TESIS DE LA UNIVERSIDAD
DE ZARAGOZA

2023

143

Surya Roca Mainer

Development and evaluation of a microservice-based virtual assistant for chronic patients support

Director/es
Alesanco Iglesias, Álvaro

http://zaguan.unizar.es/collection/Tesis

ISSN 2254-7606

Prensas de la Universidad
Universidad Zaragoza

© Universidad de Zaragoza
Servicio de Publicaciones

ISSN 2254-7606

Universidad Zaragoza
1542

Tesis Doctoral

DEVELOPMENT AND EVALUATION OF A MICROSERVICE-BASED VIRTUAL ASSISTANT FOR CHRONIC PATIENTS SUPPORT

Autor

Surya Roca Mainer

Director/es

Alesanco Iglesias, Álvaro

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

Programa de Doctorado en Tecnologías de la Información y Comunicaciones en Redes Móviles

2023

Repositorio de la Universidad de Zaragoza – Zaguan http://zaguan.unizar.es

Tesis Doctoral

DEVELOPMENT AND EVALUATION OF A MICROSERVICE-BASED VIRTUAL ASSISTANT FOR CHRONIC PATIENTS SUPPORT

Autor

Surya Roca Mainer

Director

Álvaro Alesanco Iglesias

Escuela de Ingeniería y Arquitectura

2022

Repositorio de la Universidad de Zaragoza – Zaguan http://zaguan.unizar.es

University of Zaragoza

Mobile Network Information and Communication Technologies Doctoral Program

Development and Evaluation of a Microservice-Based Virtual Assistant for Chronic Patients Support

Ph.D. Dissertation

Surya Roca Mainer

Supervisor

Dr. Álvaro Alesanco Iglesias

December 2022

Acknowledgements

I would like to thank my supervisor Dr. Álvaro Alesanco for his unforgettable inspiration, generous guidance, and supervision. I am very thankful to Dr. José García for all the support and help offered. Special thanks to Dr. Sophie Rosset for giving me the opportunity to work with a wonderful and exciting team and field. Thanks to Dr. Paul Turner for the chance to learn and work with him. Special thanks to Dr. Yolanda Gilaberte, Dr. María Luisa Lozano, and the rest of the healthcare professionals who contributed to this work.

A warm thank you to all my colleagues from the University of Zaragoza, especially to lab 2.05, where I have learned and shared beautiful moments over the years.

It is a great pleasure to thank the reviewers and all people who directly or indirectly helped me improve this work's quality. Also, I would like to thank the committee members who evaluated my research.

Additionally, I would like to acknowledge my friends for their valuable advice, company and support.

Words cannot express how grateful I feel to my family, who unconditionally supported and encouraged me. Special gratitude to my parents, Carmen and Luis, and my sister, Isis, for their entire support and love.

Agradecimientos

En primer lugar, me gustaría agradecer a mi director de tesis, el Dr. Álvaro Alesanco, por su gran inspiración, permanente guía y supervisión. Quisiera agradecer de manera especial al Dr. José García por todo el apoyo y la ayuda brindada. En especial, me gustaría agradecer a la Dra. Sophie Rosset por darme la oportunidad de trabajar con un equipo maravilloso y un campo emocionante. Gracias al Dr. Paul Turner por la oportunidad de poder aprender y trabajar con él. Un agradecimiento especial a la Dra. Yolanda Gilaberte, la Dra. María Luisa Lozano y el resto de los profesionales sanitarios que contribuyeron a este trabajo.

Quiero aprovechar para agradecer a todos mis compañeros de la Universidad de Zaragoza, en especial al laboratorio 2.05, donde he aprendido y compartido muy buenos momentos a lo largo de estos años.

Agradecer a los revisores y a todas las personas que directa o indirectamente me han ayudado a mejorar la calidad de este trabajo. También, me gustaría agradecer a los miembros del tribunal que evaluaron mi investigación.

Además, me gustaría agradecer a mis amigos por sus valiosos consejos, su compañía y su gran apoyo.

Finalmente, las palabras no pueden expresar lo agradecida que me siento con mi familia, quienes me han apoyado y animado incondicionalmente. Gracias en particular a mis padres, Carmen y Luis, y a mi hermana, Isis, por todo su apoyo y cariño.

Abstract

Virtual assistants (a.k.a. chatbots) are programs that interact with users through text or voice messages simulating a human-based conversation. Healthcare-virtual assistants offer services, tools, and interactions to the users, providing advice, help, support, and management of different diseases. The users of this type of virtual assistant can be, for example, patients, caregivers, and healthcare professionals, with different needs and requirements. Patients with chronic diseases could benefit the most from virtual assistants which can keep track of their condition, provide specific information, encourage adherence to medication, etc. To perform these functions, virtual assistants need a suitable underlying software architecture. In this thesis, **we design and introduce a virtual assistant architecture for chronic patient support**. Nowadays, people interact with each other daily using messaging platforms. To align this type of interaction with the virtual assistant architecture, we propose the use of messaging platforms for the virtual assistant-patient interaction, paying special attention to security and privacy issues (i.e. the use of secure messaging platforms with end-to-end encryption).

Virtual assistants may implement conversational systems to make interaction with patients more natural. A conversational system in complex healthcare scenarios such as disease management should be able to understand complex sentences used during the interaction. The adaptation of new methods with Natural Language Processing (NLP) may provide an improvement to the virtual assistant architecture. Word embeddings have been widely used in NLP as the input to neural networks. Such word embeddings can help in the understanding of the final objective and the keywords in a sentence. As such, in this thesis, we study the impact of different word embeddings trained with general and specific corpora using Joint Natural Language Understanding (NLU) in a Spanish medication domain. We generate data using templates for training the joint NLU model. This model is used for intent detection and slot filling. We compare word2vec and fastText as word embeddings and ELMo and BERT as language models. We use three different corpora to train the embeddings: the training data generated for this scenario, the Spanish Wikipedia as the general domain, and the Spanish drug database as specialized data. The best result was obtained with **ELMo model trained with Spanish Wikipedia**.

We provide the virtual assistant with medication management capabilities based on NLP. Thus, the impacts of slot tagging and training data length on joint NLU models for medication management scenarios using virtual assistants in Spanish are analyzed. In this study, we define the intents (purposes of the sentences) for medication management scenarios and two types of slot tags. For training the model, we generated four

datasets, combining long or short sentences with long or short slots. For the comparative analysis, we chose six joint NLU models (SlotRefine, stack-propagation framework, SF-ID network, capsule-NLU, slot-gated modeling, and joint SLU-LM) from the existing literature. The results showed that the best performance was achieved using **short sentences and short slots**. Our results suggested that joint NLU models trained with **short slots** yielded better results than those trained with long slots **for the slot filling task**.

All in all, we propose a generic microservice architecture valid for any kind of chronic condition management. The **generic prototype** offers an operative virtual assistant to manage basic information and provide a base for future extensions. Besides, we present **two specialized prototypes** to show how this new architecture allows the change, the addition, or the improvement of different parts of the virtual assistant in a dynamic and flexible way. The first specialized prototype implemented has the aim to help in the patient's medication management. This prototype will remind medication intakes through the creation of a supportive community where patients, caregivers, and healthcare professionals interact with helpful tools and services offered by the virtual assistant. The implementation of the second specialized prototype is tailored for one specific chronic disease, psoriasis. This prototype offers teleconsultation and image storage.

Finally, this thesis aims to **validate the effectiveness of the healthcare-virtual assistant** integrated within messaging platforms. For this reason, this thesis includes the evaluation of the two specialized prototypes. The first study has the aim of **improving medication adherence** in patients with comorbid type 2 diabetes mellitus and depressive disorder. For this purpose, a nine-month pilot study was designed and subsequently conducted. We analyzed the Medication Possession Ratio (MPR), measured the level of glycosylated hemoglobin (HbA1c), and obtained the Patient Health Questionnaire (PHQ-9) score in the patients before and after the study. We also conducted interviews with all participants. A total of thirteen patients and five nurses used and evaluated the proposed virtual assistant. Outcomes showed that, on average, **patients' medication adherence improved**. The second study aims to **evaluate** one year of use between the virtual assistant and psoriatic patients and dermatologists and **the impact on their quality of life**. For this purpose, a one-year prospective study was designed and subsequently conducted among patients with psoriasis and dermatologists. To measure the improvement in the quality of life, we analyzed Psoriasis Quality of Life (PSOLIFE) and Dermatology Life Quality Index (DLQI) questionnaires. Also, we conducted surveys with all participants and obtained the number of medical consultations made through the virtual assistant. A total of 34 participants (30 patients diagnosed with moderate-severe psoriasis and four healthcare professionals) were included in the study. Results showed that, on average, **the quality of life improved**.

Resumen y Conclusiones

Los asistentes virtuales (también conocidos como chatbots) son programas que interactúan con los usuarios simulando una conversación humana a través de mensajes de texto o de voz. Los asistentes virtuales destinados al cuidado de la salud ofrecen servicios, herramientas, asesoramiento, ayuda, soporte y gestión de diferentes enfermedades. Los usuarios de este tipo de asistente virtual pueden ser, por ejemplo, pacientes, cuidadores y profesionales sanitarios, los cuales poseen diferentes necesidades y requerimientos. Los pacientes con enfermedades crónicas podrían beneficiarse de los asistentes virtuales que se encargan de realizar seguimientos de su condición, proporcionar información específica, fomentar la adherencia a la medicación, etc. Para realizar estas funciones, los asistentes virtuales necesitan una arquitectura de software adecuada. Esta tesis doctoral propone el **diseño de una arquitectura específica para el desarrollo de asistentes virtuales destinados a proporcionar soporte a pacientes crónicos**. Hoy en día, las personas interactúan entre sí diariamente utilizando plataformas de mensajería. Para alinear este tipo de interacción con la arquitectura del asistente virtual, proponemos el uso de plataformas de mensajería para la interacción asistente virtual-paciente, prestando especial atención a las cuestiones de seguridad y privacidad (es decir, el uso de plataformas de mensajería seguras con cifrado de extremo a extremo).

Los asistentes virtuales pueden implementar sistemas conversacionales para que la interacción con los pacientes sea más natural. Los sistemas conversacionales en escenarios de atención médica complejos, como la gestión de enfermedades, deben ser capaces de poder comprender oraciones complejas utilizadas durante la interacción. La adaptación de nuevos métodos con el procesamiento de lenguaje natural (NLP, por su nombre en inglés, Natural Language Processing) puede aportar una mejora a la arquitectura del asistente virtual. Los word embeddings (incrustación de palabras) se han utilizado ampliamente en NLP como entrada en las redes neuronales. Tales word embeddings pueden ayudar a comprender el objetivo final y las palabras clave en una oración. Por ello, en esta tesis estudiamos el impacto de diferentes word embeddings entrenados con corpus generales y específicos utilizando el entendimiento del lenguaje natural conjunto (Joint NLU, por su nombre en inglés, Joint Natural Language Understanding) en el dominio de la medicación en español. Los datos para entrenar el modelo NLU conjunto se generan usando plantillas. Dicho modelo se utiliza para la detección de intenciones, así como para el slot filling (llenado de ranuras). En este estudio comparamos word2vec y fastText como word embeddings y ELMo y BERT como modelos de lenguaje. Para entrenar los embeddings utilizamos tres corpus diferentes: los datos de entrenamiento generados para este escenario, la Wikipedia en español como dominio general y la base de datos de medicamentos en español como datos especializados. El mejor resultado se obtuvo con **el modelo ELMo entrenado con Wikipedia en español**.

Dotamos al asistente virtual de capacidades de gestión de medicamentos basadas en NLP. En consecuencia, se analiza el impacto del etiquetado de slots y la longitud de los datos de entrenamiento en modelos NLU conjuntos para escenarios de gestión de medicamentos utilizando asistentes virtuales en español. En este estudio definimos las intenciones (propósitos de las oraciones) para escenarios centrados en la administración de medicamentos y dos tipos de etiquetas de slots. Para entrenar el modelo, generamos cuatro conjuntos de datos, combinando oraciones largas o cortas con slots largos o cortos. Para el análisis comparativo, elegimos seis modelos NLU conjuntos (SlotRefine, stack-propagation framework, SF-ID network, capsule-NLU, slot-gated modeling y joint SLU-LM) de la literatura existente. Tras el análisis competitivo, se observa que el mejor resultado se obtuvo utilizando **oraciones y slots cortos**. Nuestros resultados sugirieron que los modelos NLU conjuntos entrenados con **slots cortos** produjeron mejores resultados que aquellos entrenados con slots largos **para la tarea de slot filling**.

En definitiva, proponemos una arquitectura de microservicios genérica válida para cualquier tipo de gestión de enfermedades crónicas. El **prototipo genérico** ofrece un asistente virtual operativo para gestionar información básica y servir de base para futuras ampliaciones. Además, en esta tesis presentamos **dos prototipos especializados** con el objetivo de mostrar cómo esta nueva arquitectura permite cambiar, añadir o mejorar diferentes partes del asistente virtual de forma dinámica y flexible. El primer prototipo especializado tiene como objetivo ayudar en la gestión de la medicación del paciente. Este prototipo se encargará de recordar la ingesta de medicamentos a través de la creación de una comunidad de apoyo donde los pacientes, cuidadores y profesionales sanitarios interactúen con herramientas y servicios útiles ofrecidos por el asistente virtual. La implementación del segundo prototipo especializado está diseñada para una enfermedad crónica específica, la psoriasis. Este prototipo ofrece teleconsulta y almacenamiento de fotografías.

Por último, esta tesis tiene como objetivo **validar la eficacia del asistente virtual** integrado en las plataformas de mensajería, destinado al cuidado de la salud. Por ello, esta tesis incluye la evaluación de los dos prototipos especializados. El primer estudio tiene como objetivo **mejorar la adherencia a la medicación** en pacientes con diabetes mellitus tipo 2 comórbida y trastorno depresivo. Para ello, se diseñó y posteriormente se realizó un estudio piloto de nueve meses. En el estudio analizamos la Tasa de Posesión de Medicamentos (MPR, por su nombre en inglés, Medication Possession Ratio), obtuvimos la puntuación del Cuestionario sobre la Salud del Paciente (PHQ-9, por su nombre en inglés, Patient Health Questionnaire) y medimos el nivel de hemoglobina glicosilada (HbA1c), en los pacientes antes y después del estudio. También realizamos entrevistas a todos los participantes. Un total de trece pacientes y cinco enfermeras utilizaron y evaluaron el asistente virtual propuesto. Los resultados mostraron que, en promedio, **la adherencia a la medicación de los pacientes mejoró**. El segundo estudio tiene como objetivo **evaluar** un año de uso entre el asistente virtual y pacientes con psoriasis y dermatólogos, y **el impacto en su calidad de vida**. Para ello se diseñó y realizó un estudio prospectivo de un año de duración con pacientes con psoriasis y dermatólogos. Para medir la mejora en la calidad de vida, en este estudio analizamos los cuestionarios de Calidad de Vida de los Pacientes con Psoriasis (PSOLIFE, por su nombre en inglés, Psoriasis Quality of Life) y el Índice de Calidad de Vida en Dermatología (DLQI, por su nombre en inglés, Dermatology Life Quality Index). Además, realizamos encuestas a todos los participantes y obtuvimos el número de consultas médicas realizadas a través del asistente virtual. Se incluyeron en el estudio un total de 34 participantes (30 pacientes diagnosticados con psoriasis moderada-grave y cuatro profesionales sanitarios). Los resultados mostraron que, en promedio, **la calidad de vida mejoró**.

Scientific Contributions

The research work for this thesis was carried out in the Department of Electrical Engineering and Communications, School of Engineering and Architecture, University of Zaragoza. The main contributions presented in this thesis in Chapters 2-6 are based on the following publications:

Publications in International Journals (JCR indexed)

- ★ **Roca, S.**, Sancho, J., García, J., & Alesanco, Á. (2020). Microservice chatbot architecture for chronic patient support. *Journal of biomedical informatics*, 102, 103305.
- ★ **Roca, S.**, Rosset, S., García, J., & Alesanco, Á. (2022). A study on the impacts of slot types and training data on Joint Natural Language Understanding in a Spanish medication management assistant scenario. *Sensors*, 22, 2364.
- ★ **Roca, S.**, Lozano, M. L., García, J., & Alesanco, Á. (2021). Validation of a virtual assistant for improving medication adherence in patients with comorbid type 2 diabetes mellitus and depressive disorder. *International Journal of Environmental Research and Public Health*, 18, 12056.
- ★ **Roca, S.**, Almenara, M., Gilaberte, Y., Gracia-Cazaña, T., Morales Callaghan, A.M., Murciano, D., García, J., & Alesanco, Á. (2022). When virtual assistants meet teledermatology: validation of a virtual assistant to improve the quality of life of psoriatic patients. *International Journal of Environmental Research and Public Health*, 19, 14527.

Publications in International Conferences

- ★ **Roca, S.**, Hernández, M., Sancho, J., García, J., & Alesanco, Á. (2019, September). Virtual assistant prototype for managing medication using messaging platforms. In *Mediterranean Conference on Medical and Biological Engineering and Computing* (pp. 954-961). Springer, Cham.

