

# Towards a Personalized mHealth Model Using Intelligent Conversational Agents

Noeli Antonia Pimentel Vaz<sup>1,2</sup>, Kairo A. Lopes da Silva<sup>1</sup>, Gislainy Velasco<sup>1</sup>,  
Luciana Regina F. da Mata, Deborah S. Alves Fernandes<sup>1</sup>, Sergio T. Carvalho<sup>1</sup>

<sup>1</sup>Instituto de Informática – Universidade Federal de Goiás (UFG)  
Caixa Postal 131 – 74001-970 – Goiânia – GO – Brasil

<sup>2</sup>Instituto de Ciências Exatas e Tecnológicas - Universidade Estadual de Goiás (UEG)  
Caixa Postal 459 – 75220-400 – Anápolis – GO – Brasil

<sup>3</sup>Escola de Enfermagem – Universidade Federal de Minas Gerais (UFMG)  
Caixa Postal 1556 – 31270-901 – Belo Horizonte - Minas Gerais – Brasil

{noelivaz, kairoantonio, gislainyrisostomo}@discente.ufg.br,

lucianamata@ufmg.br, deborah@inf.ufg.br, sergiocarvalho@ufg.br

**Abstract.** *Mobile health (mHealth) applications are increasingly used for health monitoring but often struggle to engage users effectively. This study proposes a computational model to personalize mHealth experiences using intelligent conversational agents that combine Large Language Models (LLMs), vector-based knowledge retrieval, and sentiment analysis. An experiment was conducted to evaluate the model's application in assisting patients with urinary incontinence after radical prostatectomy. The results demonstrated that incorporating a specialized knowledge base significantly improves the personalization of clinical support. The study underscores the potential of LLMs to improve user experience in healthcare applications and highlights the need for further refinements in adapting interaction strategies to better meet patient needs.*

## 1. Introduction

Mobile health (mHealth) applications are essential health promotion and self-care support tools. These solutions facilitate continuous health data monitoring, including physical activity, vital signs, and dietary habits, thereby providing more significant support to users [Davies and Mueller 2020] and promoting their self-care. Despite the widespread availability of these applications, user engagement remains a critical challenge, limiting their potential impact on clinical outcomes [Mescher et al. 2024]. With the advancement of personalized medicine, which tailors care to patients based on clinical and non-clinical factors, there is a growing demand for adaptable healthcare solutions that account for the unique characteristics of each individual [Mendhe et al. 2024]. Personalization has been identified as a strategy to enhance engagement in mHealth by making interactions more aligned with individual needs [Matthews and Rhodes-Maquaire 2024]. However, traditional personalization approaches, such as those based on fixed rules, present limitations, as they do not dynamically adapt to the behavior and preferences of mHealth users [Vandelanotte et al. 2023].

Recent advances in Artificial Intelligence (AI), particularly in Natural Language Processing (NLP) and Generative AI (GenAI) through Large Language Models (LLMs),

have paved the way for personalization in mHealth. Conversational Agents (CA) based on LLMs, also known as intelligent CA, can interpret and generate natural language, enabling more adaptive and contextualized interactions [Yang et al. 2024a]. These agents dynamically adjust recommendations in real-time, providing continuous support and enhancing user engagement [Martins et al. 2024, Vandelanotte et al. 2023].

Nevertheless, challenges related to adapting to the clinical context, ensuring data security and privacy, and evaluating the effectiveness of interactions still require further investigation to support the development of practical solutions [Montagna et al. 2023]. In this context, the present study proposes a computational model for personalization in mHealth based on intelligent CA. The proposed model integrates an LLM-based CA to personalize interactions and enhance user support. This approach combines contextual and behavioral data to adapt responses and recommendations in real-time.

To address the identified challenges, an initial model version was implemented and evaluated within the intervention field of the mHealth application IUProst®. This application aims to support patients experiencing urinary incontinence resulting from prostate removal surgery (prostatectomy) due to cancer. Developed by an interdisciplinary research team, the solution integrates clinical investigations with advancements in digital technology to promote patient self-care and facilitate the execution of the clinical protocol. The experiment, conducted with the involvement of healthcare professionals from the research group, assessed the coherence and semantic relevance of the responses generated by the model, emphasizing the potential of personalization to enhance user interaction and support.

