EVALUATION OF CHATBOT TECHNOLOGY: THE CASE OF GREECE

Theodoros Papadopoulos[1], Zoi Lachana[1], Thanos Anagnou[1], Charalampos Alexopoulos[1], Yannis Charalabidis[1], Christos Bouras[2], Nikos Karacapilidis[2], Vasileios Kokkinos[2], and Apostolos Gkamas[2]

[1]Department of Information and Communication Systems Engineering, University of the Aegean, Samos, Greece
[2]Department of Computer Engineering and Informatics, University of Patras, Patra, Greece

ABSTRACT

In recent years, the field of Artificial Intelligence has made significant strides, particularly in advancing chatbots through Natural Language Processing (NLP) technology. Recently, however, there has been intense competition in this industry, with major tech companies consistently introducing new and improved solutions. Nevertheless, the Greek context introduces several unique challenges and obstacles to the adoption of modern solutions, owing to the distinctiveness and relative rarity of the Greek language, as well as the limited financial resources of the Greek economy. The objective of this study is to assess the performance of chatbots in terms of the quality of their responses, including relevance, naturalness, coherence, accuracy, and vocabulary, and to gauge user experience and satisfaction. An additional objective is to acquire a comprehensive comparative overview of chatbot performance in Greece, both on a per-question basis and when comparing related questions. The chosen method for evaluation is a guided interview employing closed-type questions. The intention is to gather structured and quantified data in an area where the typical internet user may not possess extensive familiarity or prior experience with such evaluations. Conclusions were drawn for each question, enabling a focused and comparative assessment of solution quality, identification of potential trends, and verification of response consistency.

KEYWORDS

Artificial Intelligence, Natural Language Processing, Virtual Assistants, Greek Virtual Assistants.

1. INTRODUCTION

In recent years, there has been a significant worldwide increase in the adoption of chatbots, driven by the realization among businesses and organizations of their capacity to improve operational efficiency and customer satisfaction. A pivotal factor contributing to the effectiveness of chatbots lies in their capability to comprehend and interact with customer inquiries in a manner that feels natural and akin to human communication. Achieving this necessitates the use of advanced technologies and a profound grasp of the language and communication conventions employed by the intended audience [1].

One of the most important techniques used is Natural Language Processing (NLP). Natural Language Processing employs methods both for Natural Language Understanding (NLU) and Natural Language Generation (NLG), which allows simulating of the human ability to understand and create natural language text, e.g., to summarize information or infer topics from documents. Modern NLP techniques rely mostly on machine learning to derive meaning from human

DOI: 10.5121/ijnlc.2023.12502

languages, using statistical inference to automatically learn rules through the analysis of large corpora of text [2].

1.1. Greek Chatbots

Despite their widespread acceptance on a global scale, the presence of chatbots in the Greek language remains somewhat limited. Nevertheless, there is a burgeoning interest in deploying chatbots supporting the Greek language, particularly within the realms of customer service and government sectors. In Greece, certain businesses have embraced chatbots as a means to offer round-the-clock customer support, enhance operational efficiency and reduce reliance on human resources. The outcomes of these implementations have been diverse; with some organisations reporting notable enhancements in customer satisfaction and efficiency [3], while others encountering challenges in persuading their clients to embrace this technology.

The development of chatbots for the Greek language presents several technological hurdles [4]. The absence of linguistic standardization, coupled with the ancient, complex, and multifaceted nature of Greek along with its many dialects and regional variations, contributes to the intricacies of natural language processing (NLP) algorithms. This intricacy makes it challenging for chatbots to consistently comprehend and respond to customer inquiries. Additionally, the limited availability of high-quality training datasets hinders the capacity of chatbots to provide precise and effective responses. Nonetheless, initiatives are underway to promote the use of chatbots in the Greek language, including projects such as the Pythia project [5] and the creation of language models tailored specifically for Greek, such as Greek-BERT.

Greek-BERT is a pre-trained language model for Greek based on the BERT architecture (Koutsikakis, Chalkidis, Malakasiotis, & Androutsopoulos, 2020) [6]. It has been developed by the Natural Language Processing (NLP) group of the Athens University of Economics and Business (AUEB) and it is based on the architecture presented in the paper "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding" (Devlin, Chang, Lee, & Toutanova, 2019) [7]. The model consists of an encoder and a decoder, with the encoder being modified to allow bidirectional learning. To understand the meaning and significance of Greek words and sentences, it was trained on a set of over 13.6 billion discriminative text tokens, including Wikipedia news articles and social media posts. This allows it to capture the nuances and complexities of the Greek language.

1.2. Research Focus and Objectives

Chatbots are a modern tool that enable communication between humans and machines. They mainly serve to offer replies to user queries and engage in human-like conversations. However, the accuracy and effectiveness of their responses depend on the precision of the algorithms used during their training, the precise training datasets, and the dedication of their developers to continuously maintain and improve them.

This research endeavours to evaluate the quality and performance of chatbots through subjective assessments by individuals. These chatbots are employed by companies and organizations with a primary focus on serving the Greek population.

The objectives of this study encompass an evaluation of chatbots' performance with regard to the relevance, naturalness, coherence, accuracy, and vocabulary of their responses. Furthermore, the research seeks to gauge the user experience and overall satisfaction with the functionality of these chatbots. An additional aim is to gain a comprehensive comparative perspective on how chatbots operate in Greece, both on a question-by-question basis and with similar queries.