- ★ **Roca, S.**, Rosset, S., García, J., & Alesanco, Á. (2020, November). Evaluation of Embeddings in Medication Domain for Spanish Language Using Joint Natural Language Understanding. In *European Medical and Biological Engineering Conference* (pp. 510-517). Springer, Cham.
- ★ **Roca, S.**, Rosset, S., García, J., & Alesanco, Á. (2021, July). A Natural Language Processing-based Microservice System for Healthcare Scenarios. In *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI).* Extended Abstract.
- ★ **Roca, S.**, Gilaberte, Y., García, J., & Alesanco, Á. (2021, July). Multi-platform Conversational System for Patients Information Collecting in Dermatological Scenarios. In *2021 IEEE EMBS International Conference on Biomedical and Health Informatics (BHI).* Extended Abstract.

Publications in National Conferences

- ★ Almenara, M., **Roca, S.**, Murciano, D., Navarro, A., Gracia, T., Morales. A.M., García, J., Alesanco, Á., & Gilaberte, Y. (2022, January). Asistentes virtuales en dermatología: chatbot Elena para pacientes con psoriasis. In *2022 7º Congreso de Psoriasis.* Oral communication.

Related Publications in International Journals (JCR indexed)

- ★ Hernando, D., **Roca, S.**, Sancho, J., Alesanco, Á., & Bailón, R. (2018). Validation of the Apple Watch for Heart Rate Variability Measurements during Relax and Mental Stress in Healthy Subjects. *Sensors*, 18, 2619.

Contents

List of Figures

List of Tables

List of Codes

Abbreviations

ACL	Access Control List
AI	Artificial Intelligence
AIML	Artificial Intelligence Markup Language
API	Application Programming Interfaces
BERT	Bidirectional Encoder Representations from Transformers
BLSTM	Bi-directional Long Short-Term Memory
BSA	Body Surface Area
CBOW	Continuous Bag of Words
CEICA	Comité de Ética de la Investigación de la Comunidad Autónoma de Aragón
COPD	Chronic Onstructive Pulmonary Disease
DLQI	Dermatology Life Quality Index
EHR	Electronic Health Record
ELMo	Embeddings from Language Models
FHIR	Fast Health Interoperability Resources
FN	False Negatives
FP	False Positives

GCP	Good Clinical Practice
GDPR	General Data Protection Regulation
HbA1c	Glycosylated Hemoglobin
HF	Heart Failure
HL7	Health Level Seven International
HTA	Hypertension
HTTP	Hypertext Transfer Protocol
HTTPS	Hypertext Transfer Protocol Secure
IGA	Investigator Global Assessment
IHD	Ischemic Heart Disease
IOB	Inside, Outside, Beginning
IoT	Internet of Things
JSON	JavaScript Object Notation
JWT	JavaScript Object Notation Web Token
LM	Language Modeling
MAUQ	mHealth App Usability Questionnaire
MAUQ_U	mHealth App Usability Questionnaire Usefulness
MEMS	Medication Event Monitoring System
MMAS-8	Eight-item Morisky Medication Adherence Scale
MPR	Medication Possession Ratio
NLG	Natural Language Generation
NLP	Natural Language Processing

NLU	Natural Language Understanding
OCR	Optical Character Recognition
PASI	Psoriasis Area Severity Index
PHQ-9	Patient Health Questionnaire
PHR	Personal Health Record
PSOLIFE	Psoriasis Quality of Life
RNN	Recurrent Neural Network
SD	Standard Deviation
SLU	Spoken Language Understanding
SUS	System Usability Scale
TLS	Transport Layer Security
TN	True Negatives
TP	True Positives
TSQM	Treatment Satisfaction Questionnaire for Medication
XML	eXtensible Markup Language

Chapter 1

Introduction

Advances in eHealth and medical telemonitoring have enabled healthcare professionals to provide follow-up and services remotely. Technology is continuously changing and evolving, and users' habits evolve with it. Hence, the tools used in eHealth should become with the new technology developed to be aligned with users' daily lives. Thereby, numerous challenges emerge to be addressed to provide remote healthcare following the development of technology. A virtual assistant (a.k.a. chatbot) is an autonomous intelligent computer program that automatically provides services conversing with the final user through diverse communication channels (e.g. messaging platforms, smartphone apps, etc.). Virtual assistants are currently a hot topic in eHealth scenarios. In fact, virtual assistants can help in eHealth scenarios in a wide variety of situations, such as those of patients in remote areas with limitations in healthcare facilities [1]. This chapter presents an introduction and an overview of this thesis. This thesis aims to contribute to the alignment between technology and remote healthcare.

1.1 Virtual assistants

In the last five years, fostered by the explosion of Artificial Intelligence (AI) and the enormous penetration in society of messaging platforms, virtual assistants have been gaining momentum in the eHealth world. Studies show that there exist positive results in support-virtual assistants for patients with breast cancer [2], with overall satisfaction of 93.95%. Virtual assistants can be used to help their users by playing many different roles, such as symptom checkers [3], medication reminders [4] or personal data gatherers [5]. Moreover, physicians agree with the idea that virtual assistants can help in most of the automatic simple tasks in healthcare scenarios [6]. However, their usability is limited by the algorithms behind them, their ability to share data, their scalability and the sense of security and privacy they are able to implement and transmit to their users.

One of the eHealth scenarios that is growing fast and is expected to grow even faster in the coming years is caring for patients with chronic diseases or comorbidities [7]. While research on patient-centred care has largely spawned from technical and clinical research traditions, it inevitably (not intentionally) concentrates mainly on medical staff and supporting clinical work. Traditionally, healthcare and eHealth are focused on acquiring data and information from patients and transferring them into healthcare contexts, not the other way around. Virtual assistants are the perfect tools to make possible this turnaround, being aligned with chronic patients' needs, interacting with them in the same way that patients communicate with their friends and relatives through their favourite messaging platforms. Nevertheless, current eHealth virtual assistants are not taking advantage of this great potential and are only being used in a very generic manner (nutritional disorders and neurological disorders are the areas most tracked [8]). Although these are interesting and necessary approaches, they do not realize the full potential of virtual assistants which remain quite limited for chronic patient support. To unleash this potential, virtual assistant software architecture should be grounded on three pillars: scalability that enables easy growth, standard data models that foster data sharing, and standard conversational modeling that facilitates the conversation between the user and the virtual assistant.

1.1.1 Software architectures

Scalability provides a growing ecosystem of services available to patients while they evolve (e.g. a chronic patient develops a new comorbidity). Thus, the virtual assistant would be able to provide them with solutions without the need of using other virtual assistants or even the need to go back to a specific mobile app. In other words, due to the rise in patients' comorbidities, eHealth virtual assistants should be scalable to adapt to new patients' needs, providing them with new tasks and services over time. The lack of scalability in these systems can hamper the virtual assistant growth, i.e. the offering of new services to healthcare actors (patients, doctors, health professionals in general).

To solve the problem of scalability, the virtual assistant architecture must be modular and flexible. Microservices are an architectural paradigm which emphasizes modular, lightweight services with a high degree of cohesion. This is in contrast to monolithic applications where tightly integrated components implement the applications' functionality and changing requirements affects the system as a whole. In contrast, microservices can be developed, deployed and scaled independently of other services that make up the system. The basis of microservices is to split a single application into a set of small services, each running its own process. Microservices communicate through well-defined interfaces and standard lightweight protocols such as Hypertext Transfer Protocol (HTTP) and do not need to use the same development languages or platforms. Thus, they are the perfect choice for a modular and continuously growing architecture to support the needs of chronic patients. New functionalities addressing new health conditions can be added as new microservices, fostering a care ecosystem and open to contributions from collaborative developers.

The choice of an open model for patient data is crucial to endow the virtual assistant architecture with an open nature for data sharing, avoiding the creation of isolated data silos, so dangerous for eHealth expansion [9]. In eHealth environments, the ability of standard data sharing enables all the gathered data to be integrated in other bigger structures related with the patient such as his/her Electronic or Personal Health Record (EHR/PHR). With this in mind, all the data exchanged with the virtual assistant regarding the patient's conditions would end up, eventually, in his/her EHR, enriching it and enabling carers (doctors, nurses, etc.) to complete their health assessment of the patient with important data gathered on a regular basis. Fast Healthcare Interoperability Resources (FHIR) is a next generation standards framework created by Health Level Seven International (HL7) [10]. The FHIR standard defines a list of data models which can represent a wide range of healthcare related features, both clinical and administrative. Instances of these data models, which are named resources, are used to exchange and/or store the data using different serialization formats (e.g. eXtensible Markup Language (XML) or JavaScript Object Notation (JSON)). These resources can easily be assembled into working systems making them suitable for use in a wide variety of contexts e.g. mobile phone apps, EHR-based data sharing, server communication in large institutional healthcare providers, etc [11]. All these features make it the perfect choice for data sharing and storing in the microservice-based virtual assistant architecture.

1.1.2 Communication channel

Messaging platforms have redefined the way we people communicate and interact every day. The usage of messaging platforms is a must in our interaction with other people. The number of yearly active users in chatting applications has increased by 6.2% the last year, with the usage reaching to three thousand million people worldwide by the end of 2021 [12]. Recently, the combination of AI with messaging platforms has generated a new revolution in the user's care: the emergence of virtual assistants. Virtual assistants use messaging platforms to communicate directly with users.

1.1.3 Artificial intelligence

Virtual assistants interact with the users using text-based conversations. Nowadays, virtual assistants are used for numerous tasks, such as booking plane flights, ordering food and learning new languages. The benefits offered by such virtual assistants can significantly improve the effectiveness of healthcare services, especially when offered to elderly patients and children [13]. Patients can interactively obtain support, information or medical diagnosis by chatting with a virtual assistant. For some healthcare services like reminding the medicine intake, such virtual assistants can conveniently complement a health professional or personal caretaker. These healthcare virtual assistants can perform interesting tasks for users, but without a correct understanding of the circumstances under which the services are to be offered, patients may obtain wrong information or wrong monitoring. In other words, the virtual assistants should be intelligent enough to understand the scenario and the services desired by the users. Hence, the first and most important aim of the virtual assistant is to correctly understand the requests of the patients.

Natural language processing (NLP) is a discipline that works in the processing and understanding of virtual assistant-user interactions. Thanks to NLP, the virtual assistant should be able to process the user input and to obtain a response as consistent as possible with the user conversation. Nowadays, there are many advances in NLP that offer tools and standards to allow fluent interactions with users. Among them, one appears to have a widespread acceptance, Artificial Intelligence Markup Language (AIML) [14]. AIML is a language based on XML that serves for the development of software agents and adds the capability to communicate with users in natural language. Thanks to the abilities of NLP, it is possible to obtain valuable health data from the user-virtual assistant conversation.

The dialogue system (a.k.a. conversational agent) is a computer program that uses text, speech, gestures, and other communication tools to interact with users. The components of a dialogue system are shown in Figure 1.1. First, the system receives input from the user (voice or text). Automatic speech recognition extracts the user's sentence from the speech signal. The Natural Language Understanding (NLU) identifies the domain the user is talking about from the text input. NLU is a subtopic of NLP, which is used to comprehend what an input text means and act consequently. NLU also tries to predict the intention of the user and the keywords that the user is saying in a sentence. The tasks that the NLU performs are intent detection and slot filling. Intent detection tries to label the sentence based on predefined intents [15]. Slot filling can be defined as the action of tagging the words in a sentence with the slot types. Moreover, NLU systems have a learning approach that is based on the statistical representation of each word. This representation is obtained using embeddings. Embeddings have been widely used in NLU as the input to neural networks. Such embeddings can help in the understanding of the final objective and the keywords in a sentence. There exist two types of embeddings: static word embeddings (e.g. word2vec [16] or fastText [17]) which generates the same embedding for the same word regardless of the context, and contextualized word embeddings (a.k.a. embeddings from language models) (e.g. Embeddings from Language Models (ELMo, a deep contextualized word representations) [18] or Bidirectional Encoder Representations from Transformers (BERT) [19]) which captures the word semantics in different contexts. Briefly, word embeddings are the representation of a word that is in a vocabulary into a vector of numerical values. The input of these embeddings is a text called corpus that serves to train the embedding. Corpus can be defined as a collection of structured texts used to do statistical analysis.

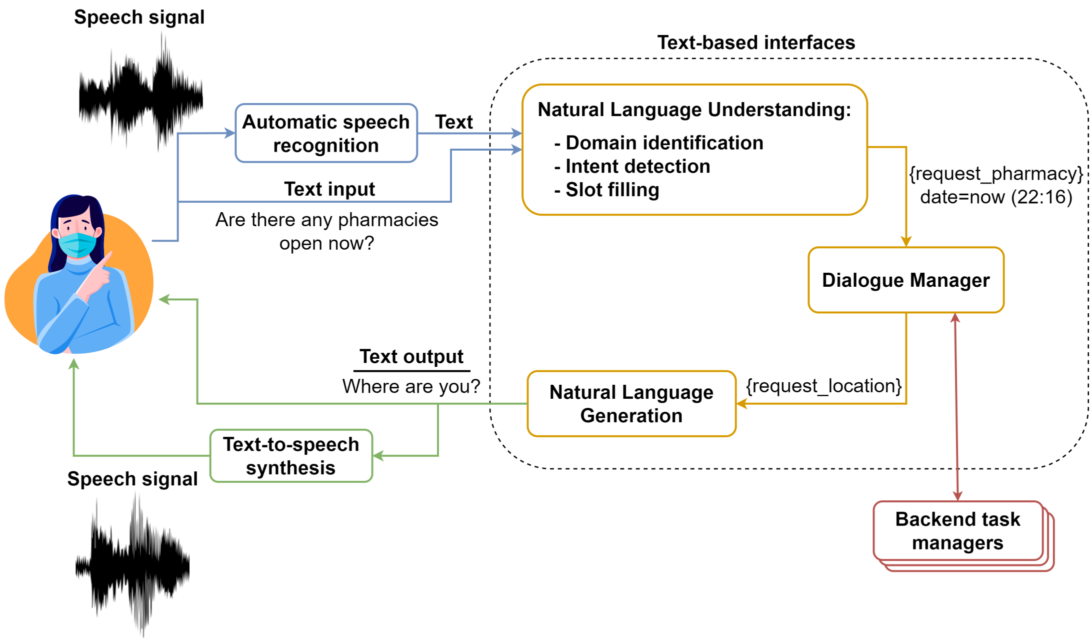


Figure 1.1: Dialog system diagram.

Once the intent and the slots are predicted, the dialogue manager decides which back-end tasks need to perform based on the user requirements, and determines the answer the user will receive. The Natural Language Generation (NLG) generates the text output of the system and, if required, the text-to-speech synthesis generates the speech signal to send to the user. Since our scenario is a text-based interface, the components of the dialogue system are NLU, dialogue manager, and NLG.

1.1.4 Challenges

A notorious rise in virtual assistant platforms has recently taken place in numerous fields, such as health tracking, help-manager in websites, and information providers. A virtual assistant performs automated tasks, such as sending alarm notifications, automatically storing information from a web server, or daily weather reports. There are a wide variety of ways a user can interact with a virtual assistant. For example, a virtual assistant can be accessed through a smartphone application or a web platform.

The interest in messaging platforms is increasing due to the fact that users use them every day. In essence, messaging platforms offer an interesting communication channel where users can interact with virtual assistants easily.

How a virtual assistant should be built does not have a direct answer. Modularity and flexibility appear to be important factors when building a virtual assistant since the services provided by the virtual assistant change and increase independently. Non-monolithic architectures such as microservices seem a good fit for virtual assistant systems.

eHealth provides healthcare services remotely. These services help in the patients' follow-up and monitoring of their diseases. Improving patients' health has a direct impact on saving healthcare costs and on the well-being of society. Research is continuously developing new techniques for eHealth, with the aim to improve patients' health and provide new tools for healthcare professionals. eHealth relies on technology. Challenges appear when technology evolves, and telemonitoring systems for healthcare systems should evolve accordingly. Furthermore, the development of new healthcare technology does not have a direct inclusion in the healthcare system. These new systems need to be evaluated in clinical cases, in order to validate their usability and utility in helping patients and healthcare systems.

This general scenario is shown in Figure 1.2. Remote patients, outside the hospital, contact healthcare professionals through a virtual assistant. The patients can interact with the virtual assistant for a wide number of possible options, such as storing their health records and asking questions, among others. On the other hand, healthcare professionals can use the virtual assistant to observe the patients' evolution, reply to patients' questions, or track relevant information over time about a patient. Specifically, users use the services offered by the virtual assistant with their technology devices (smartphones, tablets, or computers). The techniques proposed in this thesis are described for eHealth scenarios, but they can be generalized for most current virtual assistant architectures.

Figure 1.2: General scenario.

1.2 Thesis approach, hypothesis and objectives

The thesis approach is to make research contributions to the field of eHealth and mHealth scenarios, with special emphasis on virtual assistant systems and microservice architectures. Specifically, this research investigates how a new virtual assistant architecture based on microservices can be helpful in home-based caring, with special consideration for chronic diseases.

The main hypothesis of this thesis is that by designing and developing a virtual assistant (menu-based or human-like interaction) with a scalable architecture relying on medical standards, we can improve medication adherence and the quality of life of chronic patients with diseases such as psoriasis, diabetes, and depression.