This paper is structured as follows: Section 2 presents concepts related to personalization in mHealth, GenAI, and intelligent CA; Section 3 discusses related work; Section 4 describes the proposed model; Section 5, in turn, presents the experiments, results, and analyses within the mHealth application context; and finally, Section 6 provides the conclusions and outlines future directions.

## **2. Background**

### **2.1. Personalization in mHealth**

Personalization in mHealth applications refers to the customization of content, functionalities, or interfaces to meet the individual needs of users [Gosetto et al. 2023]. In contrast to the “one-size-fits-all” approach, personalization facilitates more effective and patient-centered interventions [Rivera-Romero et al. 2023].

Personalization in mHealth has been enabled by computational techniques based on client-server architectures, allowing the implementation of data-driven strategies [Vandelanotte et al. 2023, Yang et al. 2024a]. Some approaches employ recommendations based on predefined profiles and fixed rules; however, their rigidity limits the ability to adapt to user changes [Anjos et al. 2024]. Predictive models based on Machine Learning (ML) adjust interventions according to user behavior, enabling more dynamic experiences [Vandelanotte et al. 2023].

Various data can be used for personalization, including health information, user preferences, and environmental sensor data [Ammenwerth et al. 2023]. Studies have explored personalization in different health-supporting areas, such as the promotion of

physical activity [Vandelanotte et al. 2023], diabetes prevention [Rodriguez et al. 2022], mental health support [Stiles-Shields et al. 2022], among others. More recently, LLMs have enabled the dynamic adaptation of content and interactions in mHealth, adjusting responses in real-time through conversations [Yang et al. 2024a, Yang et al. 2024b]. Integrating these techniques enhances the effectiveness of mHealth applications by providing more personalized support.

## **2.2. Generative AI**

Generative AI (GenAI) encompasses techniques capable of automating the creation of various types of content, including text, images, videos, and code. In this study, the application of GenAI is focused explicitly on text generation and comprehension through LLMs. These models are trained on extensive volumes of textual data to interpret, generate, and adapt the natural language in a fluent and contextualized manner [Chang et al. 2024]. GPT-4, LLaMA, Claude, and Deepseek-R1 demonstrate high capabilities as general-purpose chatbots providing helpful responses and agents.

Using LLMs in mHealth enables dynamic, user-tailored support, facilitating effective interactions [Montenegro et al. 2019]. However, challenges include the need for reliable and secure responses, particularly for vulnerable patients [Goumas et al. 2024]. Algorithmic biases and the lack of clinical validation may compromise the acceptance of these technologies, necessitating clear safety guidelines and continuous monitoring.

In leveraging the computational power of LLMs, prompt engineering has been explored to tailor their capabilities to specific domains without retraining, a task associated with high computational and financial costs [Sahoo et al. 2024]. Techniques such as Zero-Shot, Few-Shot, Chain-of-Thought, and Retrieval-Augmented Generation (RAG) mitigate hallucinations and enhance the model's ability to adapt to sensitive contexts [Sahoo et al. 2024]. RAG integrates information retrieval with generative models, thereby improving the accuracy of responses [Li 2024].

Integrating an LLM-based engine within an application leverages the use of autonomous agents for information processing. Generative AI autonomous agents consist of several key components. The interface component manages the input and output channels, facilitating interaction with users and external systems. The LLM component is the agent's "brain," responsible for decision-making and response generation. The workflow orchestrates data management and task execution. Advanced components, such as memory, knowledge base, and tools, enhance interaction by allowing the agent to learn and refine its responses over time. Finally, the configuration component adjusts the agent's functionality, allowing customization to meet the application's specific needs [Li 2024, Qiu et al. 2024].

## **2.3. Intelligent Conversational Agents**

Intelligent conversational agents, namely CA based on LLMs, provide dynamic support, interpret dialogues, and interact through natural language to promote engagement and support self-care [Vandelanotte et al. 2023].

An intelligent CA can serve as a support interface within a mHealth application, offering personalized and continuous interactions to assist patients in managing their health. By utilizing NLP techniques and technologies, the intelligent CA can interpret

user messages and respond contextually, guiding treatments, medication reminders, and recommendations for healthy habits. Additionally, it can record patient-reported symptoms, monitor vital signs, and issue alerts to healthcare professionals in critical situations [Montenegro et al. 2019, Martins et al. 2024].