2. BACKGROUND

2.1. Chatbots History

The concept of a chatbot can be traced back to the mid-20th century. In 1950, British mathematician and computer scientist Alan Turing introduced the idea of a "machine" that could mimic human conversation in his famous "Turing Test. The first known chatbot was ELIZA, developed in 1966, whose purpose was to act as a psychotherapist returning the user utterances in a question form [8]. ELIZA operated on a set of pattern-matching rules and could simulate a psychotherapist, engaging users in conversations about their feelings. It was a simple yet impactful demonstration of natural language processing.

In the 1980s and 90s, the development of chatbots continued, with advancements in rule-based systems and scripting languages. These early chatbots were primarily used for customer support and were frequently employed in call centres. Yet, their capabilities remained quite limited, often frustrating users with their inability to understand complex queries or engage in meaningful conversations. The turn of the 21st century brought renewed interest in chatbots with the rise of the internet and the exponential growth of data. More sophisticated AI and natural language processing techniques began to emerge. Smarter, more versatile chatbots were developed, often integrated into websites and instant messaging platforms In 1995, the chatbot ALICE was developed which won the Loebner Prize, an annual Turing Test, in years 2000, 2001, and 2004. ALICE relies on a simple pattern-matching algorithm with the underlying intelligence based on the Artificial Intelligence Markup Language (AIML), which makes it possible for developers to define the building blocks of the chatbot knowledge [9].

The next step was the creation of virtual personal assistants like Apple Siri, Microsoft Cortana, Amazon Alexa, Google Assistant and IBM Watson. Those chatbots represented a paradigm shift in the chatbot landscape combining features such as speech recognition with natural language understanding, and integration with end-user devices such as smartphones and speakers. The incorporation of chatbots into messaging platforms and social media further expanded their reach. Platforms like Facebook Messenger allowed businesses to create chatbots for customer interactions, marketing, and sales. Companies started using chatbots to automate routine inquiries, reducing response times and improving customer satisfaction. These bots were particularly valuable in e-commerce, helping customers track orders, answer frequently asked questions, and even recommend products.

During the current decade, the combination of machine learning, deep learning, and natural language processing technologies has led to the creation of highly sophisticated chatbots capable of handling complex and nuanced conversations which are being employed in diverse sectors, including healthcare, education, and finance, offering new possibilities for automation and personalized services.

2.2. Chatbots Evaluation

The evaluation of chatbots plays a pivotal role in their effective development. This evaluation encompasses the assessment of their quality and performance, with a particular focus on their ability to comprehend user queries and generate appropriate responses in the form of natural and engaging conversations. The overarching goal is to ensure that chatbots effectively meet user requirements, efficiently fulfil their intended functions, and deliver a positive user experience [10]. The complexity of human communication itself introduces inherent challenges to this evaluation process.

Typically, evaluation models encompass a variety of tasks designed to assess different facets ofa chatbot's performance, including tasks such as recognizing user intent, extracting entities, and generating responses [11]. These models can be applied to different datasets and objectives, providing a comprehensive perspective on the overall quality of the chatbot. Datasets are of paramount importance for training and testing chatbot models, as well as for evaluating their performance. These datasets must exhibit diversity, accurately represent the language and context of the target users, and include relevant annotations such as intent, entities, and conversational sequences.

Evaluation metrics encompass both objective and subjective criteria [12]. Objective criteria involve metrics such as perplexity, accuracy, and the F1 score, which rely on quantitative analysis of the chatbot's performance in a particular task or dataset. In contrast, subjective criteria depend on human assessment of its performance, which can be conducted through methods such as surveys, interviews, or user studies. Among these metrics are:

Perplexity: is a frequently employed metric for assessing the performance of language models. It quantifies the uncertainty associated with predicting the next word in a sentence, with superior performance indicated by lower perplexity values [13]. This metric is computed by taking the reciprocal of the geometric mean of the probabilities assigned to each word within the sentence. For instance, if a chatbot assigns a probability of 0.8 to the word "hello" and a probability of 0.2to the word "world" in the sentence "hello world", the perplexity score would be 1/√(0,8 * 0,2) which equals 2.5. Lower scores can result in more coherent and natural responses. While perplexity is a useful metric for assessing language models, it has its limitations, as it doesn't directly measure the quality of generated text in terms of fluency, coherence, or human-like understanding. Therefore, it is often used in conjunction with human evaluations for text-generation tasks.

Accuracy: represents another commonly utilized metric for assessing the overall effectiveness of chatbots [14]. It is determined by dividing the count of accurate responses by the total number of responses. For instance, if a chatbot responds correctly to 80 out of 100 user inputs, its accuracy score would be 80%. Chatbots with higher accuracy scores can execute their designated tasks more effectively.

F1 Score: This metric is a common tool for evaluating the performance of chatbots in natural language processing tasks, such as question answering and sentiment analysis. It quantifies the trade-off between two evaluation metrics: precision and recall [15]. Precision signifies the percentage of correctly identified positive instances, essentially the proportion of accurate predictions. Conversely, recall represents the percentage of actual positive instances that the algorithm correctly recognizes, indicating the proportion of positive cases identified accurately. Balancing precision and recall underscores the challenge of achieving high values for both metrics simultaneously. When we raise the threshold for predicting positive outcomes, precision increases but recall decreases. Conversely, lowering the threshold boosts recall while reducing precision. This phenomenon occurs because increasing the threshold results in more correct predictions but fewer total predictions made by the algorithm. Conversely, reducing the threshold increases the total number of predictions but negatively impacts prediction accuracy. The F1 score is particularly valuable for evaluating the chatbot's ability to handle intricate conversations and understand the core content of these interactions.