The following objectives have been identified to sustain the hypothesis of the thesis:

1. To conduct reviews of the state of the art in general aspects of virtual assistants and their integration in healthcare systems, software architectures (with special attention to microservice architectures), natural language processing (embeddings and natural language understanding models and their performance in the biomedical domain), and clinical studies conducted on telemonitoring scenarios.

2. To design a new microservice-based virtual assistant architecture that provides modularity to give support and services in healthcare environments focused on chronic patient care. The proposed architecture should address the data storage, interaction, and security challenges.

3. To investigate and adapt the human-like interaction to provide a more specific model for the medication management domain. In order to address this objective, the following evaluations should be performed:

    - The study on the impact of using general and specific corpora to learn different configurations of embeddings. The test evaluation should be performed with data obtained from users.
    - The study on the impact of slot tagging and dataset length on intent detection and slot filling for the medication domain in Spanish and the analysis of how they can influence the performance of the NLU models.

4. To develop a generic prototype with common functionalities to provide general support for chronic diseases. Furthermore, to develop two specialized prototypes for different conditions (specifically, medication management and teleconsulting).

5. To evaluate the proposed virtual assistant architecture in different clinical environments in order to observe its modularity. Specifically, to validate the virtual assistant for improving medication adherence in diabetes and depressive patients, and to validate the virtual assistant for improving the quality of life of psoriatic patients.

1.3 Research context

This thesis, entitled "Development and Evaluation of a Microservice-Based Virtual Assistant for Chronic Patients Support" and supervised by Dr. Álvaro Alesanco Iglesias, has been carried out within the framework of Telemedicine and eHealth research line of the Communications Networks and Information Technologies (CeNIT) Group of the Aragón Institute of Engineering Research (I3A), within the Mobile Network Information and Communication Technologies doctoral program of the University of Zaragoza, Spain.

The research was developed within the funded project "Microservices and ontologies in the construction of an architecture for the secure and private management of information in the personalized follow-up of patients with psoriasis" (TIN2016-76770-R) and the doctoral grant (BES-2017-082017).

Additionally, a student mobility traineeship has been granted to the PhD candidate: Erasmus+ Student Placement European Programme and Campus Iberus, in the 2018/ 2019 academic year.

This thesis includes two research stages in collaboration with the following research groups:

- The Information Langue Écrite et Signée Group (ILES) at Laboratoire Interdisciplinaire des Sciences du Numérique (LISN) and Centre National de la Recherche Scientifique (CNRS) at the Université Paris-Saclay (France) through Dr. Sophie Rosset, Senior Researcher in Natural Language Processing discipline.

- The eHealth Services Research Group (eHSRG) at the University of Tasmania (Australia) at the School of Information and Communication Technology through Dr. Paul Turner, director of the research group.

Finally, this thesis has two collaborations with clinical partners:

- The Department of Dermatology at Miguel Servet University Hospital and Instituto de Investigación Sanitaria (IIS) Aragon (Spain) through Dr. Yolanda Gilaberte, head of the department.
- The Primary Healthcare Center Las Fuentes Norte (Spain) through Dr. María Luisa Lozano.

1.4 Thesis outline

The reminder of this document is organized as follows:

1. Chapter 2 presents the state of the art related to the main topics of this thesis. More specifically, it includes a review of virtual assistants and software architectures. It also discusses related works about Natural Language Understanding and clinical studies.
2. Chapter 3 describes the proposed system architecture, where a detailed explanation of the user interaction is included. An analysis of the modularity, standardization and security is also detailed.
3. Chapter 4 explores the Natural Language Understanding models for the medication management domain. It includes an embedding evaluation and an evaluation of the slot types and the training data.
4. Chapter 5 includes the development of a generic and two specialized prototypes. The first specialized prototype aims to manage the medication with a menu-based interaction and complemented with a human-like interaction. On the other hand, the second specialized prototype provides teleconsulting and image storage with a menu-based interaction.

5. Chapter 6 evaluates the system in two clinical scenarios. First, patients with comorbid type 2 diabetes mellitus and depressive disorder evaluated the medication adherence prototype. Then, dermatologists and patients with psoriasis evaluated the psoriasis prototype. The discussion about the results from both studies is also included in this chapter.

6. Chapter 7 presents the research objectives achieved, contributions, conclusions of the thesis and future lines of research.

Chapter 2

State of the Art

A scalable and modular software architecture that supports virtual assistants is the core of our proposal. This technology along with NLP is used in this thesis to solve the challenges observed in healthcare and telemonitoring scenarios. Moreover, clinical studies are used to validate the proposed system for different chronic conditions. This chapter presents and discusses related works with the aim to study previous research and how these technologies can be successfully applied in eHealth scenarios.

2.1 Virtual assistants

Nowadays, there is a wide range of virtual assistants offering help to patients in eHealth scenarios. Ana [20], for example, is a self-anamnesis mobile virtual assistant application for patients, where a conversational user interface is used to simulate the patient-therapist conversation. Another example is MamaBot [21] which is an AI-virtual assistant for providing support to mothers, pregnant women, and families with young children. MamaBot provides help and instructions in relevant situations through Telegram [22] (a messaging platform), that does not have end-to-end encryption and, therefore, it does not provide a secure communication channel. SWITCHes [23] is a system which has an AI-powered health virtual assistant for interaction with overweight users and it provides real-time data reception and transmission to their web server through a smartphone app and a web-based dashboard. CustomRXBot [13] is a virtual assistant developed to help healthcare professionals with custom prescriptions for dermatology through Slack [24].

An interesting approach of virtual assistants is the role of medication reminders because they could help in adherence to medication. eMMA [25] is a virtual assistant-like smartphone app to manage patient medication through a virtual assistant focused on supporting a consistent flow of medication data to increase medication adherence. Roborto [26] is a virtual assistant model system that tries to improve patients' medication adherence and inform healthcare providers about patients' conditions and improvements, using Telegram as a communication channel.

As shown in the examples explained before, most of the developed virtual assistants use a smartphone app to allow the communication between the virtual assistant and the users. However, a quarter of all downloaded smartphone applications are abandoned after the first usage [21]. Due to this fact, an alternative approach, without the usage of exclusively dedicated applications, should be used (such as messaging platforms). Developing virtual assistants that use messaging platforms as communication channels helps to provide new tools and functionalities without any update from the user side: virtual assistants are programs running in servers completely independent from the user application. Thus, any update in the virtual assistant is not going to affect the user smart devices. Furthermore, any virtual assistant system needs to provide very strict security measures to protect personal user information during all the message exchanges in healthcare scenarios, to protect the users' privacy.

The categorization of virtual assistants based on different platforms is given in Table 2.1.

Platform	Target condition	Related works
Mobile application	Self-anamnesis in music therapy	Denecke et al., 2018 [20]
	Overweight	Huang et al., 2018 [23]
	Medication adherence	Tschanz et al., 2018 [25]
Telegram	Mothers, pregnant women, and families with young children	Vaira et al., 2018 [21]
	Medication adherence	Fadhil, 2018 [26]
Slack	Dermatology	Alesanco et al., 2017 [13]

Table 2.1: Categorization based on different platforms.

2.2 Software architectures

Some research works propose the use of software architectures for eHealth scenarios which are not primarily intended for them. Catarinucci et al. [27] proposed an Internet of Things (IoT)-Aware Architecture that uses an IoT smart gateway (containing a two-way proxy, management application and secure access manager) to connect the hybrid sensing network with the user interfaces. Hill et al. [28] proposed a microservice architecture for an Internet of Things healthcare scenario to provide future requirements of scalability and resilience as the failure of a service should not adversely affect the overall system. The fundamental use cases of care provision have each action directly designed as a microservice in a microservice architecture. The microservices are described in their work, but no technical details are given about the link between the proposed system architecture and the microservices. Furthermore, no privacy protocols are studied. Another approach by O'Brien et al. [29] describes a microservice-based platform that uses activity trackers to provide a monitoring solution for health-related data. This is an interesting approach providing elasticity and scalability thanks to the microservices, but no security features are taken into consideration and there is no information about the data standardization for medical data storage. Ali et al. [30] presented an IoT platform in which a set of recommendations microservices is proposed for depressive disorders. A very attractive microservice model has been implemented, but the authors do not provide information about user interaction with the system and security issues.

The interest in using virtual assistants in eHealth environments is growing and attracting a lot of attention [3, 31–34]. Recent advances in virtual assistants show that they can improve the efficiency of healthcare delivery by performing clinical tasks that can be automated. Fadhil et al. [35] propose an AI-virtual assistant for delivering support to nutrition education. From this work, it can be concluded that virtual assistants have a lot of advantages in the eHealth domain, both for healthcare providers and patients. It uses a web application and a Telegram Application Programming Interface (API), but detailed information about the architecture components is not provided. Other healthcare virtual assistants like HealthBot [36] and Your.MD [31] are symptom checkers, providing the user with medical advice and useful tips about different medical conditions. Again, no hint about their underlying architecture is provided. Florence [4] is a virtual assistant that provides patients with medication reminders and a health tracker using patient data gathering, but without providing end-to-end encryption. These virtual assistant applications follow the same pattern: they are tailored to a specific medical condition or activity and no implementation details are provided, making it hard to evaluate their potential in terms of modularity, standard data management and standard virtual assistant-user interaction. Among those virtual assistant proposal that provide implementation details, Augellio et al. [37] propose a web-based infrastructure for virtual assistants with a modular knowledge base. Even though the division into modules is of interest, the system needs an extra effort to activate and deactivate the modules and to reload the core to apply changes upon the modules. Yan et al. [38] present a generic architecture for a virtual assistant framework built on top of a serverless computing platform. Although the approach is very interesting due to its decentralized nature, its serverless orientation together with the need to rely on IBM Watson services for virtual assistant-user interaction makes it hard to apply in healthcare scenarios.

Other research works have explored generic virtual assistant architectures. These architectures are composed of intent classification, entity recognition, candidate response generator and response selector [39]. Huang et al. [23] present an example of this type of virtual assistant architecture in healthcare scenarios, where a virtual assistant is developed for weight control and health promotion. The work focuses on how the message is processed to obtain the user response, rather than giving more information on how the software distribution is addressed to give flexibility and modularity to the system.

A comparative summary of related work in healthcare systems based on their architectures is shown in Table 2.2.

Architecture	**Related works**	**System/Purpose**
IoT[1] smart gateway	Catarinucci et al., 2015 [27]	Internet of Things
Microservices	Hill et al., 2017 [28], Ali et al., 2017 [30]	Internet of Things
	O'Brien et al., 2018 [29]	Activity tracker
Serverless	Yan et al., 2016 [38]	Virtual assistant
Web	Fadhil et al., 2017 [35], Augellio et al., 2011 [37]	Virtual assistant
Conversational approach	Rahman et al., 2017 [39], Huang et al., 2018 [23]	Virtual assistant
No provided	Symptomate, [3], Your.MD [31], Ada [32], Mediktor [33], HealthTap [34], HealthBot [36], Florence [4]	Virtual assistant

[1] Internet of Things.

Table 2.2: Categorization based on different architectures.

2.3 Natural Language Understanding

An interesting approach to studying the effect of the corpus in word embeddings was made by Wang et al. [40] where four different corpora in English were evaluated. Furthermore, Neuraz et al. [41] compared fastText with ELMo for different tasks in the clinical domain in French. Ghannay et al. [42] compared five different embeddings trained in English and French in five benchmark corpora for spoken language understanding. Moreover, Sushil et al. [43] studied the use of further language model pretraining, lexical match algorithms, supplementing lexical retrieval, and trained retriever modules in the clinical domain by finetuning BERT models; they observed that the proposed extensions did not show significant improvement and that methods should be developed to augment fundamental domain knowledge from textual corpora. There are a few notable approaches with word embeddings in Spanish medical domain. Segura-Bedmar et al. [44] in their work, proposed an approach to simplify Drug Package Leaflets. Soares et al. [45]

evaluated fastText in two different datasets. Further research in the evaluation of word embeddings in Spanish should be considered.

Furthermore, in previous years, several studies were carried out to improve and evaluate the performance of NLU tasks. Former works evaluated the impact of using different techniques to improve slot filling [46–48] and intent detection [49, 50]. Interestingly, research in joint NLU (jointly learning both intent detection and slot filling) achieved better results in both tasks [51, 52]. Moreover, Wang et al. [53] proposed a new joint model to improve the slot tagging task by combining sequence labeling with a deep bidirectional transformer.

Recent works focused on the importance of selecting the data to train the models in a low-data regime, with researchers observing that the selection criteria can have a strong influence on the performance of the model in the scenario of slot filling [54]. The data used in low-data resource scenarios, such as the biomedical domain, are generated in order to train the models, which could have different characteristics. Furthermore, the decision of slot tagging is not always direct. In fact, there are multiple ways to tag the information in a specific context.

The categorization of related work based on different methods for embeddings and NLU tasks is given in Table 2.3.

NLU[1]	Methods	Related works
Embeddings	Word2vec	Wang et al., 2018 [40], Ghannay et al., 2020 [42], Segura-Bedmar et al., 2017 [44]
	fastText	Neuraz et al., 2019 [41], Ghannay et al., 2020 [42], Soares et al., 2019 [45]
	GloVe	Wang et al., 2018 [40], Ghannay et al., 2020 [42]
	ELMo[2]	Neuraz et al., 2019 [41], Ghannay et al., 2020 [42]
	BERT[3]	Sushil et al., 2021 [43]
Tasks	Slot Filling	Wang et al., 2018 [46], Kobayashi et al., 2019 [47], Adel et al., 2017 [48], Dimovski et al., 2018 [54]
	Intent detection	Kim et al., 2016 [49], Wang et al., 2019 [50]
	Joint (slot filling and intent detection)	Goo et al., 2018 [51], Liu et al., 2016 [52], Wang et al., 2020 [53]

[1] Natural Language Understanding; [2] Embeddings from Language Models; [3] Bidirectional Encoder Representations from Transformers.

Table 2.3: Categorization based on different methods of NLU.

2.4 Clinical studies

Different diseases may benefit from the tools and services offered by virtual assistants. In particular, virtual assistants could be helpful for chronic diseases in home-based caring scenarios. The validation with patients and the analysis of their clinical variables should be studied to assess the improvement and help that virtual assistants may provide. This thesis includes such evaluations. Indeed, comorbid type 2 diabetes mellitus and depressive disorder are two major chronic conditions studied.

Diabetes is a major health issue, affecting 463 million people worldwide, and is expected to increase to 700 million by 2045 [55]. Patients with type 2 diabetes require a complex self-management of different aspects of their lives (e.g., exercise, diet, medication, or blood glucose control) [56, 57]. The relationship between diabetes and depression has been widely studied [58, 59] and it has been demonstrated that there exists a highly possible bidirectional relationship between them [60]. Depression is a common disease that affects more than 268 million people worldwide [61]. Patients with depression have mood fluctuations that affect their daily lives. These patients have problems in handling challenges at work, with family, or at school. This serious health disease, characterized by persistent sadness, is more frequent in patients with chronic diseases in general, and diabetes in particular [62], and is associated with poor medication adherence in patients with comorbidities [63].

Earlier studies have tested artificial conversational agents in relation to different diseases, such as cancer, chronic pain, or coronary heart disease [64–66]. Some of these studies examined the use of virtual assistants to improve adherence in patients with breast cancer, coronary artery disease, and other chronic diseases [67–69]. Several studies have suggested that mobile phone text messaging interventions could have considerable potential to improve medication adherence in patients with chronic diseases [70]. Studies focused on patients with type 2 diabetes mellitus using mobile applications showed that there was a moderate effect on glycemic control, with an overall difference in the mean HbA1c of -0.40% [71]. Most of these apps are focused on measuring blood-glucose using external devices in combination with a smartphone or a web-based interface. Previous depression studies have focused on providing patients with a therapeutic resource through conversational agents [72, 73]. Sarda et al. [74] studied the relationship between smartphone-sensing parameters and symptoms of depression in patients with diabetes. Bogner et al. [75] improved the medication adherence of patients with type 2 diabetes and depression using the Medication Event Monitoring System (MEMS).

Another chronic condition clinically evaluated in this thesis is psoriasis. Psoriasis is an immune-mediated disease that may cause visible signs of inflammation (e.g., raised plaques and scales on the skin) owing to inflammation caused in the body [76]. Psoriasis affects 125 million people worldwide (around 2 to 3 percent of the total population) [77]. Stress or anxiety, injury to the skin, hormonal changes, or certain infections or medications are factors that could trigger psoriasis flare-ups. Psoriasis can occur in any area of the body, such as hands, feet, nails, and the scalp [78].

The severity of psoriasis is measured based on the scales that assess physical symptoms, but it can also be measured by how the disease affects a person's quality of life. Nearly 60% of patients with psoriasis communicated their disease as a large problem in their everyday lives [79]. Roughly one-quarter of people living with psoriasis have moderate to severe cases [80]. Furthermore, psoriatic patients with moderate to severe cases experienced a greater negative impact on their quality of life [81]. Psoriasis also generates stress derived from social stigma and altered body image. To sum up, psoriasis has a substantial psychological and social impact on patients.