Intelligent CA can emulate human-like dialogues, providing emotional and informational support that sustains user engagement in well-being activities [Martins et al. 2024]. Additionally, they can guide users through structured dialogues focused on specific health goals, thereby enhancing motivation and adherence to self-care practices [Wen et al. 2024]. By analyzing user data and preferences, intelligent CA can tailor conversations to individual needs, increasing the relevance of the information provided. Although intelligent CA appears promising, ethical concerns, such as privacy and bias in the data used for pre-training LLMs, must be adequately addressed to ensure their safe implementation in healthcare settings [Wen et al. 2024]. A state-of-the-art analysis by the authors of [Goumas et al. 2024] examined the use of intelligent CA in medical and cancer contexts, suggesting that it should be employed only in specific and controlled situations where the benefits outweigh the risks. These findings underscore the need to carefully design, implement, and evaluate these technologies.

### **3. Related Work**

Various approaches have been investigated for using intelligent CA to personalize mHealth applications. The studies analyzed explore personalization with different scopes, enhance the evolution of strategies, and the challenges involved.

The ChatDiet framework [Yang et al. 2024a] employs LLMs for personalized nutritional recommendations, integrating a causal inference model with a population model to provide general information. Although it demonstrates the potential of LLMs for interpretation and interaction, its approach is limited to the nutrition domain, without extending to other aspects of self-care.

Similarly, [Vandelanotte et al. 2023] developed a digital assistant to encourage physical activity by combining reinforcement learning and GenAI. Personalization is achieved through an adaptive suggestion mechanism (nudge engine) and educational interactions; however, using LLMs is restricted to a question-and-answer functionality, limiting the potential for dynamic adaptation.

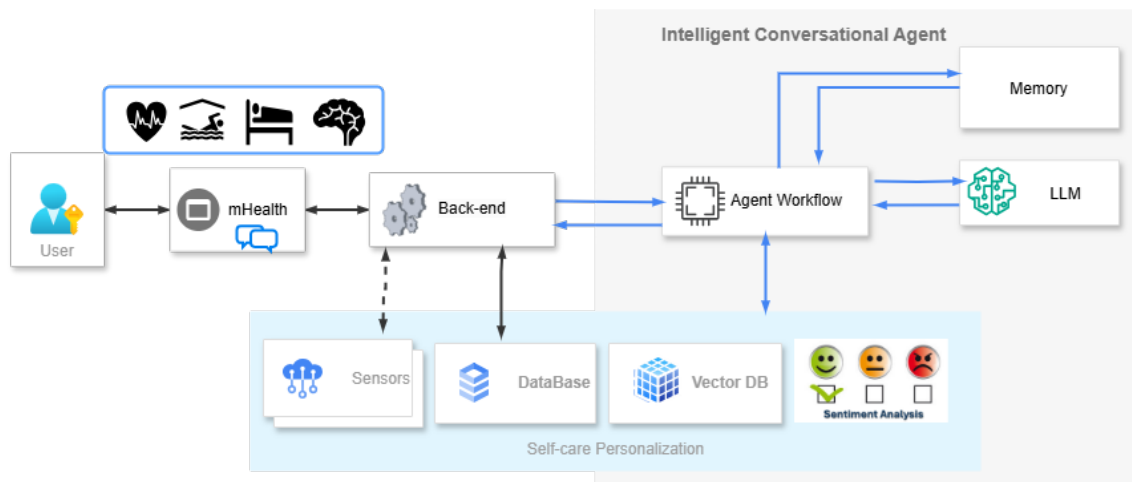
Talk2Care, presented by [Yang et al. 2024b], focuses on communication between elderly patients and healthcare professionals, using LLMs for voice interaction and organizing clinical information. While it facilitates communication and data storage, the system does not provide active support for self-care or personalized recommendations.

These studies demonstrate progress in incorporating intelligent CA into mHealth, emphasizing their relevance for personalized care. However, challenges remain to be addressed, such as adapting models to the complexity of healthcare interactions, establishing clear guidelines for clinical validation, and enhancing user trust and safety. The model proposed in this study seeks to contribute to this advancement by integrating LLMs with knowledge retrieval techniques and sentiment analysis, aiming to personalize interactions within mHealth applications.