Human evaluation: Evaluating chatbots is a crucial process that involves the assessment of their quality and performance through human subjective judgment. This evaluation primarily focuses on the quality of the chatbot's responses, considering factors like relevance, naturalness, and coherence. Additionally, it examines the overall user experience and satisfaction derived from the chatbot's functionality [16]. Various methods can be employed for this evaluation, including surveys, interviews, user studies, and expert assessments. Surveys and interviews typically gather user feedback on their interactions with the chatbot, covering aspects such as satisfaction levels, perceived usefulness, and overall performance [17]. User studies involve the observation of users engaging with the chatbot in a controlled environment, collecting data on their behaviour and feedback. These studies may involve a randomly selected group of users with diverse backgrounds and experiences or focus on expert groups with domain-specific knowledge. Several compelling reasons underscore the significance of this evaluation process. Firstly, it provides a quantifiable measure of the chatbot's performance, particularly its ability to attract users and offera positive user experience. Given that chatbots are designed to engage in natural and appealing interactions, the quality of the user experience is paramount to their success. Secondly, human evaluation can unveil both the strengths and weaknesses of a chatbot, pinpointing areas that require improvement. This feedback is invaluable to developers seeking to enhance their chatbots' performance and better cater to user needs.

Nevertheless, human evaluation does come with its own set of challenges. It can be resource-intensive and costly, particularly when a substantial number of evaluators are required. Moreover,it may introduce subjectivity influenced by factors such as individual preferences and biases. Lastly, it may not always accurately reflect the chatbot's real-world performance, as user interactions can vary in different settings or contexts. To address these challenges, researchers employ diverse techniques, including focused interviews, focus groups, crowdsourcing, and evaluations grounded in machine learning.

3. METHODOLOGY

The selected approach for evaluation involved guided interviews using closed-ended questionnaires. This method is widely recognized as an effective assessment technique for interactive systems, falling under the domains of both usability testing and predictive evaluation reviews [18]. The objective is to acquire structured and quantifiable data in an area where the typical internet user may lack familiarity and prior evaluation experience.

Closed-ended questionnaires represent a popular research tool employed within guided interviews to systematically collect data from participants [19]. This method offers several advantages that enhance its value in research. Firstly, it enables efficient data gathering. By employing a standardized set of questions, researchers can swiftly and effortlessly collect data from a multitude of participants, which proves especially beneficial in research scenarios where time and resources are constrained. Secondly, the resulting data can be easily analysed. Closed-ended questions yield quantitative data that can be readily categorized and statistically assessed, allowing for the swift identification of patterns and trends. Thirdly, closed-ended questions minimize bias in the data collection process. They eliminate the need for participants to elaborate on their responses, thus reducing the potential for misrepresentation. Lastly, this approach can be employed to gather data from a diverse range of participants, regardless of their backgrounds, experiences, or expertise.

3.1. Selected Questions

The subsequent questions were chosen based on critical criteria identified in international literature, as well as in consultation with the supervising professors.

Solution Type: *Voice vs. Text*: Chatbots can take two primary forms, either voice-based or text-based, each possessing distinct characteristics and advantages. Voice-based chatbots leverage Natural Language Processing (NLP), enabling more natural and user-friendly interactions. Users can engage with the chatbot using their voice, resembling conversations with humans, thereby enhancing engagement and enjoyment. Conversely, text-based chatbots offer the advantage of widespread accessibility across diverse devices, including smartphones, computers, and tablets. Users can interact with the chatbot by typing messages, a familiar and convenient mode of communication.

Economic Sector: *Telecommunications, Finance, Commerce, Government, Non-profit Organizations, Municipal Administration, Education*. Chatbot systems find extensive application across a range of economic sectors, supporting diverse facets of business operations. These systems are crafted to streamline customer interactions, elevate customer contentment, and generate cost efficiencies for enterprises.

Exclusive Mode of Communication: *Yes vs. No*. The concept of employing chatbot systems as the exclusive interface between businesses and their customers is a relatively recent development, yet it has garnered increasing attention in recent years. This strategy entails placing sole reliance on chatbots for customer service, as opposed to providing multiple channels like email or telephone support. The merits of this approach encompass cost reduction, round-the-clock accessibility, and enhanced operational efficiency. However, the drawbacks include limited comprehension, the absence of a human touch, and potential technical challenges.

User Experience with the Interface: *Excellent, Good, Average, Poor*. User Experience (UX) refers to the holistic encounter a user undergoes when using a product or service. It encompasses elements like user-friendliness, response time, comprehensibility of content, and the receptiveness of menus, among others.

User Experience with the Outcome: *Excellent, Good, Average, Poor*. Users are requested to assess whether they successfully accomplished their intended objective. This metric holds significant importance as it ultimately decides whether the customer attains their desired outcome or decides to discontinue their endeavour.

Overall Performance: *Excellent, Good, Average, Poor*. The user's holistic viewpoint of the performance assesses the impression left by the entire experience. Regardless of whether individual scores were lower or higher, the key question is whether the user's needs were effectively addressed according to their intended scenario.

3.2. Interview Methodology

The interviews, serving as a central element of the research, were carried out consistently for all participants. A neutral and unadorned location was chosen to maintain a standard environment, equipped with essential amenities. This space included a modern computer featuring a sizable 65-inch high-resolution (4K) display and high-speed internet access (1Gbps). The interviewer made diligent efforts to remain unbiased and refrained from suggesting any answers to the participants. Prior preparations included saving websites and phone numbers to expedite connections and minimize interruptions.

Each interview had a designated duration of 2 hours to prevent excessive fatigue or irritation, which could potentially influence the responses emotionally. The allocation of time was as follows: 20 minutes for the introduction, 45 minutes for solution analysis, a 10-minute break, and another 45 minutes for solution analysis.