Teledermatology generally uses apps and home-centered platforms for skin image monitoring and image sharing, where dermatologists can track patients' responses to therapy or receive skin image information. The use of teledermatology helps review a large number of cases in a short time. Teledermatology also helps in the communication between specialists and patients. Better treatment outcomes are obtained when the communication between doctors and patients is improved [82]. The use of teledermatology seems to be gradually taking a central place in healthcare delivery [83]. Several apps exist for skin image monitoring, such as SymTrac™ Psoriasis [84], which stores photographs of the affected areas and tracks symptoms and quality of life over time; AI Psoriasis App: Manage and Care [85], which provides the severity rating of psoriasis from a skin image provided; and Imagine Skin Condition Tracker [86], an app that can track and compare photos over time to see the symptoms progress. There are also tools to improve psoriasis well-being like Kopa [87], which offers tips and tricks for handling symptoms; Claro [88], which helps enhance emotional well-being; and MiPsoriasis [89], which monitors psoriasis through questionnaires.

Earlier virtual assistants in dermatology, such as Custom-RXBot [13], proposed virtual assistant prototypes to guide personalized medicament design. Sager et al. [90] analyzed the use of bots in health misinformation environments on Reddit's dermatology forums. The bot posted prefabricated responses when misinformation was found. Former studies on improving medication adherence with virtual assistants or smartphone applications showed a positive effect on patient adherence [91, 92]. Domogalla et al. [93] validated a disease management smartphone app for improving the mental health of patients with psoriasis in the long term. Furthermore, online care compared with in-person care showed equivalent improvements in disease severity among patients with psoriasis [94]. The improvement of quality of life for online and in-person care was also studied, which found that the online model and the in-person care had similar enhancement for psoriatic patients [95]. A comparative analysis of tools and conversational agents for healthcare scenarios based on the disease or condition is shown in Table 2.4.

Disease/Condition	Related works	Platform	Purpose/Improvement
Cancer	Greer et al., 2019 [64]	Facebook Messenger	Promote well-being
Chronic pain	Hauser-Ulrich et al., 2020 [65]	Mobile application	Pain self-management
Coronary heart disease	Chow et al., 2015 [66]	Short Message Service	Cardiovascular risk factors
Breast Cancer	Chaix et al., 2019 [67]	Web, mobile application, Facebook Messenger	Medication adherence
Acute coronary syndrome	Quilici et al., 2013 [68]	Short Message Service	
Type 2 diabetes mellitus and depression	Bogner et al., 2012 [75]	Medication Event Monitoring System	
Type 1/type 2 diabetes	Sarda et al., 2019 [74]	Passive smartphone sensing application	Early detection of symptoms of depression
Depression and anxiety	Fitzpatrick et al., 2017 [72], Fulmer et al., 2018 [73]	Instant messenger application	Therapeutic resource
Psoriasis	SymTrac™ Psoriasis [84], AI Psoriasis App: Manage and Care [85], Imagine Skin Condition Tracker [86], MiPsoriasis [89]	Mobile application	Skin image monitoring and questionnaires
	Kopa [87], Claro [88], Domogalla et al., 2021 [93]	Mobile application	Well-being and mental health
	Svendsen et al., 2018 [92]	Mobile application	Medication adherence
	Armstrong et al., 2018 [94]	Web	Disease severity
	Armstrong et al., 2019 [95]	Web	Quality of life
Dermatology	Alesanco et al., 2017 [13]	Slack	Personalized medicament design
	Sager et al., 2021 [90]	Reddit	Health misinformation

Table 2.4: Categorization of tools and conversational agents based on the disease or condition.

2.5 Conclusions

In this chapter, virtual assistants from the existing literature for eHealth scenarios were discussed, and an alternative for the communication channel was proposed. A review of software architectures used to build virtual assistants was also included. In most cases, the details of the architecture, information about data standardization, and security concerns were not discussed. Moreover, we identified future research gaps in the NLU field, which we believe will be an important target to extend the current works and add newer and more advanced capabilities to the research area. In addition, previous clinical studies that used tools and conversational agents for eHealth scenarios were analyzed. The proposed ideas associated with the architecture, the conversational aspects of virtual assistants, and their implementation and evaluation have been addressed in Chapters 3–6.

Chapter 3

System Architecture

Modularity and standardization are key points in the development of systems that should adapt over time and are involved in medical scenarios, where the interoperability of the systems and the data is a must. Microservices have drawn attention in software architecture mainly for the modularity offered. This chapter describes the new microservice-based virtual assistant architecture proposed.

3.1 Microservice architecture

The application of a microservices-based architecture in the design of a virtual assistant architecture for healthcare scenarios is proposed in Figure 3.1. The proposed architecture is intended to provide new services and functionalities for evolving user needs. The logic based on microservices in the proposed architecture serves to process the user information and perform automatic tasks to provide scalability in a healthcare virtual assistant ecosystem. Furthermore, modularity is covered by the proposed microservice structure, allowing a new microservice to be built independently and without needing to modify the developed virtual assistant. Moreover, user communication with the virtual assistant architecture is carried out using messaging platforms. In a health environment scenario, the most important feature to consider while choosing the messaging platform is to guarantee the privacy of the data that is exchanged between the user and the virtual assistant. Signal [96], with its end-to-end encryption, is used as a messaging platform in this architecture.

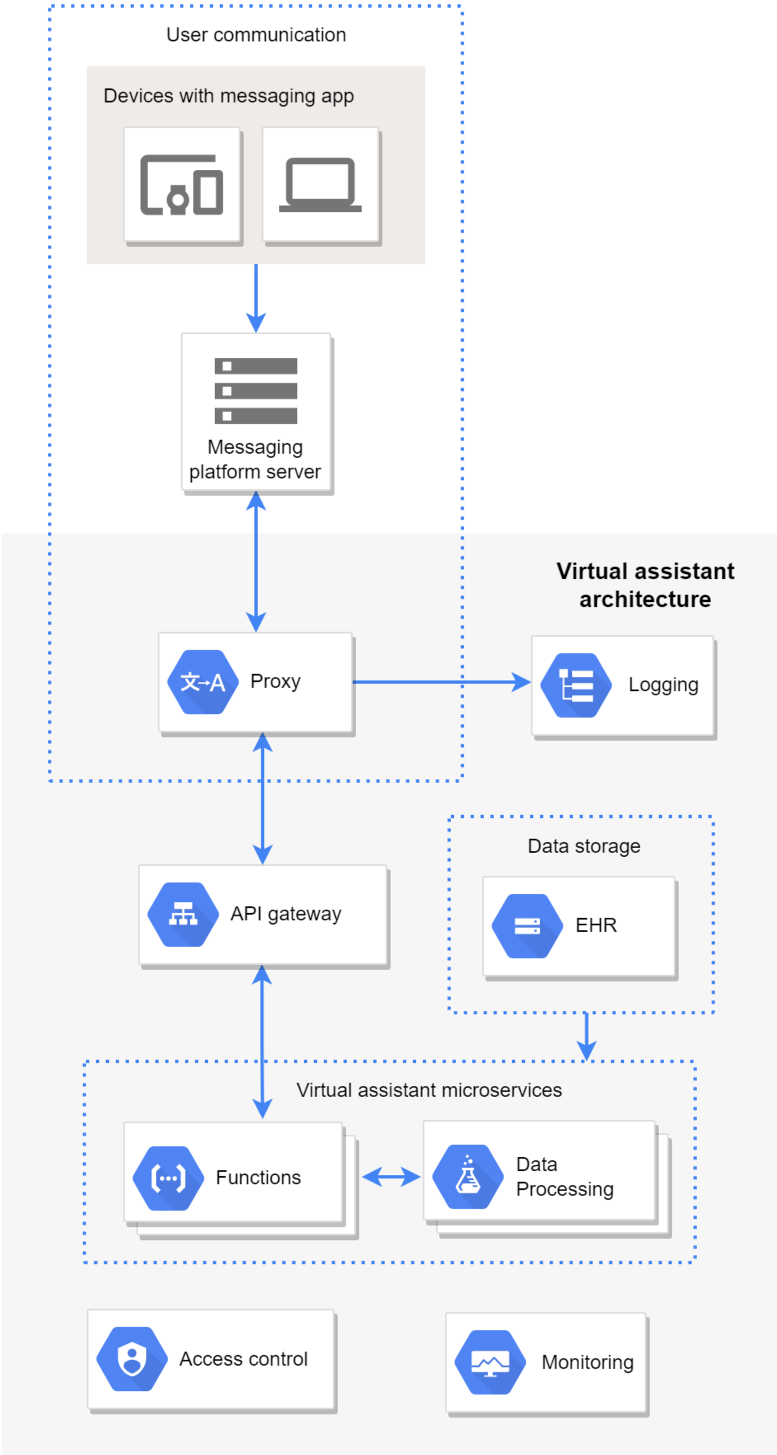


Figure 3.1: Overview of the virtual assistant internal structure.

The standardization of user data and the modeling of user conversations are added within the proposed microservice structure, to provide interoperability to the system. The structure of the architecture allows having different types of databases depending on the necessities of the virtual assistant storage. In order to be able to manage the entire amount of data that is generated and provide interoperability between different healthcare systems, the architecture supports EHR for medical data storage. Furthermore, each microservice has its own small database to store all the information needed during the run time.

The proposed architecture includes microservices intended for the correct functioning of the system (monitoring, authentication, and logging). One microservice is dedicated to checking if the system is running correctly. Another microservice is dedicated to providing authentication between all the microservices running in the platform. Finally, there is a microservice that provides logging facilities. The number of microservices can be increased according to the needs that arise within the system.

The coordination and distribution of the system tasks among the microservices of the architecture is essential to be able to perform a fluent user-virtual assistant interaction. The architecture has two core microservices: the proxy and the API gateway. The proxy translates the message received from the user through the messaging platform into the internal standard format of the architecture (e.g., a Signal format message translated into the internal standard modeled in the JSON Schema that contains user information, message body, attachment content, and timestamp). The API gateway is the smart gateway that decides, with the information obtained from the message body and the current state information of the system, which virtual assistant microservice is best suited to receive the user message. Apart from the core microservices, there is a virtual assistant microservices pool where each microservice has independent and specific tasks to perform. The virtual assistant microservices pool allows the virtual assistant to offer different functionalities to the user. Each microservice is completely independent from others, giving the opportunity to personalize the scenario by developing specific tasks for each group of patients' needs. Moreover, an advantage of using microservices is that new functionalities can be tested separately of the rest of the functionalities, added into the system without causing any disturbance to the user (the prototype does not need to be restarted in order to add new pieces of code into the microservice architecture, in contrast with monolithic architectures). One example of functionality is the microservice that allows users to check the opening hours and the location of their health clinic. Within the virtual assistant microservices, there are two different types: functions (microservices that have an interaction with the users) and data processing (microservices that perform more complex tasks, such as processing the relevant information of an image).

The microservices are based on an internal structure that consists of three main components designed for this architecture (that may appear in the microservice or not). The first component of the microservice is the part that communicates with other microservices. The communication uses JSON to exchange data between all microservices. Furthermore, the communication is established using the secure version of HTTP (Hypertext Transfer Protocol Secure, HTTPS). The second component is the interaction with the databases of the architecture. These interactions use HTTPS for the communication, in some cases using the standardization needed for some specific data types. More specifically, FHIR is used in the architecture for medical exchange data. The last component is the part of the microservice dedicated to the interaction with the user. Along this thesis, the user interaction is designed with workflows and modeled with two different types of interactions: a menu-based interaction using AIML and a human-like interaction using NLP.

3.1.1 Menu-based interaction

To represent the virtual assistant's brain, we use Artificial Intelligence Markup Language (AIML). AIML is a recursive language based on XML that defines the stimulus-response message of the interaction and is used to create artificial intelligent agents. AIML matches the user text input into the virtual assistant knowledge to obtain the response to send to the user. We decided to use AIML due to it is lightweight and easy to configure [97].

The AIML interaction is divided between the different microservices of the architecture. This separation of the conversation among different microservices generates the problem of knowing to which microservice any message is directed in order to obtain the user's response from his AIML file. To solve this problem, we propose an expansion of AIML with a new element called *Microservice*. This new element, *Microservice*, can be found within the AIML tag *Category*, providing the necessary information to know if this base unit of knowledge is in the middle of a conversation with the user or the conversation has ended to allow the user to initiate a new conversation with another microservice.

3.1.2 Human-like interaction

NLP is used in the integration of a human-like interaction into the conversational microservice. The structure of the proposed NLP-based conversational microservice can be seen in Figure 3.2. The proposed structure also uses the element *Microservice* to know when the user is in the middle of a conversation with the microservice. The elements inside the microservice are described as follows:

- **HTTPS Server:** The HTTPS server receives the petitions from the API gateway and checks the access permissions. Then, it sends the message received to the dialog manager and waits for the answer to send to the user.

- **Access Filter:** This element checks the authentication token received in the header of the HTTPS request.

- **Dialog Manager:** The dialog manager receives the user's sentence and obtains the intent and the slots (tags of the words in a sentence) from the Joint Natural Language Understanding (NLU) component. Based on the intent, it decides which tasks should be done and the rules to generate the response. It also decides if the user is in the middle of a conversation with the current microservice, and it provides the correct value to the *Microservice* element accordingly.

- **Joint Natural Language Understanding:** The Joint NLU component obtains the intent and the slots from the user's sentence. The Joint NLU component uses an NLU model to obtain the predictions. The choice of this model is not direct and a detailed analysis of the model and the used corpus needs to be performed. Due to the extension of this work, we decided to address these analyses in an independent chapter (Chapter 4).

- **Task Manager:** The task manager performs the database petitions and the tasks asked by the dialog manager.

- **Generation Manager:** This element generates the answer to the user with a rule-based system.

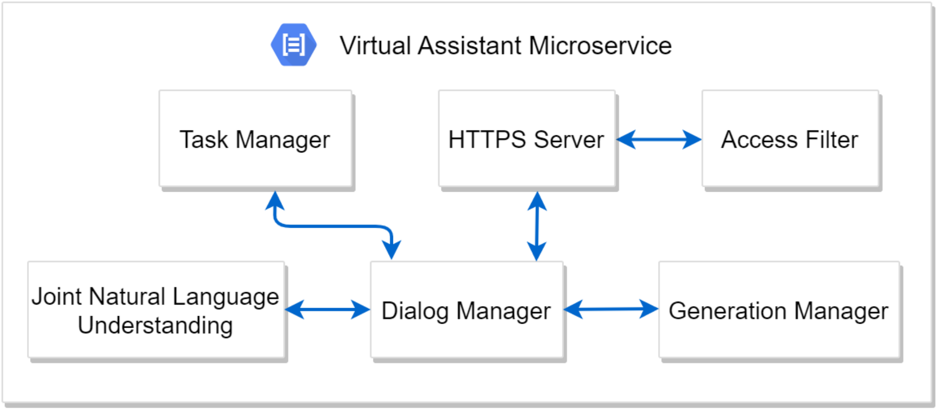


Figure 3.2: Internal NLP-based conversational microservice structure.

3.1.3 Analysis

Various aspects and challenges are addressed in the proposed architecture. We classify these in three categories: Modularity, Standardization and Security.

3.1.3.1 Modularity

The microservices architecture pattern accomplishes a level of modularity that in practice is extremely difficult to achieve with a monolithic architecture [98]. An important point in a microservice architecture is the service size. In the proposed architecture, the service size offers a virtual assistant split into tasks called functionalities. These functionalities are offered to users in the main menu with numeric options. More in detail, the main menu is created by the API gateway based on the available functionalities. The menu can adapt in real time to the functionalities running at that moment, to add modularity in the expansion of the microservices. Thanks to this feature, new microservices related to new functionalities can be added dynamically and independently in the architecture by informing the API gateway of the existence and nature of a newly added microservice in the system.

Furthermore, when a microservice with interaction is incorporated into the virtual assistant, the interaction is added inside that specific microservice. The new interaction can be added independently as a behavior on the virtual assistant even when the core system is running thanks to the new element, called *Microservice*. This provides the possibility of splitting the virtual assistant conversation into different components in the system architecture, knowing which microservice conversation the program should search to obtain the user response.

3.1.3.2 Standardization

Standardization in scenarios where data is stored and used is fundamental to guarantee the interoperability between different systems. Furthermore, standardization is fundamental to guarantee robustness against the difficulties faced during the deployment of the architecture. The main drawback is the initial effort needed from the developers to learn and adapt the system to the different standards used.

In a patient support scenario, medical data integration is required to guarantee interoperability. The proposed virtual assistant architecture supports the storage and sharing of medical information strictly aligning with HL7 FHIR standards. Furthermore, the standard AIML has been used as the menu-based conversation modeling to give homogeneity in the development of the interaction of the user with different microservices. In addition, NLP has been used as human-like conversational modeling to bring fluency in user conversations. Finally, a JSON Schema is proposed to standardize the communication between the microservices that require interaction with the user. The schema is shown in Code 3.1.

```
{
  "type": "object",
  "properties": {
    "user": { "type": "string",
      "description": "Unique user identifier." },
    "platform": { "type": "string",
      "description": "Indicates the platform through which the
    communication is being made." },
    "message": {
      "type": "object",
      "properties": {
        "timestamp": { "type": "integer",
          "description": "Includes all the relevant information of the
    message." },
        "attachments": {
          "type": "object",
          "properties": {
```

```
            "contentType": { "type": "string",
              "description": "Specifies the type of attachment." },
            "data": { "type": "string",
              "description": "Contains the attachment (base64 encoded)."
    },
            "size": { "type": "integer",
              "description": "Number of bytes of content." }
          }
        },
        "body": { "type": "string",
          "description": "Contains the text of the message." }
      },
      "required": [ "timestamp" ]
    }
  },
  "required": [ "user", "platform", "message" ]
}
```

Code 3.1: JSON Schema: message format.