#### 4. Proposed Model

The model proposed in this study incorporates an autonomous LLM-based agent integrated into mHealth applications to process input and output during conversational interactions. This approach emphasizes personalizing the user experience by combining contextual and behavioral data sources. In addition to conversation modeling, the agent is connected to a personalization system that leverages stored mHealth application data and sentiment analysis. As a result, the model manages communication while adapting responses and recommendations based on the user's profile, aiming to provide more dynamic and context-aware support.

Providing personalized responses for tasks requiring empathy and analytical precision is essential in application contexts such as healthcare [Li et al. 2024]. Figure 1 presents the model for personalization in mHealth based on intelligent CA, including the operational flow involving a mHealth application.



**Figure 1. Model for personalization in mHealth based on intelligent CA.**

In the proposed model, the application user communicates with the intelligent CA of the mHealth application through the back-end. Integrated with this back-end, the Agent Workflow receives the input information and utilizes data sources for personalization by accessing specific information related to the mHealth application context to understand the user better. These data sources consist of the mHealth application's data (Database) and the sensors it employs (if applicable), as well as the vector database (Vector DB). The Vector DB stores the knowledge base related to the healthcare context and language models for sentiment analysis. This database supports the efficient retrieval of information, enabling the agent to provide contextualized responses tailored to the user's needs. The LLM processes these data, generates a response, and maintains a memory history to offer more personalized and coherent interactions. Below is a detailed overview of the model components.

- User: interacts with the mHealth application to record information, receive recommendations, and manage their self-care.
- mHealth: an application available on a mobile device that facilitates communication between the user and the application's back-end by collecting and processing the user's health information, including cognitive data, physical activity, and clinical conditions.

- **Back-end:** the component responsible for processing data from the mHealth application and managing the interaction with the intelligent CA.
- **Self-care Personalization:** encompasses the components used for personalization (sensor and application data), a vector database storing the knowledge base, and the sentiment analysis component.
- **Intelligent Conversational Agent:** an autonomous unit capable of performing tasks independently, guided by specific instructions and contextual understanding.
  - **Agent Workflow:** processes user requests and defines how to interact with the LLM. This workflow is a pre-coded plan responsible for decision-making.
  - **Memory:** retains contextual information to ensure continuous and coherent interactions.
  - **LLM:** a GenAI engine that generates responses and supports the personalization of interactions.

The model is supported by three essential approaches aimed at enhancing personalization and the efficiency of the intelligent CA's interaction: (i) dynamic flows based on real-time data, allowing continuous adjustments to responses according to user behavior; (ii) RAG, ensuring more accurate and contextualized responses based on a specialized knowledge base; and (iii) response adjustment according to sentiment analysis, promoting more empathetic interactions aligned with the user's emotional state.

The intelligent CA's personalization relies on its ability to interpret real-time data, including information collected from sensors (when available), application usage logs, and user feedback. This continuous data flow enables responses and recommendations to evolve according to the user's needs, allowing for dynamic adjustments in the interaction. For example, if a patient exhibits an increased urinary frequency recorded in the application, the agent may suggest preventive measures or recommend seeking medical assistance. Similarly, patterns of low adherence to the rehabilitation protocol may prompt the agent to reinforce personalized reminders or adjust the tone of the messages to encourage engagement. This approach enables more responsive and adaptable support for the user's health status changes.

The RAG approach enables the agent to consult a specific knowledge base before generating responses, ensuring greater accuracy and contextualization. Unlike methods that rely solely on the pre-trained knowledge of the LLM, RAG combines information retrieval and language generation, allowing the generated content to be grounded in reliable sources. In the proposed model, a vector database stores medical guidelines, clinical documents, and information related to patient self-care. During the interaction, the agent retrieves the most relevant data and incorporates it into the LLM's prompt, thereby reducing the risk of hallucinations and improving response accuracy. This strategy strengthens the reliability of recommendations and enhances personalization, supporting contextualized interactions and providing safe information.

The incorporation of sentiment analysis enables the agent to adjust the tone and content of its responses based on the emotions detected in the user. By identifying signs of frustration, anxiety, or demotivation, the agent can provide encouraging messages or modify the type of recommendation presented, thereby making the interaction more empathetic. Studies indicate that considering users' emotional states can enhance engagement and improve treatment adherence [Gandy et al. 2025]. A sentiment taxonomy was

developed with healthcare professionals to guide response generation, ensuring an appropriate balance between informational detail and user comprehension.