The interviews were conducted based on the following scenario: Entry of the interviewee and familiarization with the space and equipment. Collection of the participant's profile data: *Gender, Age, Education level, whether they use a computer, and if they own a smartphone*. Demonstration and testing of ChatGPT as a benchmark for other solutions. Review of the text evaluating the questions. Commencement of evaluation.

3.3. Presentation of the Interview Approach

The participant selection process aimed to mirror, to the greatest extent possible, potential Greek users who might opt for chatbots as a means of communication rather than opting out. Consequently, individuals who lacked internet familiarity or held a negative view of technology were not included as candidates. Additional factors considered for participant selection encompassed their availability and willingness to partake, given the necessary time commitment of 120 minutes. The group's characteristics are detailed in the table below:

Table 1: Characteristics

No	Gender	Age	Educationallevel	Computer user	Smartphoneowner
1	F	42	University	Yes, often	Yes
2	M	48	University	Yes, often	Yes
3	M	50	Elementary	No	Yes
4	M	52	High school	No	Yes
5	F	80	Postgraduate	Yes, often	Yes
6	F	51	Postgraduate	Yes, often	Yes
7	F	17	High school	Yes, often	Yes
8	M	23	University	Yes, often	Yes
9	M	37	Postgraduate	Yes, often	Yes
10	M	29	University	Yes, often	Yes

4. RESULTS

The table provided below showcases the solutions under examination, distinguishing between those in the private and public sectors, along with their associated economic sectors:

Table 2: Chatbot Solutions

Company – Organization	Solution type	ApplicationDomain	Economic sector
2103288000 - Piraeus	Voice	Private	Financial
Winbank - Piraeus	Text	Private	Financial
Vodafone.gr - tobi	Text	Private	Telecommunications
13888 - Cosmote	Voice	Private	Telecommunications
Alpha Bank	Voice	Private	Financial
Eurobank	Voice	Private	Financial
National bank of Greece	Voice	Private	Financial
Attika bank	Text	Private	Financial

Hellenic Development Bank	Text	Private	Financial
Leroy merlin	Text	Private	Commercial
ikea	Text	Private	Commercial
eco-mat	Text	Private	Commercial
pennie	Text	Private	Commercial
ledison	Text	Private	Commercial
xtr	Text	Private	Commercial
acs	Text	Private	Commercial
coca-cola	Text	Private	Commercial
goldmall	Text	Private	Commercial
Market4you	Text	Private	Commercial
ReBrain Greece	Text	Public	State entities
oasa	Text	Public	Public utility
deddie	Text	Public	Public utility
dei	Text	Public	Public utility
eydap	Text	Public	Public utility
eopyy	Text	Public	State entities
dypa	Text	Public	State entities
Region of Attika	Text	Public	State entities
Region of Stereas Elladas	Text	Public	State entities
Municipality of Papagou-Hollargou	Text	Public	Municipalities
Municipality of Kalamaria	Text	Public	Municipalities
Municipality of Patmos	Text	Public	Municipalities
Municipality of Moschato-Tavros	Text	Public	Municipalities
Municipality of Filis	Text	Public	Municipalities
Municipality of Kastellorizo	Text	Public	Municipalities
Municipality of West Lesvos	Text	Public	Municipalities
Municipality of Platanias	Text	Public	Municipalities
Municipality of Agia	Text	Public	Municipalities
Municipality of Visaltia	Text	Public	Municipalities
Elecectrical & Computer, Engineering Dept - UOP	Text	Public	Education
University of West Attika	Text	Public	Education

Regarding the open-ended questions, we have presented the most provided response by the interviewees, which we consider to be indicative of the prevailing opinion among participants.

Conclusions will be drawn individually for each question to assess the performance of the solutions in a targeted and comparative manner, to identify potential patterns or trends.

4.1. Presentation of Interview Methodology Analysis

The evaluation of functionality and user experience was conducted based on the responses provided by the study participants.

Sector of economy: The country's economic sector assumes a crucial role in this context. The economic background of each entity that adopts a chatbot solution provides valuable information regarding the prevalence of these solutions within the country's economy. Nevertheless, when considering the proportion relative to the number of sites searched to locate them, they representa rather modest fraction, approximately 15%.

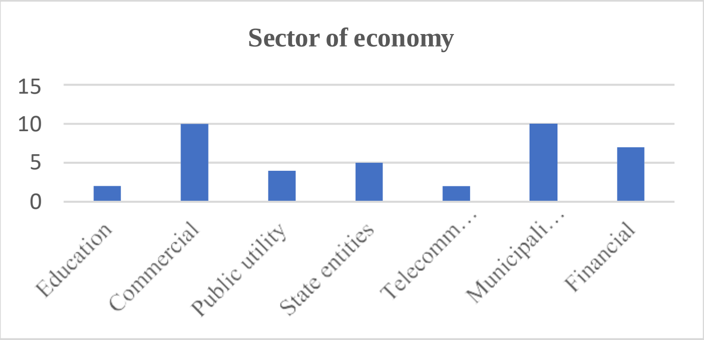

Figure 1: Sector of Economy

Most implementations (10) were discovered within the commercial sector, which is not surprising given the proliferation of online shops in the post-Covid era. An unexpected finding was the presence of 10 implementations in the local government sector, which appear to be integrated intoa pre-existing software package. Other sectors such as financial services, telecommunications, and public utilities also exhibited a notable presence relative to their smaller numbers. In contrast, the education sector and central government had lower percentages of implementations.

Exclusive Communication Channel: This aspect serves as a significant indicator of the extent to which a company or organization has relied on chatbots for customer service, reflecting their level of investment in the development of this technology.