3.1.3.3 Security

Besides all the things that have to be taken into account in any system, such as general security considerations (e.g. database configuration, known service vulnerabilities, etc.), some specific security issues have been considered in the proposed architecture. These issues are related with access control (authentication and authorization) for both users and microservices, so that the privacy of the user's data meets the expected requirements.

Inside the architecture, services authentication relies on Transport Layer Security (TLS) for both client and server. Since all communications are secured using HTTPS and all services in the architecture play at least the role of server (most of them are also clients), each service already have its own digital certificate. Relative to service authorization, each service in the architecture is provided with an Access Control List (ACL) at configuration time. This list includes the services that would need to perform calls to its API in a normal way of operation.

Furthermore, a user's identity is verified as soon as a new message reaches the system. User authentication relies on Signal's user management which is based on public key architecture where a user's identity key pair is generated at the time the application is installed and the public keys are exchanged the first time a communication is established between two users. To ensure that there are no man-in-the-middle attacks when the session is established, the users can check the virtual assistant's public key fingerprint which is shared with them beforehand. Once the system has verified the user identity, a JSON Web Token (JWT) is issued which is included in further requests between services. Moreover, authorization is performed in a finer-grained way using the XACML framework for policies definition and evaluation.

This security and privacy solution is based on standard security mechanisms that are typically used in microservices architectures, providing a complete vision of how existent technology could be well-suited to address security and privacy issues in the proposed architecture (critical when clinical data are involved). Notwithstanding, the proposed solution is not unique since other standard solutions already exist to that end (e.g. OAuth) and ad-hoc mechanism might be designed to be used in place.

3.1.4 Discussion

In order to use the virtual assistant, the user needs to be connected to the Internet and be familiar with the chatting interaction (not a difficult requirement taking into account the current penetration of messaging platforms). Obviously, being connected is also a necessary condition to receive reminders and notifications.

Privacy of the messaging platform is paramount. We have chosen Signal because of its end-to-end encryption capabilities and its open source code approach. If the communication is not end-to-end encrypted, medical information can be eavesdropped by the messaging platform servers. It is interesting to note that any messaging platform using end-to-end encryption for user communication can be used for the deployment. Nowadays, many platforms have developed specific APIs for an easy virtual assistant development (e.g. Telegram) but these APIs do not use end-to-end encryption at this moment (messages are decrypted at messaging platform servers and then encrypted again to be forwarded to the destination user). Hence, such APIs should not be used for secure virtual assistant development. This makes the choice of platforms for virtual assistant communication harder since not many platforms provide with an open way to use their end-to-end encryption mode. Signal is one of the few that provides end-to-end encryption.

3.2 Conclusions

The purpose of this chapter was to offer a solution based on microservices to provide personalized eHealth functionalities and data storage using virtual assistants in healthcare scenarios. The internal virtual assistant architecture allows the addition of new services and tools over time, splitting of the conversation and development of a wide range of healthcare functionalities as microservices. This architecture is designed to suit any chronic patient, overcoming the shortcomings of other mobile applications that are only intended for a specific type of disease [99, 100]. Standardization, security and fluent interaction are considered in the proposed architecture. The architecture provides the advantages of flexibility, modularity, and expansiveness in a virtual assistant scenario, enabling the virtual assistant to be modeled following the specifications of patients' illnesses. The virtual assistant architecture and the use of microservices provide a good flexible solution for personalized monitoring services. This architecture has been proposed to address the challenges of personalization and data storage in mHealth scenarios.

Chapter 4

Embeddings and design choices evaluation in Natural Language Understanding

NLU is a critical component in the fluent and successful conversations between virtual assistants and their users. The interest in NLU has recently increased, as proven by the increasing number of conversational systems in numerous fields that attempt to improve our quality of life [101, 102]. In Chapter 3, Section 3.1.2, we presented the human-like interaction for the virtual assistant architecture. We described the microservice elements, and we noted that the study of the Joint NLU component could fit better in a new chapter. Therefore, this chapter studies the Joint NLU component for medication management scenarios in a human-like interaction (extending the work presented in Section 3.1.2).

For this reason, this chapter focuses on what is best in a medication scenario using the Spanish language, the analysis being two-fold: (1), we evaluated different embeddings trained with three different datasets, to observe the impact of specific datasets in the model response; (2) the best combination of shorter concepts compared to having fewer but longer concepts and datasets with long and short sentences for joint NLU models.

4.1 Embeddings evaluation

Inside a specific domain, the NLU involves the tasks of slot filling and intent detection. Slot filling attempts to tag the words in a sentence with the slot types [15] and intent detection can be defined as the action of labeling the sentence based on predefined intents. An example of intent detection and slot filling in an NLU model that receives, as an input, the sentence "I would like to add paracetamol" is to obtain the intent ADD and extract the key-value slot {medication = paracetamol}.

4.1.1 Scenario overview

In existing NLU techniques, there are different embedding configurations, corpora schemes and learning models. To find out the best configuration of embeddings and corpora, we use different embeddings and datasets to obtain the best results for the medication domain. The different embeddings and datasets and the NLU system are shown in Figure 4.1. In a medication domain, due to privacy issues, there is no access to corpus obtained from patients. Therefore, we have generated a total of 30,185 sentences based on templates and slot filling for a medication scenario.

The proposed scenario uses the predictor multiLSTM [103]. This predictor uses bi-LSTM-CRF for joint NLU tasks (both slot filling and intent detection). multiLSTM uses a corpus in the Inside, Outside, Beginning (IOB) format to determine the intents and to detect the slots in a sentence. The original multiLSTM works with word2vec embeddings, training the word2vec with the generated data. We have adapted multiLSTM to be able to use fastText, ELMo and BERT as embeddings for the input of the neural network, using different embeddings trained with the proposed datasets for the comparison.

For our goal of comparing different embeddings and corpora to establish the best one, we have fixed the generated data and the model which predicts the intents and the slots. The study compares word2vec and fastText as word embeddings and ELMo and BERT as language models. The baseline is a word2vec implementation with continuous skip-gram model which has been trained with only the training data generated for the medication scenario.

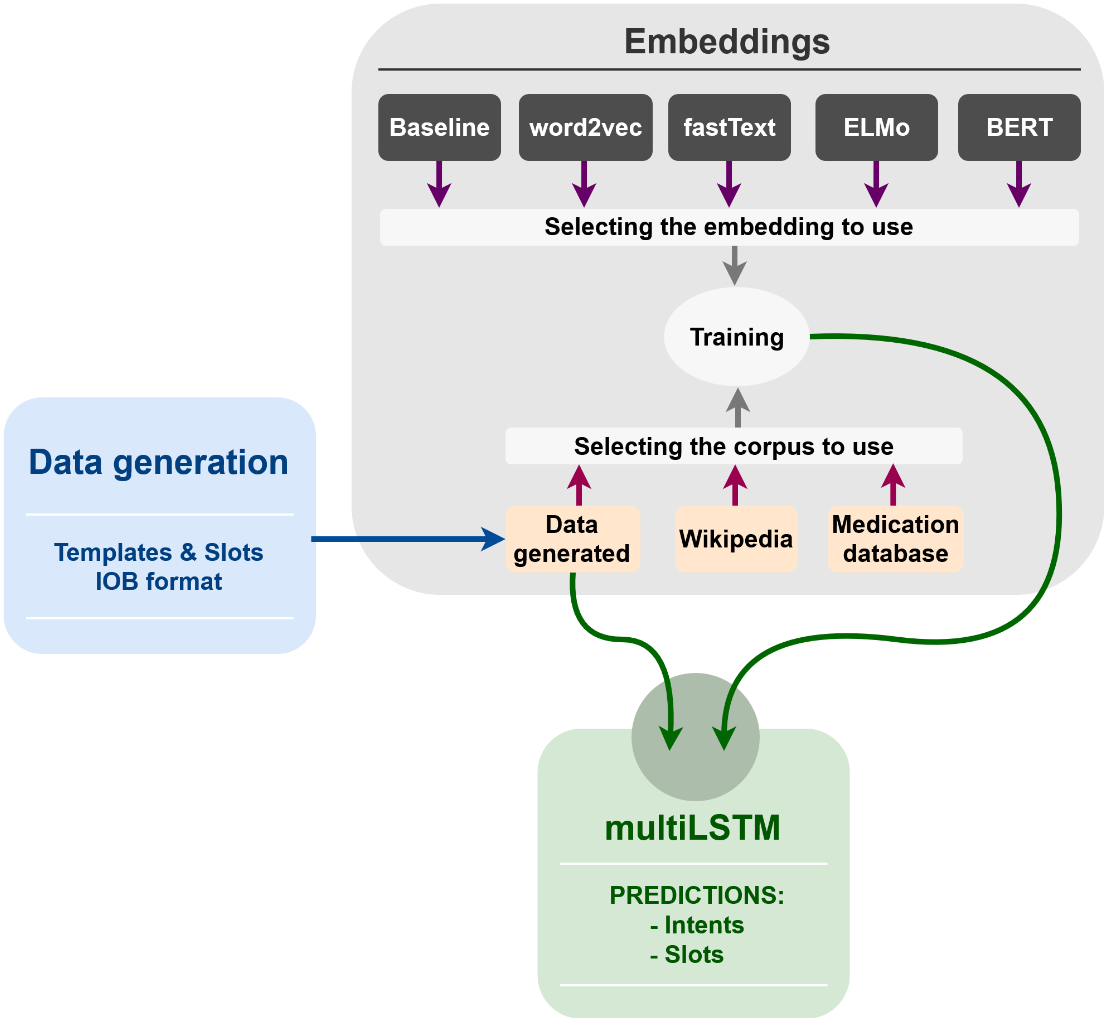


Figure 4.1: Training scenario overview.

4.1.1.1 Methods

In this section, we explain in detail the methods we use to generate data and different embeddings, learning models and datasets used in this work.

4.1.1.2 Data generation for intent detection and slot filling

The data was generated using a set of 13,881 templates in total, for five different predefined intents. Also, we have defined six files, one for each slot type defined in our scenario. We have generated 20,000 sentences from these slots and templates for the training set (used during the learning process to fit the parameters of the model). Additionally, we have generated 5,000 sentences for the development set (used to evaluate the model while tuning the model's hyperparameters). Both generations was obtained using the algorithm proposed by Boulanger, described in [104]. We have further added additional sentences for seven basic intents, creating a total of 24,270 sentences for the training set and 5,915 sentences for the development set. These generated data have been used as an input to the multiLSTM, as well as a corpus to train the embeddings.

4.1.1.3 Training dataset

In this work, we have utilized three different training datasets for comparison. The first dataset is the data generated for training described in Section 4.1.1.2. The training data has a vocabulary with size 4,291 and with 274,114 train tokens. The second dataset was obtained from the Spanish Wikipedia. This dataset has a vocabulary with size 2,968,376 and with 683,501,282 train tokens. To be able to observe if a specific corpus in the domain has a positive impact on the results, we use the third dataset obtained from the Spanish medication database [105]. This dataset has a vocabulary with size 152,393 and with 143,741,382 train tokens. We have obtained the information pamphlet from 12,736 drugs and the technical datasheet from 12,813 drugs.

4.1.1.4 Embeddings

This work analyzes 13 different configurations of embeddings and corpora. As a baseline, we have a word2vec implementation with continuous skip-gram model of 300 dimensions trained with the generated training set. Also, we have a word2vec implementation with continuous bag of words (CBOW) model of 300 dimensions trained with the generated training set, the Wikipedia corpus, and the Spanish medication database. fastText embedding (using 300 dimensions and CBOW model) was trained with the generated training set, the Wikipedia corpus, and the Spanish medication database. ELMo embedding of 512 dimensions was trained with Wikipedia corpus and ELMo embedding of 1,024 dimensions was trained with the generated training set and Spanish medication database. Finally, we have BERT embedding of 768 dimensions trained with the generated training set, the Wikipedia corpus, and the Spanish medication database.

4.1.1.5 Test data

We have developed a virtual assistant with a specific conversation about patient medication management to collect the test data. We have obtained a total of 456 sentences from 14 people between 22 and 66 years old. We have manually filtered the sentences to obtain a final number of 382 sentences as test data. This filtering was made as not all the sentences fit the proposed intents defined in our scenario of the medication domain. The test data has a vocabulary with size 291 and with 1,093 tokens. This vocabulary has 105 unknown words compared with the generated data for the training set. Comparing with Spanish Wikipedia, this vocabulary has 11 unknown words. Finally, this vocabulary has 34 unknown words compared with the Spanish medication database.

4.1.1.6 Evaluation

We have used a micro F_1-score with a chunk (a group of words grouped with the same tag that has discrete grammatical meanings)-level in the slot filling and an accuracy score for intent detection and sentence-level semantic frame to evaluate the results of the different configurations of embeddings and corpora. The sentence-level semantic frame indicates the general performance of the model, taking into consideration both tasks: intent detection and slot filling need to be correct in a sentence to consider the sentence correctly predicted. The F_1-score (defined by the Equation 4.1) considers the true positives (TP, the model predicts the chunk correctly), the false positives (FP, the model predicts a chunk where “there is not”), and the false negatives (FN, the model predicts there is not a chunk where there is one). Precision is the division between the number of correct positive results (TP) and the number of all positive results obtained by the model (TP and FP). Recall is the division between the number of correct positive results (TP) and the number of all samples that should have been identified as positive (TP and FN). The best value of the F_1-score is 1, and the worst is 0. Accuracy is the ratio of the correct predictions (TP and true negatives (TN, the model predicts where there is not a chunk correctly)) to the total predictions made (Equation 4.2). For accuracy, 1 is the best result, and 0 is the worst.

$$F_1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{4.1}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.2}$$

4.1.2 Results and discussion

The results obtained from the simulations performed with the different configurations are shown in Table 4.1. The best result was obtained with ELMo model trained with the Spanish Wikipedia, obtaining an F_1-score of 74.32% for slot filling, an accuracy of 71.99% for intent detection, and a sentence-level semantic frame accuracy of 62.30%. The results of ELMo configuration trained with our generated data and Spanish Wikipedia were obtained using the average of three layers and the result of ELMo configuration trained with the Spanish medication database was obtained using the first LSTM hidden layer, as we observed that they had the best results. The best results for BERT configuration trained with our generated data and the Spanish medication database were obtained by summing together the last four layers (each layer with 768 dimensions). The best result for BERT configuration trained with Spanish Wikipedia was obtained by concatenating the last four layers.

Comparing the results obtained with different training datasets, we observe the best results are obtained with the models that are trained with Spanish Wikipedia except for BERT. It seems that BERT has better performance when the training data is specific for the scenario. Nevertheless, BERT is usually the top model among the compared set in other languages and domains [106–108]. This idea agrees with the results obtained for BERT trained on generated data and the medication database. However, when trained on Wikipedia (the model was trained from scratch), it is almost ten points worse than the baseline. Therefore, those results deserve double-checking for a potential bug. The evaluation of BERT with Wikipedia requires further study and should be compared with a pre-trained Spanish BERT to obtain a conclusion. The worst results are obtained with the medication database using word embeddings. These results may be due to the fact that the medication database uses frequently a technical language, and in this context, the vocabulary used by the user with the virtual assistant is not very technical; usually the name of the medication is the only technical term in the conversation.

Method	**Slot filling $\mathbf{F_1}$-score**	**Intent detection accuracy**	**Sentence-level semantic frame accuracy**
Baseline on generated data	68.64	67.02	53.14
Word2vec CBOW[1] on generated data	62.50	57.33	47.12
Word2vec CBOW[1] on Wikipedia	72.04	69.37	58.38
Word2vec CBOW[1] on medication database	62.32	54.71	43.98
fastText on generated data	64.43	63.87	48.95
fastText on Wikipedia	69.59	68.85	55.50
fastText on medication database	59.96	58.12	43.19
ELMo[2] on generated data	64.70	64.66	52.09
ELMo[2] on Wikipedia	**74.32**	**71.99**	**62.30**
ELMo[2] on medication database	64.17	64.92	54.19
BERT[3] on generated data	71.35	64.92	49.74
BERT[3] on Wikipedia	58.94	55.76	44.76
BERT[3] on medication database	71.40	60.73	50.79

[1] Continuous Bag of Words; [2] Embeddings from Language Models; [3] Bidirectional Encoder Representations from Transformers.

Table 4.1: Results for Joint NLU (in %).

The main limitation of this model is the fact that our system predicts one of the defined intents and is not able to detect any other intent if the sentence is off-topic. Hence, as mentioned in Section 4.1.1.5, we need to filter the sentences obtained from the users to obtain only sentences that fit the intents of this scenario.