## 5. Experiments and Results

The experiment was conducted within the context of the mHealth application IUProst®<sup>1</sup>, developed to provide support to patients undergoing treatment for urinary incontinence resulting from radical prostatectomy, performed as part of prostate cancer treatment. The application offers essential functionalities for monitoring and managing the patient's health condition. Through these features, the mHealth application encourages the patient's active participation in self-care, aiming to improve quality of life and rehabilitate urinary continence in the postoperative period. Available for download on the Google Play Store platform, the mHealth application IUProst® has approximately one thousand registered users [Vaz et al. 2024]. It is supported by published clinical studies demonstrating its effectiveness [Estevam et al. 2023, Anjos et al. 2024].

The experiment involves developing an intelligent CA for the integration into the IUProst® application, incorporating the RAG component, which utilizes a Vector Database (Vector DB). This implementation represents a partial realization of the proposed model, aimed at validating its initial version. The experimental setup includes designing the agent's workflow, utilizing clinical knowledge base materials from the IUProst® application, and integrating the RAG component to enrich the contextual understanding, allowing the LLM to generate personalized responses. This experiment aims to enable the preliminary evaluation of the RAG-based approach proposed in the model, as presented in Figure 1, which depicts the interaction between the Agent Workflow and the Vector DB. Evaluating the interactions performed by healthcare professionals is a critical step in refining the modular components of the intelligent CA, before the full deployment of the proposed model.

The Make framework<sup>2</sup> was used to develop the CA. This tool enables the integration of various services and automation through external application programming interfaces (APIs). Using these APIs, the RAG model was configured for information retrieval from the knowledge base, user communication via WhatsApp, and the implementation of the Llama 70B. In configuring the intelligent CA, three main prompts were employed: one for the classification task, another to refine and enhance the response retrieved by the RAG, and a third for the direct generation of responses. The latter was designed to address general patient inquiries unrelated to treatment, utilizing a logical prompt that did not require few-shot learning. This architecture enabled the dynamic generation of contextualized and relevant responses tailored to the clinical domain of the intelligent CA.

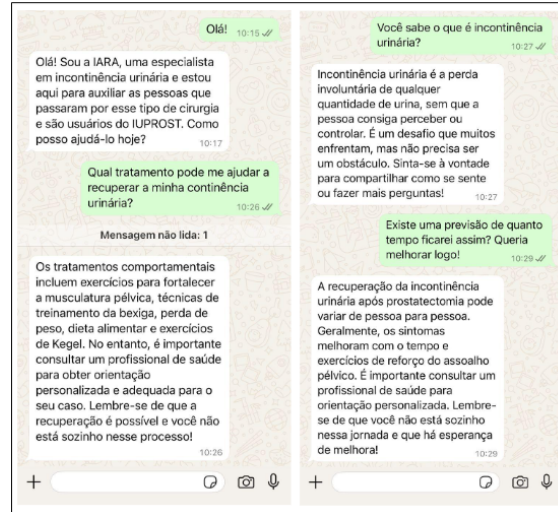
Eight healthcare professionals specializing in urinary incontinence participated in the experiment, which was conducted in November 2024. The interactions were carried out via WhatsApp, utilizing an intelligent CA designed to provide information based on a knowledge repository focused on the treatment of urinary incontinence and prostate cancer. The primary objective of this phase was to evaluate the coherence and relevance of the model's responses, ensuring their alignment with clinical needs and their effectiveness

---

<sup>1</sup> Approved by the Research Ethics Committee of the University IUProst®, approval number -.

<sup>2</sup> Available at: [www.make.com](http://www.make.com)

in supporting patient care. Figure 2 presents the original interface (with dialogues in Portuguese) and illustrates two interactions between the CA and the healthcare professional.



**Figure 2. Interactions in Portuguese with the intelligent CA.**

The dataset analyzed comprises 40 frequently asked questions posed by patients regarding urinary incontinence following prostatectomy. These questions were formulated based on typical interactions with healthcare professionals and clinical reference materials on the subject. The responses generated by the CA were then compared to reference answers extracted from academic sources and clinical guidelines provided by the project's clinical team.