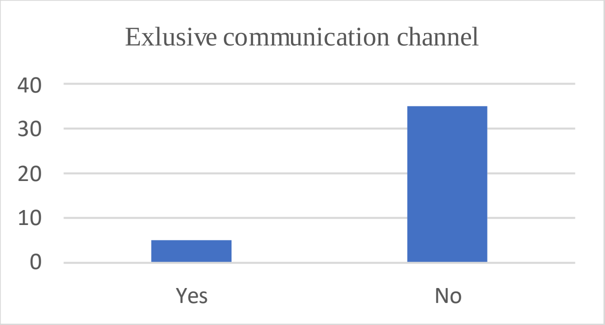

Figure 2: Exclusive Communication Channel

In most cases (35 out of 40), the implementations primarily function as supplementary tools. Consequently, most of these implementations have not made significant investments in harnessing the full potential of chatbots. An exception exists with large organizations that have meticulously evaluated the potential advantages this communication protocol can deliver to them.

User Experience with the Interface. Irrespective of the sophistication of chatbot technology, if the interface through which users interact is not user-friendly, swift, and clear, it becomes difficult to assess it positively.

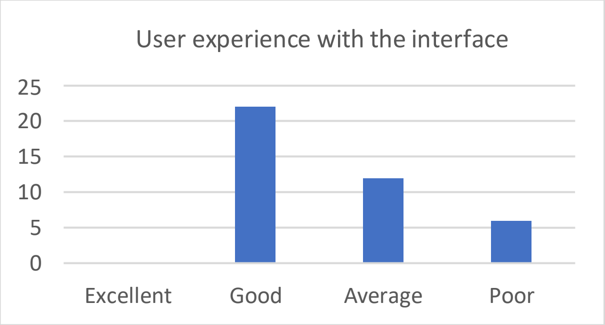

Figure 3: User Experience with the Interface

Our findings indicate that in the majority of instances (34 out of 40), the reception to the method of interacting with the chatbot is deemed acceptable, falling within the range of good to average. The average user's familiarity with messaging applications plays a role in comprehending its operation. However, the neutral tone of the responses and the absence of intuitive elements have hindered exceptional evaluations.

User Experience with the Outcome. The primary user expectation when interacting with a chatbot is to receive comprehensive assistance with minimal effort. Additionally, the nature and significance of the matters users seek to address through the application also exert a substantial influence. For example, the level of service quality expected from a bank or telecommunications provider differs from that anticipated from a retail store or municipality.

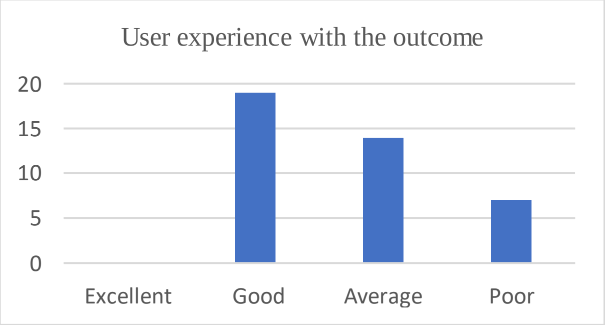

Figure 4: User Experience with the Outcome

Based on the results, it is observed a moderate to low satisfaction level (33 out of 40) with the outcomes. The chatbot is predominantly utilized as a search tool within the site rather than as a solution-generating engine. The absence of top-rated outcomes is not surprising, as there were no instances of the "magic" of intuitive results characteristic of a well-functioning AI.

Overall Performance: During this assessment, the interviewee is tasked with providing an overall evaluation of the solution they tested. The primary criterion remains the extent of personal effort and discomfort required for service. However, it is important to differentiate this from the prior evaluation of the outcome. If the overall experience was poor, the user might not have progressed to that point unless they were participating in a study. Conversely, they may not have achieved the expected result, but the overall experience might not have been unfavourable.

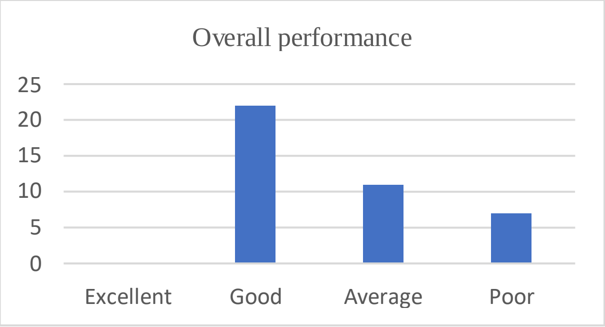

Figure 5: Overall Performance

The pattern observed in the preceding questions remains consistent here. Predominantly, users reported good to moderate performances (33 out of 40) because they did obtain some results, albeit often requiring substantial effort. Achieving high performance was rare, as the aspiration for a fully intuitive system was far from being realized. Conversely, it appears that there was a degree of informal compromise and leniency in assessments when evaluating smaller organizations or businesses.

4.2. Comparative Analysis

Drawing upon the feedback provided by participants in the survey and subsequent assessments of functionality and user experience, we performed a comparative analysis to identify emerging patterns and validate the consistency of responses.

Solution Category with Overall Performance: This comparison enables us to assess the satisfaction rate within each solution category and evaluate it accordingly.

