factors influence their effectiveness and integration into existing business frameworks.

One notable observation is the limited impact of virtual assistants on in-store foot traffic. Several possible reasons contribute to this phenomenon, including the assistant blending into the surrounding environment, lack of promotional activities to raise awareness of its presence, and pre-planned visits to the store. Understanding these dynamics is crucial for optimizing the deployment of virtual assistants and maximizing their influence on customer engagement.

The placement of virtual assistants within the retail environment plays a pivotal role in their effectiveness. In this experiment, some individuals mistook the assistant for a ticket machine, highlighting the importance of clear signage and intuitive user interfaces to attract attention and convey the assistant's function effectively. Strategic placement should be considered as part of the overall customer experience design, ensuring alignment with the assistant's intended role and functionalities.

For virtual assistants to effectively engage customers and prolong interaction duration, they must be capable of handling end-to-end use cases seamlessly. Frequent redirections to in-store personnel due to informational limitations hindered the depth of interactions and led to predominantly short and one-sided exchanges.

Multilingual support emerges as a critical aspect for reaching a wider audience. A significant proportion of conversations, particularly among non-Polish speakers, revolved around operational topics related to the store's offerings. By catering to diverse linguistic needs, virtual assistants can enhance accessibility and ensure effective communication with a broader customer base.

Moreover, the scalability of virtual assistants across multiple locations offers a consistent and reliable customer service experience. While initial implementation costs may be high, the potential for long-term cost savings in more repetitive in-store information, or sales routines as well as through value generation when extending work hours of human staff, underscores the value proposition of virtual assistants in retail settings.

Integration with existing systems and data collection mechanisms is paramount for maximizing the utility of virtual assistants. Long-term data collection provides valuable insights into customer preferences and behaviors, enabling data-driven decision-making to enhance sales effectiveness and customer satisfaction.

7 CONCLUSIONS

This study provides insights into the implementation of digital assistants in retail environments, focusing on their impact on customer engagement and problem resolution, as well as the challenges and opportunities of VUI technologies. After analyzing the collected data, we return to answer the research questions we formulated at the beginning of this paper (Section 1.)

RQ1: The digital assistant demonstrated its potential to enhance customer engagement, particularly through multilingual support and interactive features that increased accessibility and inclusivity for non-native speakers. However, capturing customer attention in a retail environment crowded with digital screens proved challenging, limiting the assistant's ability to sustain interest. Although it effectively distributed marketing information, its limited impact on problem resolution and sales often necessitated redirection to human staff. Enhancing user engagement and addressing these limitations will be essential for realizing the full potential of digital assistants in retail.

RQ2: The study identified key challenges and opportunities in deploying VUI technologies. Challenges included an unintuitive push-to-talk interface that hindered usability and technical issues, such as inadequate intent categorization and frequent interruptions during longer responses. These limitations underscore the need for more intuitive interaction mechanisms and advanced technical capabilities. On the other hand, opportunities lie in personalization to tailor interactions to individual users and integrating technologies like large language models (LLMs) to improve adaptability and conversational depth. Real-time adaptability, driven by continuous data analysis and feedback, is also crucial to maintaining relevance and satisfaction.

To fully leverage digital assistants in retail, future efforts should focus on improving user interface design, aligning it with client needs and complete customer journeys. Personalization and advanced technology integration, combined with naturalized interaction styles, can transform digital assistants from marketing tools into essential elements of customer relationships and operational success.

ACKNOWLEDGEMENTS

Emilia Lesiak is responsible for writing the first part of the manuscript, i.e. the introduction, state of the art, experiment setup, and conclusions. Grzegorz Wolny focused on the sections concerning experiment results and data analysis, their discussion and contributed to the conclusions. Bartosz Przybył provided technical details regarding the experiment setup. Michał Szczerbak shaped the whole manuscript, the objectives of the experiment, and conclusions.