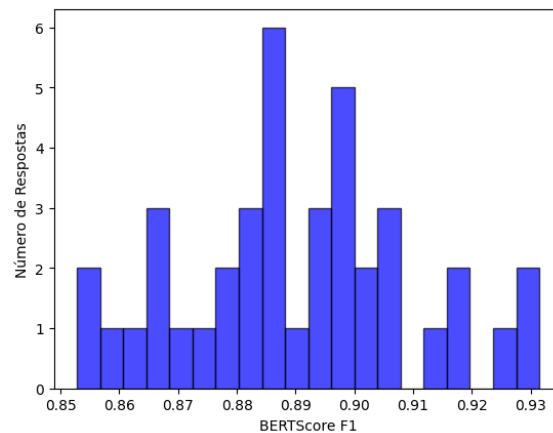
The evaluation of the proposed model requires a comprehensive approach that balances technical performance, usability, personalization, and clinical safety. While usability is a crucial factor for the success of mHealth applications utilizing conversational agents (CAs), a clear standard procedure for assessing the usability of health-related CAs has yet to be established [Denecke and May 2022]. Therefore, systematically analyzing response quality, user engagement, and clinical reliability is essential to ensure the model's effectiveness and safety [Deniz-Garcia et al. 2023].

Thus, the BERTScore metric (Precision, Recall, F1-Score), which measures the semantic similarity between the responses generated by the intelligent CA and a set of reference responses [Zhang et al. 2019], was employed to assess the quality of the CA's outputs. Unlike traditional metrics such as BLEU and ROUGE, BERTScore leverages contextual embeddings to capture the semantic proximity between responses [Chuayrod et al. 2024], making it more appropriate for evaluating coherence and accuracy in specialized domains like urinary incontinence and post-prostatectomy.

The average F1-Score obtained for the set of responses was 0.89, indicating a high semantic similarity between the responses generated by the intelligent CA and the reference responses. Figure 3 presents the distribution of the F1-Scores for the analyzed responses. The analysis of this distribution reveals that most responses with scores above 0.88 are semantically aligned with the references. Responses with scores ranging from 0.85 to 0.87 suggest some omission of terms or variations in terminology. Scores below 0.85 correspond to responses that may exhibit incompleteness or reformulation that



partially alters the intended meaning.



**Figure 3. F1-Score Distribution.**

In addition to the quantitative analysis, a qualitative examination of selected responses generated by the intelligent CA was conducted with the support of a healthcare professional from the clinical team. The qualitative analysis revealed that the intelligent CA was able to provide appropriate responses in the interactions analyzed, particularly when the knowledge base contained detailed and accurate information. One example of such an interaction is as follows:

**Question:** *When will my urinary incontinence go away?*

**Answer:** *Urinary incontinence after prostatectomy varies from person to person. Some patients regain continence within a few months, while others may take longer. The main treatment approaches include behavioral interventions, medications, and, in more severe cases, surgical procedures. Since each case is unique, it is essential to follow the guidance of a healthcare professional for personalized support. Are you feeling any better? Would you like to discuss this further?*

On the other hand, when questions lack sufficient contextual information, it was observed that the CA might generate generic, incomplete, or incorrect responses. The following interaction illustrates this issue:

**Question:** *Where are the pelvic floor muscles located?*

**Answer:** *The pelvic floor muscles.*

In this case, the response is imprecise and lacks the necessary anatomical detail and contextualization with the topic under discussion, which may compromise the user's understanding. Similarly, in more specific inquiries, the response presents an evidence-based technical recommendation that is accurate and useful. However, it is noted that this response could be enhanced through a more personalized and empathetic approach, considering factors such as the user's level of understanding regarding the exercises, potential difficulties in execution, and the emotional impact of the rehabilitation process. The following interaction illustrates this issue:

**Question:** *How many exercises should I do?*

**Answer:** *15 to 20 Kegel exercises per set, three sets a day.*

The quantitative analysis indicates that the majority of responses achieve an F1-Score above 0.88, highlighting the CA's ability to generate semantically accurate replies. However, some inconsistencies suggest limitations in the contextual retrieval and response generation process of the LLM, emphasizing the need for further improvements to the RAG model.