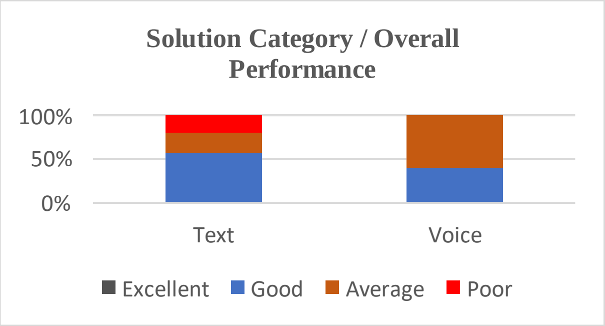

Figure 6: Solution Category with Overall Performance

Voice-based solutions, while achieving only good and average performance, received a lower proportion of votes for good performance compared to the average votes. Conversely, text-based solutions, despite encompassing poor overall experiences, still garnered a satisfaction rate exceeding 50%. It is worth considering that the voice portals were provided by large corporations, which may have elevated user expectations and led to more stringent evaluations. Another contributing factor could be the pressure of responding to questions naturally and swiftly, as required in spoken conversation.

Overall Performance by Economic Sector. We perform this comparison to comprehend the interviewees' ultimate perception about the economic sector that each solution serves.

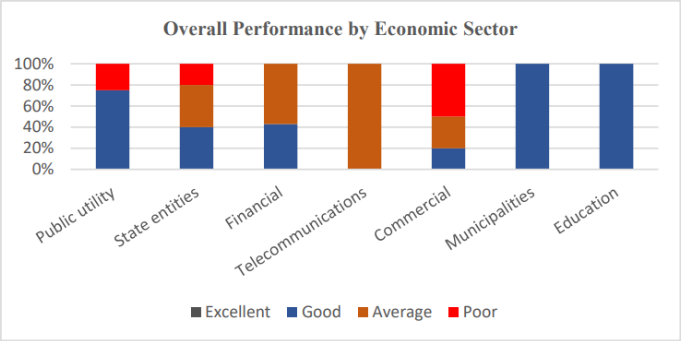

Figure 7: Overall Performance by Economic Sector

Local governance and education receive favourable reviews. They primarily offer information, and the expectations placed on them are not particularly high. Conversely, the commercial sector receives a relatively low rating, which is justified given that the implemented solutions seem to serve as mere supplements to their websites, lacking substantial investments in their functionality.

User Experience in Relation to Overall Performance. Our objective is to assess the consistency of the obtained results, considering the user interface consistently exerts a substantial influence on the overall user experience.

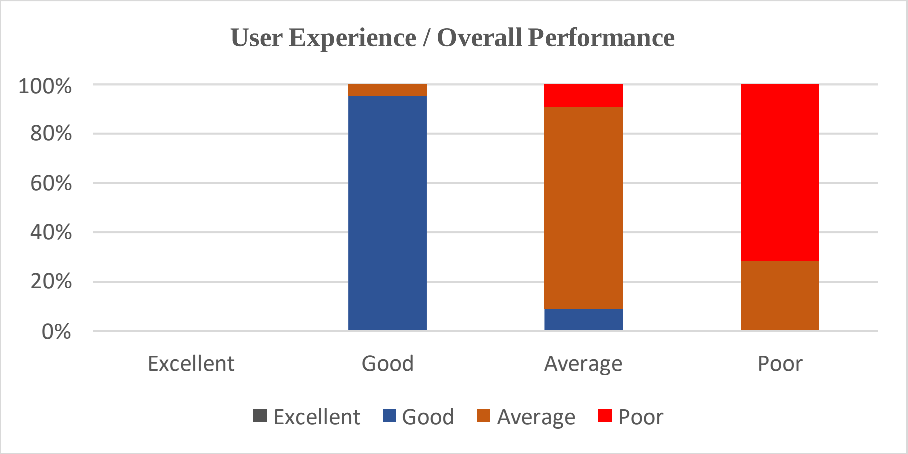

Figure 8: User Experience in Relation to Overall Performance

We notice that, with slight variances, a well-designed interface contributes to a successful overall performance. This correlation is entirely logical, as it would be unlikely for substantial resourcesto have been invested in the technological foundation without a commensurate effort in presentingthe outcome.

User Experience in Relation to Overall Performance: Our objective is to assess the consistency of the obtained results, considering that the interface consistently exerts a significant impact on the user's overall experience.

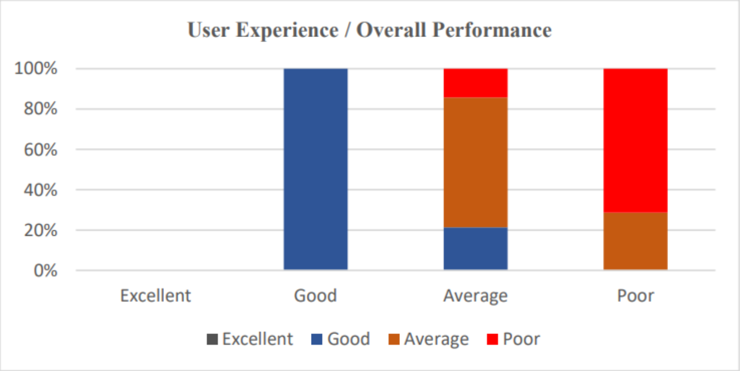

Figure 9: User Experience in Relation to Overall Performance

It is notable that with minimal deviations, a well-designed interface correlates with a favourable overall performance. This alignment is entirely logical, as it would be improbable for significant investments to be made in the technological infrastructure without an equivalent effort in presenting the result.

5. CONCLUSIONS

Looking ahead, it's evident that chatbots will assume an increasingly significant role in our daily lives, with their influence on society and the economy continuing to expand. The ongoing advancements in artificial intelligence and natural language processing technologies will empower chatbots to comprehend and respond to customer queries in a manner closer to human interaction, delivering more precise and personalized answers. Consequently, people will become more comfortable using them, fostering wider adoption.