4.2 Slot types and training data evaluation

4.2.1 Materials and methods

4.2.1.1 Study overview

Given a sentence consisting of a collection of n words defined as $w = \{w_1, w_2, \ldots, w_n\}$, the main purpose of our work is to predict the correct intent i, where $i \in \{$ADD, SEE, ..., BYE$\}$ and the set of slots $s = \{s_1, s_2, \ldots, s_n\}$ associated with each word w_k, where $s_k \in \{$O, B-medication, I-medication, ..., B-posology, I-posology$\}$ if we are in the long-slot configuration or $s_k \in \{$O, B-medication, I-medication, ..., B-timeExpressions, I-timeExpressions$\}$ if we are in the short-slot configuration. Slots use the IOB format.

The system used in this work (shown in Figure 4.2) consists of two components: data generation and the joint NLU model. To determine the best configuration of data characteristics, we generated different datasets in the medication domain and tested them with different joint NLU models. Generally, it is quite challenging to gather a big corpus in the medical domain due to privacy issues (there is no access to corpus obtained from conversations with real patients in real scenarios). For this reason, we resorted to generating datasets that were close to the ones in real life.

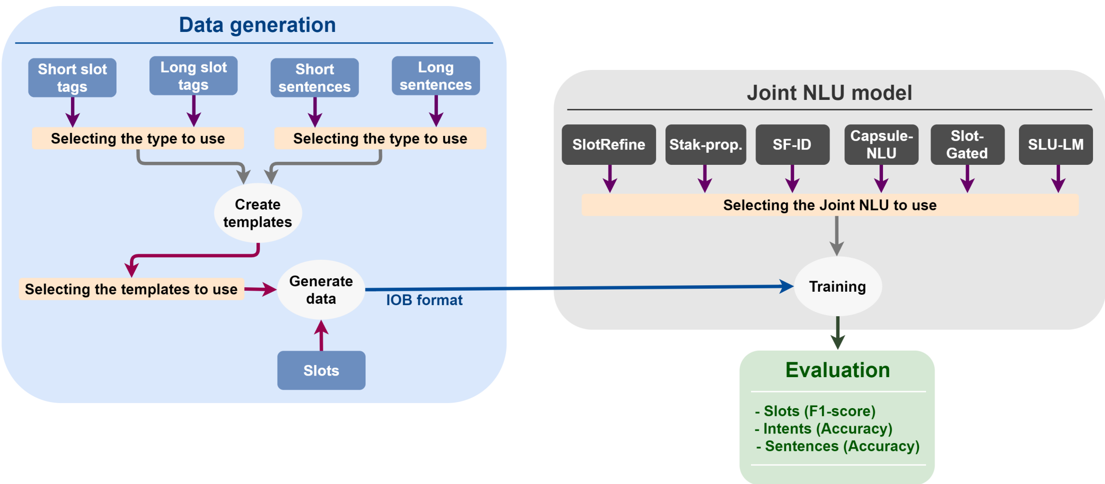


Figure 4.2: Overview of the system.

The details of the two components are described in the following subsections. Additionally, we present in detail the intents and slots proposed for this scenario and the real test data obtained from a group of real users.

4.2.1.2 Data generation

We generated a total of four datasets in Spanish to estimate the influence of slot tagging and sentence length. The four datasets are: two datasets with long sentences, one tagged with long slots and the other tagged with short slots, and two datasets with short sentences, one tagged with long slots and the other tagged with short slots. Table 4.2 summarizes the characteristics of the different datasets generated. These datasets were used as the input of the joint NLU models. All sentences are tagged using the IOB format.

We wanted to compare two different slot types: 1) short; 2) long but fewer slots. Therefore, we used templates created for each dataset with the corresponding slot tagged in each template. Then, we defined one file for each slot, and we used the method proposed by Boulanger [104] to generate the training and the development dataset for each combination. The method is based on a simple filling pattern for obtaining all of the possible sentence combinations and then sampling the generated sentences through a series of Euclidean divisions. The number of templates used to generate each dataset is shown in Table 4.2. An example of a long sentence is "I want to add $medication $duration", where a word after $ character means it is a slot, and an example of a short sentence is "Add $medication". After using the pattern filling, an example of a long sentence generated is "I want to add ibuprofen for 3 weeks" and an example of a short sentence is "Add ibuprofen".

We also wanted to compare two different sentence lengths. Hill et al. [109] compared the differences between human–human and human–virtual assistant interactions in the general domain, showing that a human–human conversation has a sentence length average of 7.95 words, whereas a human–virtual assistant conversation has a sentence length average of 4.29. We used these length averages as references, creating datasets with short sentences a bit shorter (4.1) and long sentences longer (between 9.1 and 10.9), with an average length difference of 5.9 between short and long sentences. Training datasets have, in total, 24,270 sentences, while development datasets have 5915 sentences.

Dataset	**Number of templates**	**Training sentence length (average)**	**Training total words**	**Development sentence length (average)**	**Development total words**
Long sentences and short slots	13,942	10.79	261,946	10.92	64,599
Long sentences and long slots	1756	9.06	219,949	9.23	54,590
Short sentences and short slots	367	4.06	98,574	4.11	24,337
Short sentences and long slots	367	4.10	99,428	4.13	24,403

Table 4.2: Datasets summary.

Examples of the long sentences include the following (examples are translated from Spanish for ease of reading):

- The doctor told me to take risedronato.
- Four puffs before going to bed for twenty-five days.
- The nurse told me to take rilast.
- Until 24 February 2023, two drops every two weeks.

Short sentence examples include the following:

- Delete medication.
- Summary of 8 days.
- Add doxidina.
- Yes, I took it.

4.2.1.3 Test data

For the test data, we used the real data described in Section 4.1.1.5. The real data was collected from a total of 14 users through the interactions with a virtual assistant. The virtual assistant had basic knowledge to ask questions related to medication management and continue the conversation with the user. The total number of sentences in the test data was 382, with a sentence length average of 2.70, and the total number of words was 1031. As for the vocabulary of the test data—there were 11 unknown words compared to the Spanish Wikipedia corpus. The unknown words were due to typographical errors, i.e., holaa (hi), ningunooo (none), pliss (please), and medication names (some with typographical errors): pharmagrip, eutirox, noctamid, norotil, traumel, urorec, uroret, espidifen.

Some examples from test data are the following (translated from Spanish):

- See my medications.
- Well I want to add paracetamol.
- Delete frenadol.
- Can you show me my medication?

4.2.1.4 Intent detection

We defined intent related to the medication scenario and additional intent related to the comprehension of the user's answers in a virtual assistant conversation. As our scenario focused on the task of medication management, we defined the intent as shown in Table 4.3.

Intent	**Category**	**Example**
ADD	Medication management	I want to add medication
SEE	Medication management	I want to see my medication
DEL	Medication management	I want to delete a medication
PRO	Medication management	I want to see my summary of medication in-takes
INT	Medication management	I have taken the medication of today
OPT	Help	What can I do?
YES	Answer	Yes
NO	Answer	No
NEUTRAL	Answer	Whatever you want
BCK	Answer	Cancel
HELLO	Greetings	Hi
BYE	Greetings	Have a nice day

Table 4.3: Intent defined for the medication scenario.

4.2.1.5 Slot filling

One of the purposes of this work was to determine the best slot tagging strategy, either having shorter concepts or having fewer but longer concepts. For our medication management scenario, we defined two slot configurations (shown in Table 4.4). Long slots were designed to cover posology (i.e., the branch of pharmacology that determines the appropriate doses of drugs and medicines) information in one slot, whereas short slots split the posology into different concepts, such as quantity, unit medication, unit frequency, and time expressions. The duration of the treatment was also defined as one slot in a long slot configuration, whereas a short slot configuration split the duration in quantity and unit frequency. The medication and date slots were the same for both configurations. The date slot refers to a specific day (for example, to indicate when the treatment starts). An example of these two slot configurations is shown in Figure 4.3, where posology ("20 drops every 8 h") can be expressed using quantity ("20" and "8"), unit medication ("drops"), and unit frequency ("hours"); duration ("for a week") can be expressed using quantity ("a") and unit frequency ("week"). As mentioned previously, the sentences are tagged using the IOB format.

Short slot tags	**Examples**	**Long slot tags**	**Examples**
medication	ibuprofen	medication	ibuprofen
date	today, last Monday	date	today, last Monday
quantity	1, 2, one	duration	for 3 weeks
unitMedication	ml, cl, mg	posology	1 mg every 8 hours
unitFrequency	days, hours, weeks		
timeExpressions	at breakfast		

Table 4.4: Short and long slots proposed for the medication management scenario.

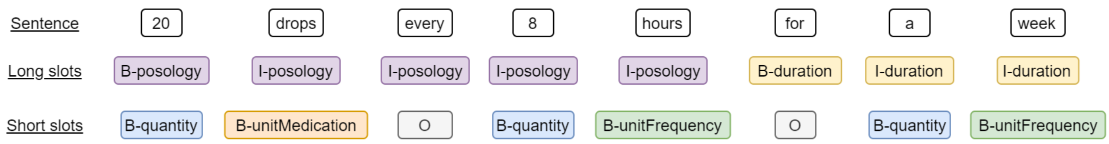


Figure 4.3: Example of slot filling with long and short slot configurations (using the IOB format).

4.2.1.6 Joint NLU model

In this work, we used joint NLU models proposed in the literature to evaluate the effects of the generated datasets in intent detection and slot filling. The maximum number of epochs (one pass of the entire training dataset through the algorithm) was set to 100 for all simulations in this work. The rest of the hyperparameters not specified are with the default value. We used the following joint NLU models with their respective configurations.

- SlotRefine [110]: is a two-pass refine mechanism that uses the B-tags of the first pass as input for the second pass. The results were obtained with a batch size of 32 and 64, a learning rate of 0.0025 and 0.001, a hidden size of 80 and 96, a filter size of 80 and 96, and a number of heads of 8 and 16. Furthermore, the attention dropout was set to 0.05, patience set to 0, remove numbers was set to true, and the multiply embedding mode was set to "none".

- Stack-propagation framework [111]: is a joint learning scheme, which first uses a self-attentive encoder, performs a token-level intent detection, and finally leverages the token-level intent information to improve the slot filling decoder. The results were obtained with a word embedding dimension of 32 and 256.

- SF-ID network [112]: has a model architecture that is based on bi-directional LSTM (BLSTM) and uses two subnets: SF subnet and ID subnet. SF-ID network also includes a CRF layer to jointly decode the labels of the sentence. We obtained the results with four configurations: SF-first with and without CRF, and ID-first with and without CRF. The early stop was disabled.

- Capsule-NLU [113]: combines dynamic routing and re-routing processes with three different types of capsules (WordCaps, SlotCaps, and IntentCaps) to accomplish intent detection and slot filling. We obtained the results with two configurations: using dynamic routing and re-routing processes and using only dynamic routing processes. The early stop was disabled.

- Slot-gated modeling [51]: is an attention-based recurrent neural network (RNN) model. We obtained the results with the two different models proposed in the slot gate mechanism: slot attention, and intent attention approach, and only intent attention approach. The early stop was disabled.

- Joint SLU-LM [114]: is a joint online spoken language understanding (SLU) and language modeling (LM) model with RNN. We obtained the results with two different configurations: recurrent intent context and both local and recurrent intent context. The maximum sequence length was set to 50, DNN at the output layer was set to true, steps per checkpoint were set to 1017, and the maximum training steps were set to 101,700.

4.2.1.7 Evaluation

The results were evaluated using the micro F_1-score with a chunk-level in the slot filling and the accuracy score for intent detection and sentence-level semantic frame. Furthermore, the Wilcoxon signed-rank test (a non-parametric statistical hypothesis test that compares two paired groups to assess whether the sets of pairs are significantly different from each other) was used to evaluate the statistical significance of the outcomes for continuous variables. Null hypothesis: there was no significant difference in F_1-score or accuracy results between short and long slots or sentences. If the p-value is lower than

0.05, the null hypothesis is rejected, and the results are considered significantly different. All statistical analyses were conducted using R software, version 4.0.3 [115].

The error of each slot was calculated using Equation 4.3. Error-values of 0 or close to 0 mean that the model was able to predict the presence and the absence of the slot in the sentences, and values of 1 or close to 1 show a model that is not able to predict the slot correctly.

$$Error = \frac{FP + FN}{TP + TN + FP + FN} \tag{4.3}$$

4.2.2 Results

The results for slot filling, intent detection, and sentence-level semantic frame are summarized in Tables 4.5–4.7, where the best results for each model are highlighted in bold. The SlotRefine results shown in the tables were the best results obtained for each dataset from all possible configurations. Stack-propagation results used a word embedding dimension of 32. The best results for SF-ID were obtained with ID-first and without the CRF layer. Capsule-NLU results used dynamic routing and re-routing processes. The best results for the slot-gated modeling were obtained with the intent attention approach. Joint SLU-LM results used both local and recurrent intent contexts. Overall, the best results were obtained using short sentences and short slots for most joint NLU models.

	Short sentences		**Long sentences**	
Method	**Short slots**	**Long slots**	**Short slots**	**Long slots**
SlotRefine	**76.4**	59.8	75.2	56.0
Stack-Prop.	**51.1**	33.1	49.5	31.6
SF-ID	**53.7**	29.3	49.9	31.0
Capsule-NLU	47.4	27.4	**53.1**	27.7
Slot-Gated	**57.8**	36.1	52.5	33.0
SLU-LM	38.7	19.5	**50.9**	20.2

Table 4.5: Results for slot filling with joint NLU (in %).

	Short sentences		**Long sentences**	
Method	**Short slots**	**Long slots**	**Short slots**	**Long slots**
SlotRefine	79.3	79.1	76.7	**81.4**
Stack-Prop.	59.2	**59.9**	**59.9**	59.7
SF-ID	**61.0**	60.5	56.8	57.6
Capsule-NLU	**45.0**	44.5	32.2	44.2
Slot-Gated	60.7	**61.0**	56.3	58.9
SLU-LM	46.9	45.6	**47.9**	46.6

Table 4.6: Results for intent detection with joint NLU (in %).

	Short sentences		**Long sentences**	
Method	**Short slots**	**Long slots**	**Short slots**	**Long slots**
SlotRefine	**68.6**	68.3	65.7	66.5
Stack-Prop.	36.9	**37.7**	35.9	36.9
SF-ID	**39.3**	35.9	32.2	34.3
Capsule-NLU	19.9	19.6	20.4	**21.5**
Slot-Gated	**40.6**	38.7	34.3	35.3
SLU-LM	**18.1**	15.7	17.3	14.1

Table 4.7: Results for sentence-level semantic frame with joint NLU (in %).

We performed statistical testing for slot tags and sentence lengths. If the p-value is lower than 0.05, we can conclude that the results are significantly different.

Slot tags:

- Results for long sentences and long slots (fifth column of the three tables) versus long sentences and short slots (fourth column of the three tables) ($p = 0.28$) are not significantly different.
- Results for short sentences and long slots (third column of the three tables) versus short sentences and short slots (second column of the three tables) ($p = 0.004$) are significantly different.

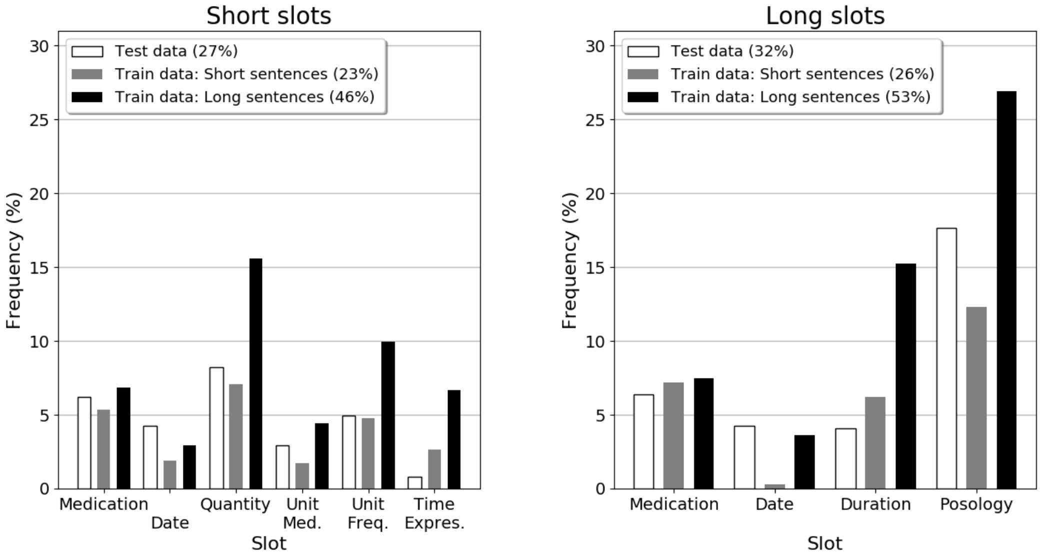
Sentence length:

- Results for long sentences and long slots (fifth column of the three tables) versus short sentences and long slots (third column of the three tables) ($p = 0.12$) are not significantly different.
- Results for long sentences and short slots (fourth column of the three tables) versus short sentences and short slots (second column of the three tables) ($p = 0.04$) are significantly different.

When we compare long to short sentences, and short to long slots, the best results for slot filling and sentence accuracy are achieved with short sentences and short slots when SlotRefine is used (76.4 and 68.6%, respectively), and the best result for intent detection are achieved with long sentences and long slots when SlotRefine is also used (with an accuracy of 81.4%).