The qualitative analysis complemented this evaluation by uncovering nuances that could not be fully captured through quantitative indicators alone. The intelligent CA showed satisfactory performance when delivering contextualized responses, particularly when the knowledge base contained detailed information, as seen in interactions involving explanations of the therapeutic approaches. However, the responses were often generic, incomplete, or inaccurate for vague questions or those with limited contextual information. This highlighted the CA's reliance on the quality of input to generate accurate responses, which raises concerns regarding the characteristics of the target audience interacting with mHealth applications.

Furthermore, while certain technical responses are accurate and evidence-based, they could benefit from a more empathetic and personalized approach. For example, exercise recommendations could be enhanced by including information tailored to the user's knowledge level, execution challenges, and the emotional impact of the treatment. These findings suggest that incorporating mechanisms to adjust the language and tone of responses according to the user's profile may improve the overall user experience and promote a higher level of engagement.

## **6. Conclusion**

This study proposed a computational model for personalization in mHealth applications based on intelligent CA. The model integrates data from mHealth applications, information retrieval from a specialized knowledge base, and sentiment analysis to enhance user interaction. A partial implementation of the model, focusing on the RAG component, facilitated the evaluation of the LLM's ability to generate coherent and contextually relevant responses to support patients experiencing urinary incontinence following prostatectomy.

The evaluation of the proposed model requires a comprehensive approach, balancing technical performance, usability, personalization, and clinical safety. Although usability is a key factor for the success of intelligent CA in mHealth, there is no consolidated standard protocol for its assessment yet [Denecke and May 2022]. Additionally, response quality, user engagement, and clinical reliability must be systematically analyzed [Deniz-Garcia et al. 2023]. The model validation involves evaluating automated tests to measure response accuracy and contextual adaptation; engagement metrics to assess user adherence and user experience; and a clinical review to ensure that the agent's recommendations align with medical guidelines, thereby minimizing the risk of providing incorrect information.

The quantitative experiments demonstrated that the intelligent CA achieved a high level of semantic similarity with the reference responses. Concurrently, the qualitative analysis revealed that the responses were more accurate when the knowledge base contained detailed information. Limitations were also identified, such as generating generic responses to questions with limited contextualization. These findings highlight the importance of active collaboration with the clinical team and suggest that the model should be

continuously refined based on evaluation results, fostering iterative optimization cycles.

The initial results reinforce the model's potential to deliver more personalized and humanized support within mHealth applications. Moving forward, the next steps of this research involve implementing the components responsible for the agent's workflow and integrating mHealth application data, sensors, and sentiment analysis. These enhancements aim to improve the dynamic adaptation of responses and the personalization of interactions. Furthermore, a more comprehensive evaluation will be conducted, incorporating engagement and usability metrics to validate the solution's effectiveness within the clinical context.

In light of these findings, opportunities for refining the model have been identified, underscoring the importance of including the subsequent components that address the recognized needs. These improvements could significantly strengthen the role of conversational agents in rehabilitation and self-care support within mHealth applications, thereby promoting patient well-being and encouraging adherence to treatment protocols.

## References

- Ammenwerth, E. et al. (2023). Personalization in mHealth: Innovative informatics methods to improve patient experience and health outcome. *Journal of Biomedical Informatics*, 147:104523.
- Anjos, F. M. S. d. et al. (2024). Analyzing the impact of gamification on a mhealth application for treating urinary incontinence in prostate cancer patients. *Journal on Interactive Systems*, 15(1):728–740.
- Chang, Y. et al. (2024). A survey on evaluation of large language models. *ACM TIST*.
- Chuayrod, P. et al. (2024). The Impact of Prompt Engineering on Large Language Models: A Case Study of Sustainable Development Goals. In *2024 19th iSAI-NLP*, pages 1–6.
- Davies, A. and Mueller, J. (2020). Introduction to mHealth. pages 1–24. Springer.
- Denecke, K. and May, R. (2022). Usability assessment of conversational agents in health-care: a literature review. *Challenges of Trustable AI and Added-Value on Health*.
- Deniz-Garcia, A. et al. (2023). Quality, usability, and effectiveness of mHealth apps and the role of artificial intelligence: current scenario and challenges. *JMIR*, 25:e44030.
- Estevam, F. E. B., , et al. (2023). Desenvolvimento e análise de qualidade do app iuprostop para controle da incontinência urinária em prostatectomizados. *Revista Remecs-Revista Multidisciplinar de Estudos Científicos em Saúde*, pages 44–44.
- Gandy, L. M. et al. (2025). Public Health Discussions on Social Media: Evaluating Automated Sentiment Analysis Methods. *JMIR Formative Research*, 9(1):e57395.
- Gosetto, L. et al. (2023). Personalization of Mobile Apps for Health Behavior Change: Protocol for a Cross-sectional Study. *JMIR Research Protocols*, 12(1):e38603.
- Goumas, G. et al. (2024). Chatbots in Cancer Applications, Advantages and Disadvantages: All that Glitters Is Not Gold. *Journal of Personalized Medicine*, 14(8):877.
- Li, C. (2024). Designing LLM-Based Agents: Key Principles (Part 1). <https://medium.com/binome/designing-llm-based-agents-key-principles-part-1-7e8c6fe3ddaf>.