In the context of Greece, the predominantly average to low levels of satisfaction reflect the challenges presented by the Greek language and underscore the necessity for increased investments. Despite these hurdles, Greece should not lag in technological progress. The country boasts a skilled workforce with the requisite educational background, and there is potential for direct implementation of various solutions in the tourism sector.

In the short term, the focus should be on enhancing NLP technologies and expanding the pool of specialized personnel for chatbot training. In the medium term, investments in new technologies that align better with our linguistic nuances are imperative. Equally critical is the promotion of collaborations between universities, technology companies, the government, and private entities.
Based on the research findings about the private sector, there is a notable demand for more comprehensive electronic services. These services should be accessible 24/7, enabling increased workflow efficiency without a corresponding increase in payroll costs. Furthermore, merely redirecting users to pre-existing product websites without the integration of AI-driven guidance and solutions is deemed unsatisfactory. This approach fails to make chatbots appealing or enticing for users. While digital customer service within large companies is deemed satisfactory, it has not reached an exceptional level, indicating the need for further emphasis on its enhancement. It is advisable to initiate collaborations between academic and research institutions and the private sector to develop and train the first "Greek digital sales assistant," harnessing emerging AI platforms like ChatGPT. Another intriguing prospect would be the creation of a voice-text hybrid

to facilitate complex processes such as contract and agreement signing.

As for the public sector, the research suggests that chatbots primarily serve an informational role, with a complete disconnection from providing tangible services. Additionally, the research team's low expectations imply that the average citizen may resort to more traditional methods if they seek assistance, forfeiting the advantage of 24/7 availability. Immediate funding is recommended for the integration of all administrative processes into the National Registry of Administrative Procedures. There should also be a requirement for lawmakers to model each new process incrementally before submitting it for parliamentary approval. Lastly, the creation of a unified "digital assistant for administrative processes" spanning the entire Greek public sector could be beneficial. This assistant would aid citizens in completing applications, guide them to the appropriate services, and keep them informed about the outcomes, potentially paving the way for the comprehensive digitization of the Greek state.

ACKNOWLEDGEMENTS

This research has been co-financed by the European Union and Greek national funds through the Operational Program "Competitiveness, Entrepreneurship and Innovation", under the call "RESEARCH - CREATE - INNOVATE (2nd Cycle)" (project code: T2EDK-01921).

REFERENCES

[1] Knight, S. (2020). NLP at work. London: Hachette Book Group.

[2] Theodoros Papadopoulos, Yannis Charalabidis. 2002. What do governments plan in the field of artificial intelligence? Analysing national AI strategies using NLP techniques. In Proceedings of the 13th International Conference on Theory and Practice of Electronic Governance (ICEGOV 2020), 23-25 September 2020, Athens, Greece, 12 pages. https://doi.org/10.1145/3428502.3428514

[3] vodafone.gr. (2022, 06 28). Retrieval from https://www.vodafone.gr/vodafone- ellados/digital-press-office/deltia-typou/20220628-tovi-o-psifiakos-voithos-tis- vodafone-pio-exypnos-kai-apotelesmatikos-apo-pote/

[4] Mageira, K., Pittou, D., Papasalouros, A., Kotis, K., Zangogianni, P., & Daradoumis, A. (2022, 03 22). mpdi.com. Retrieval https://doi.org/10.3390/app12073239

[5] The Pythia.Project (2023). Retrieval 09 24, 2023, https://thepythiaproject.com/en/

[6] Koutsikakis, J., Chalkidis, I., Malakasiotis, P., & Androutsopoulos, I. (2020, 09 03). Cornell University. Retrieval arxiv.org: https://arxiv.org/pdf/2008.12014v2.pdf

[7] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre- training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

[8] Weizenbaum, J.: ELIZA—a computer program for the study of natural language communication between man and machine. Commun. ACM 9, 36–45 (1966). https://doi.org/10.1145/365153.365168

[9] Adamopoulou, Eleni, and Lefteris Moussiades. (2020). "An overview of chatbot technology." IFIP International Conference on artificial intelligence applications and innovations (pages. 373 – 383). Springer.

[10] Tullis, T., & Albert, B. (2008). Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics. Burlington: Morgan Kaufmann.

[11] Shawar, A., & Atwell, E. (2007). Different measurements metrics to evaluate a chatbot sys-tem. Proceedings of the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog (pages. 89-96). Seattle: NAACL.

[12] Stent, A., Marge, M., & Mohit, S. (2009). Evaluating Evaluation Methods for Generation in the Presence of Variation. International Conference on Intelligent Text Processing and Com-putational Linguistics (pages. 351-354). Berlin: Springer.

[13] Jelinek, F., Mercer, ., & Salim, . (1991). Principles of Lexical Language Modeling for Speech Recognition. Advances in Speech Signal Processing. New York: Dekker Publishers.

[14] Liu, B. (2012). Sentiment Analysis and Opinion Mining. Synthesis Lectures on Human Lan-guage Technologies. San Rafael: Morgan & Claypool Publishers.

[15] Manning, C., Raghavan, P., & Schütze, H. (2008). Introduction to Information Retrieval. Introduction to Information Retrieval: Cambridge University Press.

[16] Luger, E., & 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 (σσ. 5286-5297). New York: Association for Computing Machinery.

[17] Radziwill, N., & Benton, M. (2017, 06). Cornell University. Retrieval arxiv.org: https://arxiv.org/ftp/arxiv/papers/1704/1704.04579.pdf

[18] Koutsampasis, P. (2015, 8 21). The University of Aegean, Department of Product and systems design engineering. Retrieval eclass.aegean.gr: https://eclass.aegean.gr/modules/document/file.php/511265/merged_document5.pdf

[19] Oppenheim , A. (1992). Questionnaire Design, Interviewing and Attitude Measurement. Lon-don: Continuum International Publishing.