4.2.2.1 Datasets

In order to observe the differences and similarities in the datasets, Figure 4.4 shows the frequency of the slots in the test and generated datasets. The percentage of the number of slots in the datasets is also included. We observed that short sentences have a percentage of the number of slots in the dataset similar to the test data, compared to long sentences.



(a) Short slots datasets.

(b) Long slots datasets.

Figure 4.4: Composition of the datasets in terms of slot tags and tagged slot proportion (the legend percentage).

4.2.2.2 Errors

The percentage of errors for slot filling can be observed in Tables 4.8 and 4.9. Observing the slots that are in both configurations, the medication slot shows the highest value of errors; short sentences with short slots had the best configuration for the prediction of this slot for 4 out of 6 models. The date slot had the highest errors when using short slots and short sentences.

We can observe that, comparing duration and posology with quantity, unit medication, unit frequency, and time expressions, long slots have an average error of 9.96% compared to 6.18% for short slots. We observe that, for long slots, short sentences have an average error of 40.60%, and long sentences have 39.85%. Furthermore, for short slots, short sentences have an average error of 32.60%, and long sentences have 27.13%. Based on these results, we can conclude that joint NLU models trained with long slots yield worse results than the models trained with short slots.

Slots	Sentences	SlotRefine	Stack-Prop.	SF-ID	Capsule-NLU	Slot-Gated	SLU-LM
Medication	Short	8.5	7.3	6.7	23.1	16.6	42.0
	Long	11.0	21.7	24.1	10.7	20.9	14.1
Date	Short	2.0	17.3	16.9	6.6	5.1	5.4
	Long	1.2	4.1	5.4	4.9	4.2	4.4
Quantity	Short	1.0	3.2	2.9	3.9	1.9	2.6
	Long	1.1	3.0	3.1	2.8	2.5	3.5
Unit Medication	Short	0.8	1.9	1.5	1.9	1.0	1.8
	Long	0.9	1.6	1.0	1.8	1.0	1.0
Unit Frequency	Short	1.0	2.4	1.6	0.9	1.1	3.0
	Long	0.6	1.6	1.7	2.1	1.2	2.8
Time Expressions	Short	0.4	1.1	0.4	0.3	0.5	1.0
	Long	0.1	0.5	0.4	0.5	0.5	1.0

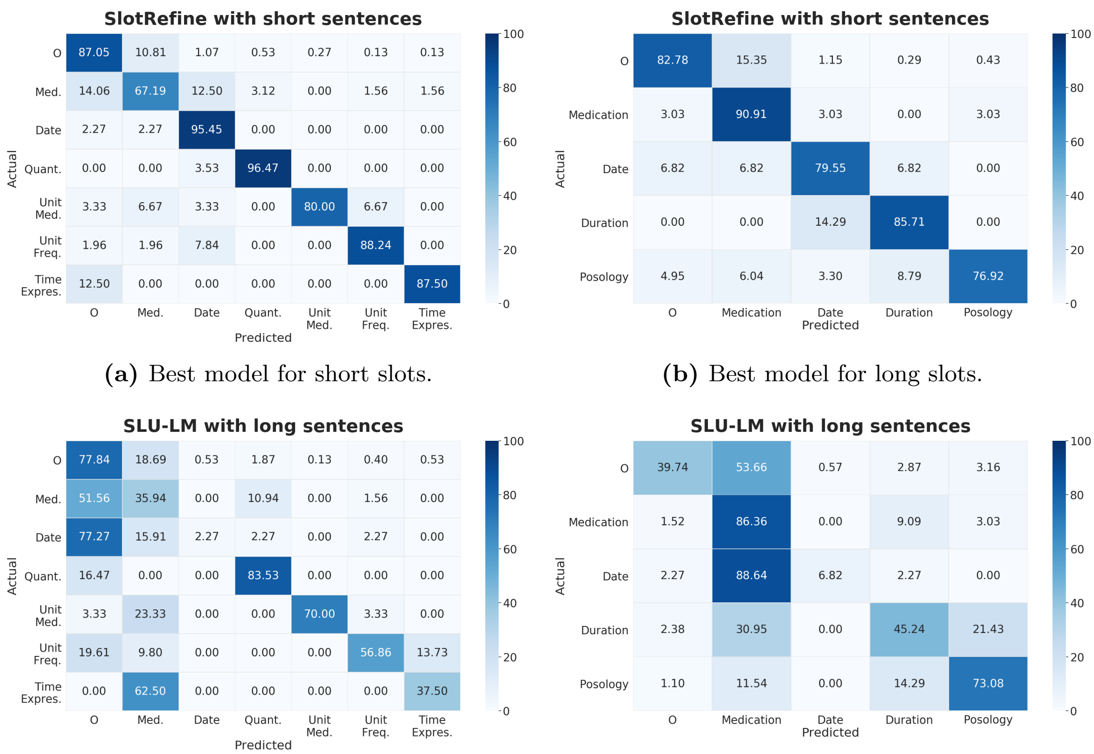
Table 4.8: Errors in slot tagging for short slots (in %).

Slots	Sentences	SlotRefine	Stack-Prop.	SF-ID	Capsule-NLU	Slot-Gated	SLU-LM
Medication	Short	11.3	19.7	22.9	26.6	21.3	46.3
	Long	15.9	24.0	25.2	28.1	21.8	41.5
Date	Short	2.6	10.9	6.1	4.8	3.8	5.5
	Long	1.1	4.5	5.1	4.9	4.3	5.0
Duration	Short	1.6	2.7	4.6	4.3	4.5	10.0
	Long	2.2	3.4	3.0	4.3	4.4	6.4
Posology	Short	4.4	4.1	4.9	5.7	4.7	10.2
	Long	4.8	4.6	5.0	6.0	6.0	7.7

Table 4.9: Errors in slot tagging for long slots (in %).

4.2.2.3 Model performance

The performances of the models were analyzed using the confusion matrix and the t-SNE representation. The confusion matrices for the best and the worst models are shown in Figure 4.5. We observe that, for the best model trained with short slots, medication and unit medication were the slots that had more errors, whereas date and medication were the slots that had more errors for the worst model. In addition, the best model trained with long slots showed more errors in posology slots, and the worst model trained with long slots had more errors in date slots. Observing the best model predictions for short slots, nada (nothing) was wrongly predicted ten times, ninguno (none) six times, medicamentos (medications) five times, and Hola (Hi) five times. All of them were tagged as medication and should have been tagged as “O”. It seems that the model interpreted frequent words, such as “nothing” or “Hi”, as medication. This fact may increase the number of mistakes for this slot type.



(a) Best model for short slots.

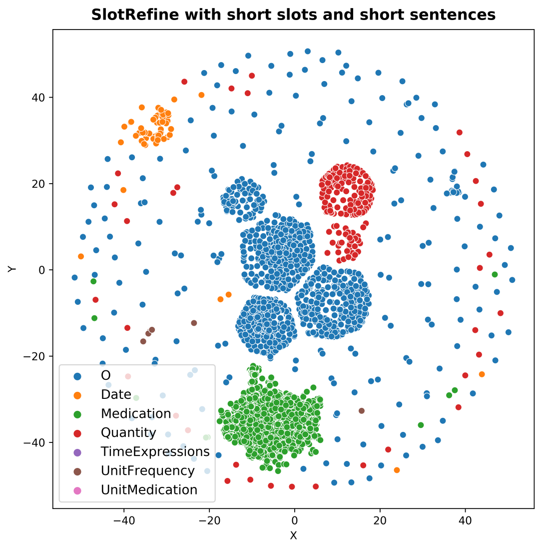
(b) Best model for long slots.

(c) Worst model for short slots.

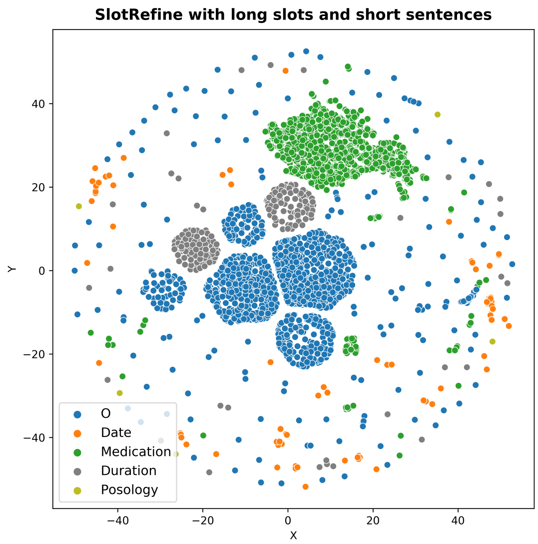
(d) Worst model for long slots.

Figure 4.5: Confusion matrices for the best and the worst models for short and long slots (in %).

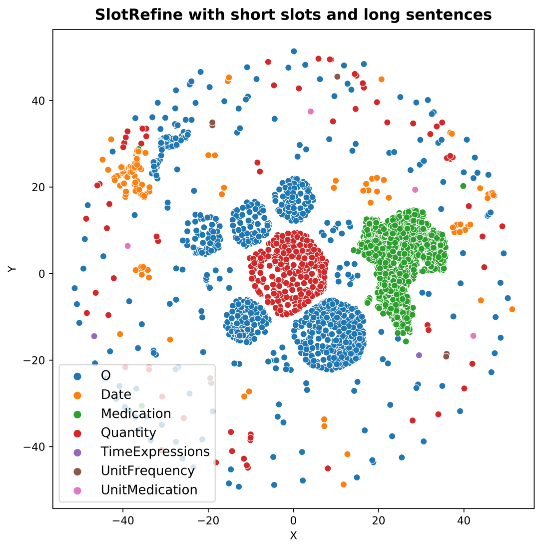
t-SNE representations are shown in Figure 4.6. We used the vocabulary from training, development, and test data to represent their transforms with the different models. We compared the best model for each combination of slots and sentences. With this visual evaluation, we can observe how scattered the vocabulary is in each model. The models trained with short sentences show the tags as more compact and clustered, which may be the reason why these models obtained the best results.



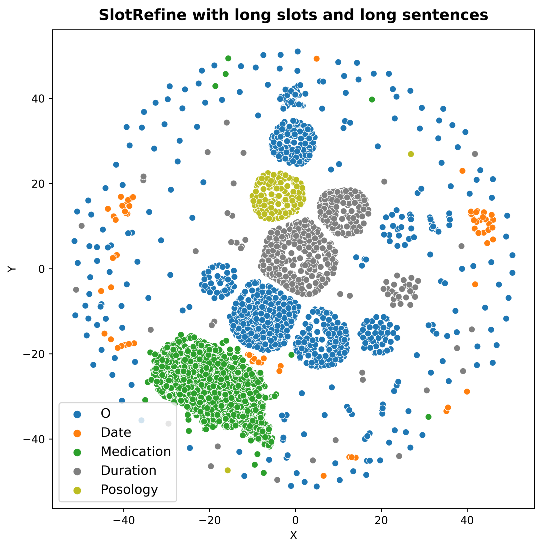
(a) Best model for short slots and short sentences.



(b) Best model for long slots and short sentences.



(c) Best model for short slots and long sentences.



(d) Best model for long slots and long sentences.

Figure 4.6: t-SNE representations for the best models for each combination of slots and sentences.

4.2.3 Discussion

In this work, we evaluated six joint NLU models trained with four different datasets. In all of the models, short slots (with an average of 54.7) showed better results in the case of slot filling compared to long slots (average of 33.7). This result may be due to the fact that short slots are more specific and avoid ambiguities that may appear in long slots. Sentence length did not seem to affect slot filling results, with an average of 44.2 for both short and long sentences. It seems that slot tags and sentence lengths were slightly significant variables for intent detection since short slots have an average of 56.8, long slots 58.3, short sentences 58.6, and long sentences 56.5. In the case of sentence-level, the outcomes showed that short sentences (with an average of 36.6) had slightly better performance than long sentences (average of 34.5), and 5 out of 6 models yielded better results with short sentences. The slot tags did not affect the sentence level, with an average of 35.8 for short slots and 35.4 for long slots. To summarize, the slot type seems to be a variable that affects the slot filling outcome but not intent detection or sentence-level outcome. Sentence length seems to be a variable that slightly affects sentence-level and intent detection outcomes, but not the slot filling outcome.

As observed in Figure 4.4, short sentences and test data have a similar frequency of slots in the sentences, whereas long sentences have a significantly high frequency. Long sentences have more combinations of slots in the same sentence, which may increase the complexity and show a slightly negative impact on the performance of the models. We noted that, for the slots, such as quantity, unit medication, unit frequency, and time expressions, the percentage of error was more or less the same while using different joint NLU models (as observed in Table 4.8). This can be explained by the fact that short slots may decrease the complexity of the task, observing a better performance in the joint NLU models, whereas long slots, such as duration or posology, increase the difficulty, with different error percentages between models.

Impact on dialogue system

Another fact to consider is that the NLU model is part of dialog management. After obtaining the relevant information in the sentence by the NLU model, the slots tagged need to be processed to save the information accordingly. In this case, if, for example, the posology is tagged with long slots, it needs complex post-processing with regular expressions where the different values of the posology need to be obtained. Nevertheless, short slots can simplify the post-processing task because they are already divided into small pieces of information, being a better choice for the complete task, the dialog system.

4.3 Conclusions

The purpose of this chapter was to obtain the best result and discuss different configurations of embeddings with different corpora for a medication management scenario. Such comparative study can help in understanding the effectiveness of existing NLU techniques in medical domain and facilitate more advanced and intelligent features in virtual assistants. In our analysis, the NLU models in the medication domain show better results with the ELMo model trained with the Spanish Wikipedia. Moreover, most of the models trained with the Spanish Wikipedia yield better results compared with those trained with our generated data and the Spanish medication database.

Furthermore, this chapter analyzed the impact of slot tagging and training sentence length using four different datasets generated specifically for this work. The findings contributed toward understanding the effects of slot tagging and the characteristics of the generated data in joint NLU models designed to be used in virtual assistants, specifically for medication management scenarios in Spanish. A total of six joint NLU models were used to analyze the influence of the training data. In our analysis, short slots significantly yielded better results for slot filling outcomes. It seems that training sentence length slightly affects the performance of the models at intent detection and the sentence level. Our analysis obtained the best outcomes with short slots and short sentences using SlotRefine as the NLU model.

Chapter 5

Prototype Development

This chapter describes the generic prototype developed and its adaptation into two specific profiles: type 2 diabetes mellitus and depressive disorder, and psoriasis. Type 2 diabetes mellitus and depressive disorder profile provides medication management tools (reminders, notifications, and summaries). The functionalities included in this prototype are implemented using the menu-based interaction. Additionally, we complemented this prototype with an NLP-based interaction that can intelligently process and understand complex sentences in order to facilitate medication management with the aim to explore and offer two types of interactions to healthcare professionals. Psoriasis profile provides teleconsultation and image storage. Based on the dermatologists' opinion and their interaction preferences, this prototype is fully developed with menu-based interactions. These two prototypes are evaluated in Chapter 6. Appendix A includes two more prototypes as examples of the flexibility of the proposed virtual assistant architecture, which have not been evaluated.

5.1 Generic prototype

This prototype development aims to create a generic virtual assistant including the most common functionalities for chronic disease management. In Lasierra et al. [116], eleven different chronic diseases were studied in order to obtain the data gathering requirements so as to implement a complete telemonitoring scenario based on ontologies. This work has provided us with a highly valuable study in order to know the data gathering requirements as well as patient interactions for a very complete set of chronic conditions.

These requirements were obtained working with primary care physicians, creating a questionnaire that contains all the needs for the supervision of each chronic disease. Moreover, we have added additional requirements for psoriasis with the collaboration of dermatologists and their patients. A summary of these needs is shown in Table 5.1.

Studying the needs of different chronic diseases, some common functionalities can be obtained, as well as specific functions for concrete illnesses. The need of monitoring some vital constants such as weight or height are integrated into the same microservice, called *Questionnaires*, which can ask periodically the users' vital constants, giving the possibility to monitor their data. We have modeled this microservice to allow physicians to create customized questionnaires, generating automatically the alerts and the interaction with the patients. Other more specific microservices, such as one that offers deeper tools in monitoring the quantity and the type of food eaten, are not included in the generic prototype because not all diseases need it.

In order to validate the viability of the architecture, we have developed a generic virtual assistant with the most common functionalities. The rest of specific functionalities can be independently and easily added to the generic virtual assistant developed. To validate the flexibility of the scenario, after building a generic virtual assistant, we have modeled and developed new functionalities that are specific for different scenarios, described in the following sections.

The overview of the scenario implemented is shown in Figure 5.1. The Proxy microservice is the only microservice that interacts directly with the messaging platform server, as shown in Figure 3.1. In this prototype, Signal is the messaging platform used. The microservice Proxy sends and receives the client messages through the Signal server, using the Java libraries signal4j [117] version 1.0.4 and signal-service-java [118] version 2.3.1_unofficial_1 (both libraries have been modified because they are not updated with the latest changes of the Signal official repository [119]). The Proxy microservice has been developed in Android, based on a sample bot [120]. Despite having developed the prototype with the Signal messaging platform, the Proxy is designed so that new messaging platforms can easily be added to the architecture.