- Li, Y. et al. (2024). Quantifying ai psychology: A psychometrics benchmark for large language models. *arXiv preprint arXiv:2406.17675*.
- Martins, A. et al. (2024). Unlocking human-like conversations: Scoping review of automation techniques for personalized healthcare interventions using conversational agents. *International Journal of Medical Informatics*, page 105385.
- Matthews, P. and Rhodes-Maquaire, C. (2024). Personalisation and Recommendation for Mental Health Apps: A Scoping Review. *B & I Technology*, pages 1–16.
- Mendhe, D. et al. (2024). AI-Enabled Data-Driven Approaches for Personalized Medicine and Healthcare Analytics. In *2024 Ninth ICONSTEM*, pages 1–5. IEEE.
- Mescher, T. et al. (2024). Mobile Health Apps: Guidance for Evaluation and Implementation by Healthcare Workers. *Technology in Behavioral Science*, pages 1–12.
- Montagna, S. et al. (2023). Data decentralisation of LLM-based chatbot systems in chronic disease self-management. In *ACM GoodIT*, pages 205–212.
- Montenegro, J. L. Z. et al. (2019). Survey of conversational agents in health. *Expert Systems with Applications*, 129:56–67.
- Qiu, J. et al. (2024). LLM-based agentic systems in medicine and healthcare. *Nature Machine Intelligence*, 6(12):1418–1420.
- Rivera-Romero, O. et al. (2023). Designing personalised mHealth solutions: An overview. *Journal of Biomedical Informatics*, page 104500.
- Rodriguez, D. V. et al. (2022). PAMS-A Personalized Automatic Messaging System for User Engagement with a Digital Diabetes Prevention Program. In *2022 IEEE 10th ICHI*, pages 297–308. IEEE.
- Sahoo, P. et al. (2024). A systematic survey of prompt engineering in large language models: Techniques and applications. *arXiv preprint arXiv:2402.07927*.
- Stiles-Shields, C. et al. (2022). Digital Screening and Automated Resource Identification System to Address COVID-19–Related Behavioral Health Disparities: Feasibility Study. *JMIR Formative Research*, 6(6):e38162.
- Vandelanotte, C. et al. (2023). Increasing physical activity using an just-in-time adaptive digital assistant supported by machine learning: A novel approach for hyperpersonalised mHealth interventions. *Journal of Biomedical Informatics*, 144:104435.
- Vaz, N. A. P. et al. (2024). O uso da metodologia crisp-dm para apoiar a análise de dados no aplicativo mhealth iuprost. In *Simpósio Brasileiro de Computação Aplicada à Saúde (SBCAS)*, pages 448–458. SBC.
- Wen, B. et al. (2024). Leveraging Large Language Models for Patient Engagement: The Power of Conversational AI in Digital Health. In *2024 IEEE ICDH*, pages 104–113.
- Yang, Z. et al. (2024a). ChatDiet: Empowering personalized nutrition-oriented food recommender chatbots through an LLM-augmented framework. *Smart Health*.
- Yang, Z. et al. (2024b). Talk2Care: An LLM-based Voice Assistant for Communication between Healthcare Providers and Older Adults. *ACM IMWUT*, 8(2):1–35.
- Zhang, T. et al. (2019). Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.