AUTHORS

Theodoros Papadopoulos (male) is a PhD candidate in Machine Learning and Artificial Intelligence. He has been working in ICT since 1998, having acquired and practiced Software development, business and systems analysis, consulting, research, architectural and management skills. He participated in demanding projects and mastered cutting edge technologies, with a strong focus on Data Modeling, Research and Development, Enterprise technology and Open Source frameworks. He worked in highly regarded companies and organizations and he is passionate about eGovernment.

Zoe Lachana (female) is a PhD candidate in the University of the Aegean. She holds a Computer Engineering Degree and a Master of Science in Digital Governance from University of the Aegean, Department of Information and Communication Systems Engineering. Her research interests include Digital Government, Artificial Intelligence, Text and Data Mining. She has experience in teaching lab courses and supervising student projects while also taking part in ERASMUS, HORIZON, CEF and National research projects on Digital Governance.

Thanos Anagnou (male) is a Postgraduate Student in “Digital Governance ”, Department of Information and Communication Systems Engineering, University of the Aegean, Greece. Thanos holds a Bachelor Degree from Hellenic Open University, Department of Informatics.

Dr. Charalampos Alexopoulos (male) holds a PhD diploma from the Department of Information and Communications Systems Engineering at the University of the Aegean on open data. He is a Senior Researcher and Project Manager at the Information Systems Laboratory of the same department, working on European and National funded research and pilot application projects for governments and enterprises. Harris is a computer science graduate from the University of Peloponnese with an MSc in Management Information Systems from the University of the Aegean. In 2016, Harris was ranked as one of the most prolific researchers in open data research worldwide by Hossain, Dwivedi and Rana (2016). He has organised and participated as a mentor in more than 10 Entrepreneurial contests and hackathons.

Yannis Charalabidis (male) is Professor of Digital Governance in the Department of Information and Communication Systems Engineering of the University of Aegean. In parallel, he serves as Director of the Innovation and Entrepreneurship Unit of the University, designing and managing youth entrepreneurship activities, and Head of the Digital Governance Research Centre, coordinating policy making, research and pilot application projects for governments and enterprises worldwide. He has more than 25 years of experience in designing, implementing, managing and applying complex information systems as project manager, in Greece and

Europe. He has been employed for 10 years as an executive director in SingularLogic Group, leading software development and company expansion in Greece, Eastern Europe, India and the US. In 2018, Yannis was nominated among the 100 most influential people in Digital Government worldwide, according to the Apolitical Group.

Christos Bouras (male) is Professor in the University of Patras, Department of Computer Engineering and Informatics. Also he is a scientific advisor of Research Unit 6 in Computer Technology Institute and Press - Diophantus, Patras, Greece. His research interests include 5G and Beyond Networks, Analysis of Performance of Networking and Computer Systems, Computer Networks and Protocols, Mobile and Wireless Communications, Telematics and New Services, QoS and Pricing for Networks and Services, e-learning, Networked Virtual Environments and WWW Issues. He has extended professional experience in Design and Analysis of Networks, Protocols, Telematics and New Services. He has published more than 450 papers in various well-known refereed books, conferences and journals. He is a co-author of 9 books in Greek and editor of 2 in English. He has been member of editorial board for international journals and PC member and referee in various international journals and conferences. He has participated in R&D projects.

Nikos Caracapilidis (male) is professor of Management Information Systems at the Department of Mechanical Engineering and Aeronautics, University of Patras, Greece. Before joining University of Patras (August 2000), he worked as Visiting Assistant Professor at the Dept. of Computer Science, University of Cyprus, and as Research Associate at EPFL – LITH (Switzerland), INRIA – Sophia Antipolis (France), GMD – AiS (Germany), and Queen Mary & Westfield College (UK). He is also an editor of a recent Springer book on Mastering Data-Intensive Collaboration and Decision Making. He was the Scientific Coordinator of the Dicode FP7-ICT project. His work is aimed at supporting and augmenting the synergy of human and machine reasoning in the areas of Computer- Supported Collaborative Work, Business Intelligence and Technology-Enhanced Learning.

Dr. Vasileios Kokkinos (male) was born in Ioannina, Greece in 1981. He obtained his diploma from the Physics Department of the University of Patras on October 2003. Next, he was accepted in the postgraduate program “Electronics and Information Processing” in the same department and on March 2006 he obtained his Master Degree. In 2010 he received his PhD on Power Control in Mobile Telecommunication Networks from the Computer Engineering and Informatics Department. He currently works as Computer and Informatics Engineer and Manager at the Distributed Systems and Telematics Lab of the Computer Engineering and Informatics Department. His research interests include data networks, 3rd/4th/5th generation mobile telecommunications networks, multicast routing and group management, radio resource management and internet of things. He has published several research papers in various well-known refereed conferences and articles in scientific journals.

Apostolos Gkamas (male) obtained his Diploma, Master Degree and Ph.D from the Computer Engineering and Informatics Department of Patras University (Greece). He is currently Assistant Professor (with tenure) in University Ecclesiastical Academy of Vella, Ioannina (Greece). His research interests include Computer Networks, Telematics, Multimedia transmission and Cross Layer Design. He has published more than 70 papers in international Journals and well-known refereed conferences. He is also co-author of three books (one with subject Multimedia and Computer Networks one with subject Special Network Issues and one with subject IPv6). He has participated in various R&D project (in both EU and national) such as IST, FP6, FP7, Intereg eLearning, PENED, EPEAEK, Information Society.