The experiment described in this paper was financed by the Orange Innovation Research and B2C Customer Journey departments in Orange and was technically prepared by the AI Skills Center department in Orange Innovation Poland under Paweł Tuszyński's lead. The authors wish to thank the following designers, developers, translators, and researchers who also contributed operationally to the experiment (in alphabetical order): Artur Bajll, Damian Boniecki, Mikołaj Doliński, Damian Fastowiec, Piotr Gołabek, Maciej Jonczyk, Robert Kołodyński, Adam Konarski, Łukasz Krajewski, Anna Kraśkiewicz, Izabella Krzemińska, Valeriia Majcher, Tomasz Michalik, Filip Olechowski, Przemysław Pietrak, Robert Warzocha, Wojciech Zieliński, and Katarzyna Żubrowska.

Writing of this manuscript was supported by the use of Large Language Models for editorial-stylistic corrections only.

REFERENCES

Amershi, S., Weld, D., Vorvoreanu, M., Fourney, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., et al. (2019). Guidelines for human-ai interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems*, pages 1–13.

Balakrishnan, J. and Dwivedi, Y. K. (2024). Conversational commerce: entering the next stage of ai-powered digital assistants. *Annals of Operations Research*, 333(2):653–687.

Breazeal, C. (2004). *Designing sociable robots*. MIT press.

Brill, T. M., Munoz, L., and Miller, R. J. (2022). Siri, alexa, and other digital assistants: a study of customer satisfaction with artificial intelligence applications. pages 35–70.

Cassell, J., Sullivan, J. W., Prevost, S., and Churchill, E. F. (2000). Embodied conversational agents. *Social Psychology*.

López, G., Quesada, L., and Guerrero, L. A. (2018). Alexa vs. siri vs. cortana vs. google assistant: a comparison of speech-based natural user interfaces. In *Advances in Human Factors and Systems Interaction: Proceedings of the AHFE 2017 International Conference on Human Factors and Systems Interaction, July 17- 21, 2017, The Westin Bonaventure Hotel, Los Angeles, California, USA 8*, pages 241–250. Springer.

Luger, E. and Sellen, A. (2016). "like having a really bad pa" the gulf between user expectation and experience of conversational agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 5286–5297.

McTear, M. F., Callejas, Z., and Griol, D. (2016). *The conversational interface*, volume 6. Springer.

Mik, E. (2019). Smart contracts: A requiem. *Journal of Contract Law, Forthcoming*.

Morgan, D. P. and Balentine, B. (1999). How to build a speech recognition application: A style guide for telephony dialogues. Enterprise Integration Group San Ramon.

Oviatt, S., Schuller, B., Cohen, P. R., Sonntag, D., Potamianos, G., and Krüger, A. (2018). *The Handbook of Multimodal-Multisensor Interfaces: Signal Processing, Architectures, and Detection of Emotion and Cognition-Volume 2*. Association for Computing Machinery and Morgan & Claypool.

Park, K., Chax, M., and Rhim, E. (2018). Positivity bias in customer satisfaction ratings. In *Companion Proceedings of the The Web Conference*.

Van Kleef, E. and van Trijp, H. C. (2018). Methodological challenges of research in nudging. In *Methods in Consumer Research, Volume 1*, pages 329–349. Woodhead Publishing.

Velkovska, J. (2019). When an emotional robot meets real customers exploring hri in a customer relationship setting. *Proceedings of Mensch und Computer*, pages 350–352.

Wilpon, J. G., Rabiner, L. R., Lee, C., and Goldman, E. (1990). Automatic recognition of keywords in unconstrained speech using hidden markov models. In *IEEE Transactions on Acoustics, Speech, and Signal Processing*.

Xie, B., He, D., Mercer, T., Wang, Y., Wu, D., Fleischmann, K. R., Zhang, Y., Yoder, L. H., Stephens, K. K., Mackert, M., et al. (2020). Global health crises are also information crises: A call to action. *Journal of the Association for Information Science and Technology*, 71(12):1419–1423.

Xu, C., Li, Z., Zhang, H., Rathore, A. S., Li, H., Song, C., Wang, K., and Xu, W. (2019). Waveear: Exploring a mmwave-based noise-resistant speech sensing for voice-user interface. In *Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services*, pages 14–26.