Profile	Needs to ask
Chronic Obstructive Pulmonary Disease (COPD)	Height, weight, pulse rate, blood pressure, SpO2, FEV1, glucose, medication, number of cigarettes, patient location, room temperature, room humidity, health test questionnaire (COPD specific).
Obesity	Height, weight, pulse rate, blood pressure, SpO2, glucose, medication, liquids quantity, food quantity, body fat, body water, health test questionnaire (Obesity specific).
Thyroid Disorders	Height, weight, pulse rate, blood pressure, SpO2, temperature, glucose, medication, liquids quantity, food quantity, health test questionnaire (Thyroid specific).
Ischemic heart disease (IHD)	Height, weight, pulse rate, blood pressure, SpO2, Glucose, medication, number of cigarettes, health test questionnaire (Ischemic specific).
Asthma	Height, weight, pulse rate, blood pressure, SpO2, FEV1, Temperature, Glucose, patient location, room temperature, medication, number of cigarettes, room humidity, health test questionnaire (asthma specific).
Hypertension (HTA) or high blood pressure	Height, weight, pulse rate, blood pressure, SpO2, Glucose, medication, number of cigarettes, health test questionnaire (HTA specific).
Osteoporosis	Height, weight, medication, health test questionnaire (Calcium), health test questionnaire (Falls).
Heart failure (HF)	Height, weight, pulse rate, blood pressure, SpO2, medication, liquids quantity, health test questionnaire (Liquids), health test questionnaire (HF).
Diabetes mellitus	Height, weight, pulse rate, blood pressure, Glucose, medication, liquids quantity, food quantity, number of cigarettes, HbA1c, health test questionnaire (Diabetes).
Dyslipidemia	Height, weight, pulse rate, blood pressure, medication, number of cigarettes, health test questionnaire (Dyslipidemia).
Osteoarthritis	Height, weight, blood pressure, temperature, medication, number of cigarettes, health test questionnaire (Osteoarthritis).
Psoriasis	Height, weight, pulse rate, medication, patient location, sleep hours, stress level, health test questionnaire (Psoriasis).

Table 5.1: Chronic patient profiles with their needs.

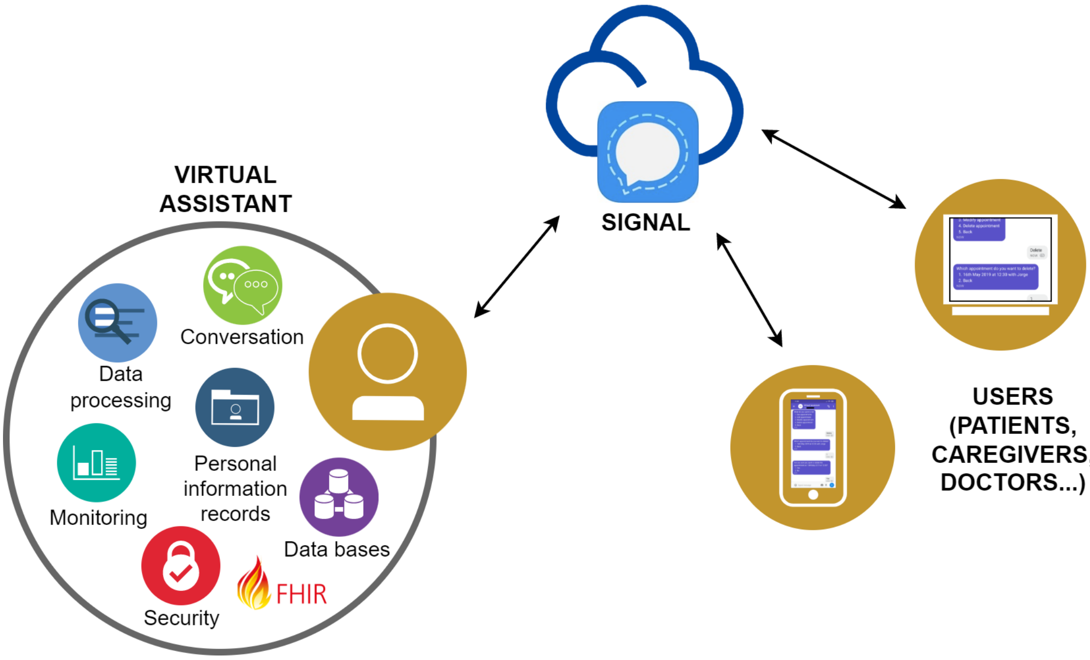


Figure 5.1: Overview of the scenario.

The pool of microservices for the virtual assistant architecture developed for this prototype is described in Table 5.2. As mentioned before, we have the *Questionnaires* microservice to ask users about their healthcare data. Also, generic microservices such as *Register*, *Unsubscribe* or *Verification* are included due to the necessities associated to the use of the virtual assistant. *Specialist* and *Patients* help to see the different actors that use the platform. Another tool that is shared between all chronic diseases is *Medical appointments* to manage the appointments with the healthcare professionals.

Furthermore, a role is assigned to each user. There are three different roles: patients, healthcare professionals, and informal caregivers (e.g., patient's family). Moreover, patients can create care groups to provide access to their data to other users, such as healthcare professionals or informal caregivers. Thereby, patients are the center of a care group.

Microservice	**Tasks**
Proxy	1. Pull messages from Signal server 2. Translate the messages into a formatted JSON 3. Send the messages to the API gateway
API gateway	1. Choose the best microservice to send the message to 2. Forward the message to the chosen microservice
Verification	Register the user in the virtual assistant
Personalize	Enable or disable the functionalities in the menu
Modify profile	Modify personal data
Reminder	1. Check the reminder queue 2. Send the reminders to the users
Unsubscribe	Delete all the data related to the user
Specialist	1. Show the specialists list 2. Change the data permissions for the specialists
Patients	1. Show the permissions granted to a specialist 2. Show a summary of the patient's activity
Register	Register, by another user, a new user in the virtual assistant
Medical appointments	1. Show the medical appointments 2. Add new medical appointments 3. Modify medical appointments 4. Delete medical appointments 5. Set reminders related to the appointments
Questionnaires	1. Create questionnaires 2. Modify questionnaires 3. Show questionnaires 4. Delete questionnaires 5. Fill in questionnaires 6. Generate AIML files based on questionnaires 7. Set reminders related to fill in the questionnaire 8. Export results
Suggestions	Receive the feedback from users

Table 5.2: Prototype: generic virtual assistant.

The microservices in this project have been developed in two principal programming languages: Java (using Java 8 Update 121) and Python (using Python 3.7). The microservices developed in Java use the open source Jersey framework supporting JAX-RS APIs to implement the RESTful service and client development, and the Grizzly framework to implement the HTTPS server. The microservices developed in Python use the Flask framework for developing small server applications and Gunicorn is used to serve the Flask application at a production level. All the microservices developed in this work have an asynchronous HTTPS server that relies on TLS for clients authentication. In order to provide a way to add new microservices easily into the generic virtual assistant, we have created a template in Java and another in Python (the templates in both programming languages are available in `https://github.com/ehealthz-lab`). Thanks to these templates that contain the basic code to develop a microservice which works in the proposed architecture with menu-based interaction, new functionalities related to specific illnesses could be added into the generic virtual assistant. The developer needs to model the conversation flow and create the AIML file with the conversation of this new functionality of the generic virtual assistant. After, the AIML file and the programmed tasks of this specific microservice are added in the template, creating the new microservice. We have used these templates with three students, that have developed new functionalities inside the virtual assistant without a deep knowledge of the architecture itself, obtaining good results about the easiness of the usage of the templates.

We use the Docker platform to build lightweight containers, each one with a different microservice of the architecture, allowing dynamic deployment. This allows us to have a modular, independent and flexible microservice architecture with the possibility of adding new microservices without compatibility problems with other programming languages or versions.

In order to provide interoperability with other eHealth systems, the user healthcare information is stored in FHIR resources, that contain all the relevant data for each case. The FHIR resources used to cover the information storage needs in this scenario are shown in Figure 5.2. The user information is stored in *Patient*, *Practitioner* and *RelatedPerson* resources, depending on the type of the user. The other resources shown in Figure 5.2 are for storing the relevant information that each user provides during the conversation with the virtual assistant.

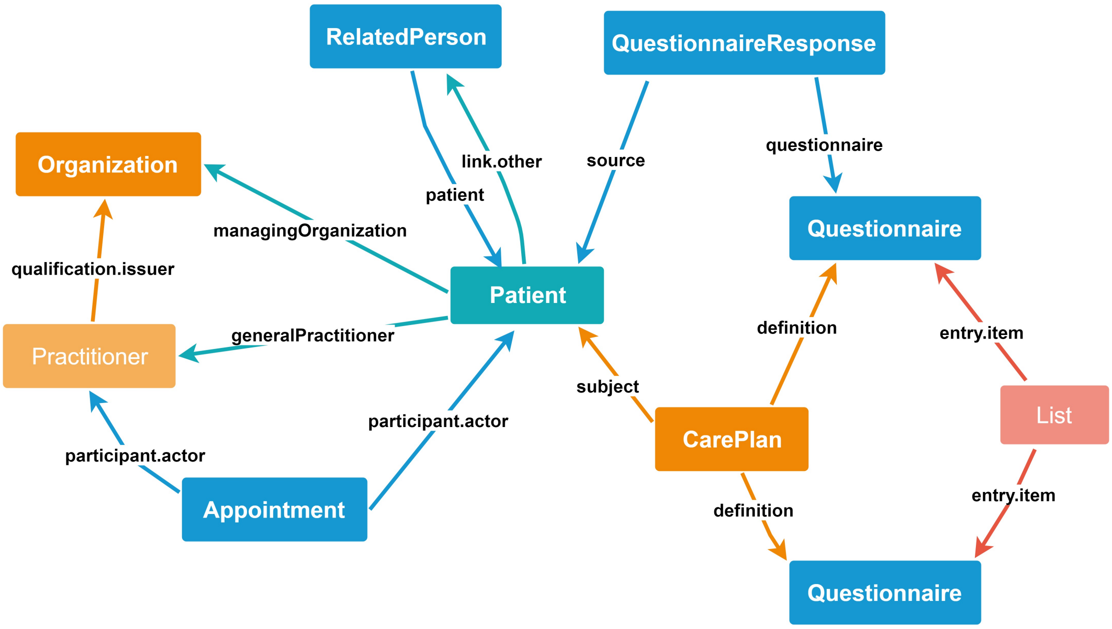


Figure 5.2: Relations of the FHIR resources.

AIML version 2.0 is used in this prototype as a language to define the virtual assistant behavior. Thus, the library of the Program AB [121] has been used in this prototype as our AIML 2.0 interpreter that obtains the best match between all AIML categories stored in the AIML files of the project. The conversation patterns in the generic virtual assistant have been developed in Spanish, but examples are in English for the sake of understandability. In order to provide personalization in the conversation with the user based on the information stored in the FHIR resources, the creation of AIML files must be dynamic and flexible to adapt to each specific case. The AIML files used generically in the virtual assistants are created in a static manner, not offering this degree of personalization based on the information stored in databases. This work tries to solve this problem by offering an AIML file generator that, based on the FHIR resources available in the database, generates a new file with the interaction to be produced with the user, resulting in a more personalized interaction.

The use of this AIML file generator can be useful in generating a user interaction for a questionnaire. This questionnaire is stored in the database with the statement of the questions and the type of data stored for each question as a *Questionnaire* resource (an example is shown in Code 5.1). This *Questionnaire* resource can be created by the doctor or by another user. The AIML file generator is able to read this resource and create a new file with the user interaction where the user has to fill out the complete questionnaire. An example of an AIML file created by the AIML file generator is shown in Code 5.2. This file is generated from the resource *Questionnaire* shown in Code 5.1. The file contains the user-virtual assistant interaction generated from the questionnaire. First, when the user asks for the questionnaire, the virtual assistant asks the first question to the user (in this case, "How many hours did you sleep last night?"). Then, with the response of the user, the virtual assistant saves the response thanks to the oob tag, knowing that it should be a decimal answer ("<oob>SAVE_DECIMAL</oob>"). Then, the virtual assistant finishes the questionnaire because there are no more questions in the FHIR resource. All this conversation is automatically created thanks to the AIML file generator.

```json
{
  "resourceType": "Questionnaire",
  "id": "45467",
  "meta": {
    "versionId": "1",
    "lastUpdated": "2019-01-14T10:14:25.000+00:00"
  },
  "title": "Sleep quality",
  "subjectType": [
    "Patient"
  ],
  "item": [
    {
      "linkId": "1",
      "text": "How many hours did you sleep last night?",
      "type": "decimal",
      "required": true
    }
  ]
}
```

Code 5.1: FHIR resource example: Questionnaire.

```xml
<?xml version="1.0" encoding="UTF-8"?>
<aiml version="2.0">

<category><pattern>QUESTIONNAIRE45467</pattern>
<template>How many hours did you sleep last night?</template>
<microservice>questionnaire</microservice>
</category>

```

```
<category><pattern>*</pattern>
<that>How many hours did you sleep last night?</that>
<template>Thanks for answering the questions<oob>SAVE_DECIMAL</oob></template>
</category>

<category><pattern>*</pattern>
<template><random>
<li>Sorry, I have not understood you</li>
<li>Can you repeat it to me in another way?</li>
</random></template>
</category>
</aiml>
```

Code 5.2: AIML example: Questionnaire auto-generated.

This AIML generator has been built based on the patterns used to build AIML files. These patterns have been programmed in Java (using Java 8 Update 121) to be able to create the conversation flow in the case of surveys. Each *Questionnaire* resource is linked to its AIML file by the name of the file and also with the first pattern of the first category of the AIML file, where the *id* of the resource in FHIR is set to know which AIML file should use the virtual assistant (as shown in Code 5.1, line 3, with the FHIR id 45467). Also, the *Microservice* AIML tag extension is used to know when the user is in the middle of a conversation and when the questionnaire is finished.

In this example, we have shown how the translation from an FHIR resource to an AIML file is performed, showing one of the many cases in which this generator can be used.

All the user interaction and the usage and the state of the microservices can be shown in the monitoring microservice thanks to the logging facility. In this prototype, Kibana [122] has been used as a tool to display the relevant information such as, for example, the messages sent over time or the average usability values of microservices (shown in Figure 5.3).

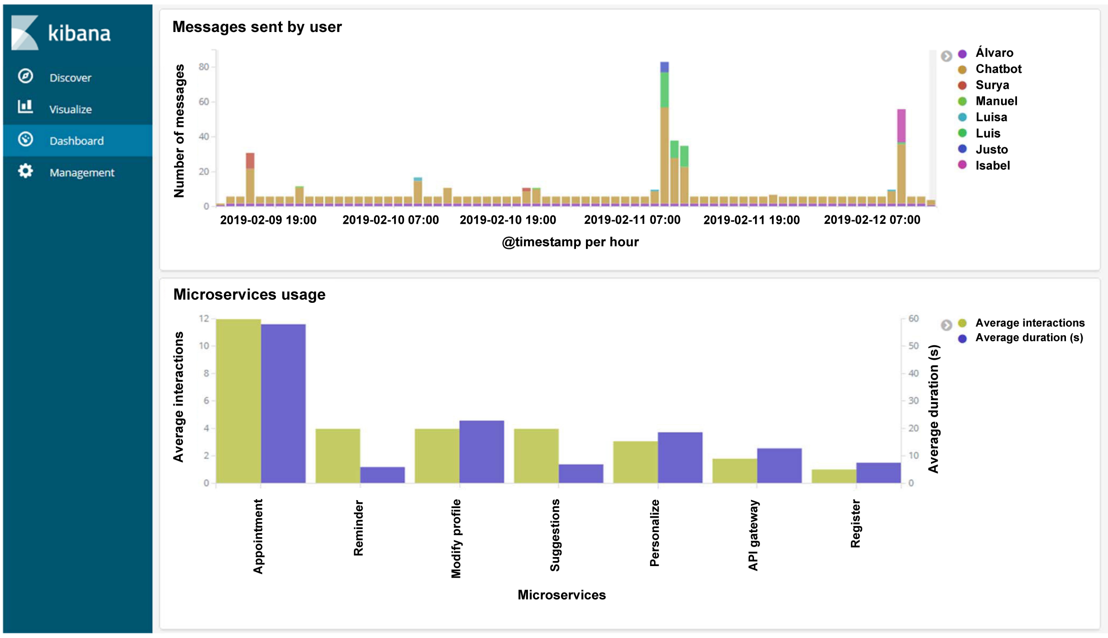


Figure 5.3: Logging snapshot: users' messages and microservices usage.

The privacy of the data is preserved due to all the data exchanged in the scenario are encrypted. Because of Signal is used as a messaging platform in the scenario, the conversation between the end-user and the virtual assistant has end-to-end encryption. Inside the architecture proposed, all the data exchange between microservices use the HTTPS protocol, which encrypts the data, providing privacy. All the sensible data related to personal data and illnesses are stored in FHIR, which uses the HTTPS protocol to exchange the information and provide privacy. The FHIR data are stored in an encrypted hard disk. The periodic backups are configured to cipher the data and save them in an external encrypted hard disk. Applying all these measures of privacy and user's rights, we follow both national data protection law LO 03/2018 [123] and European General Data Protection Regulation (GDPR) [124] due to the sensitivity of the data stored in the system.

5.2 Type 2 diabetes mellitus and depressive disorder profile

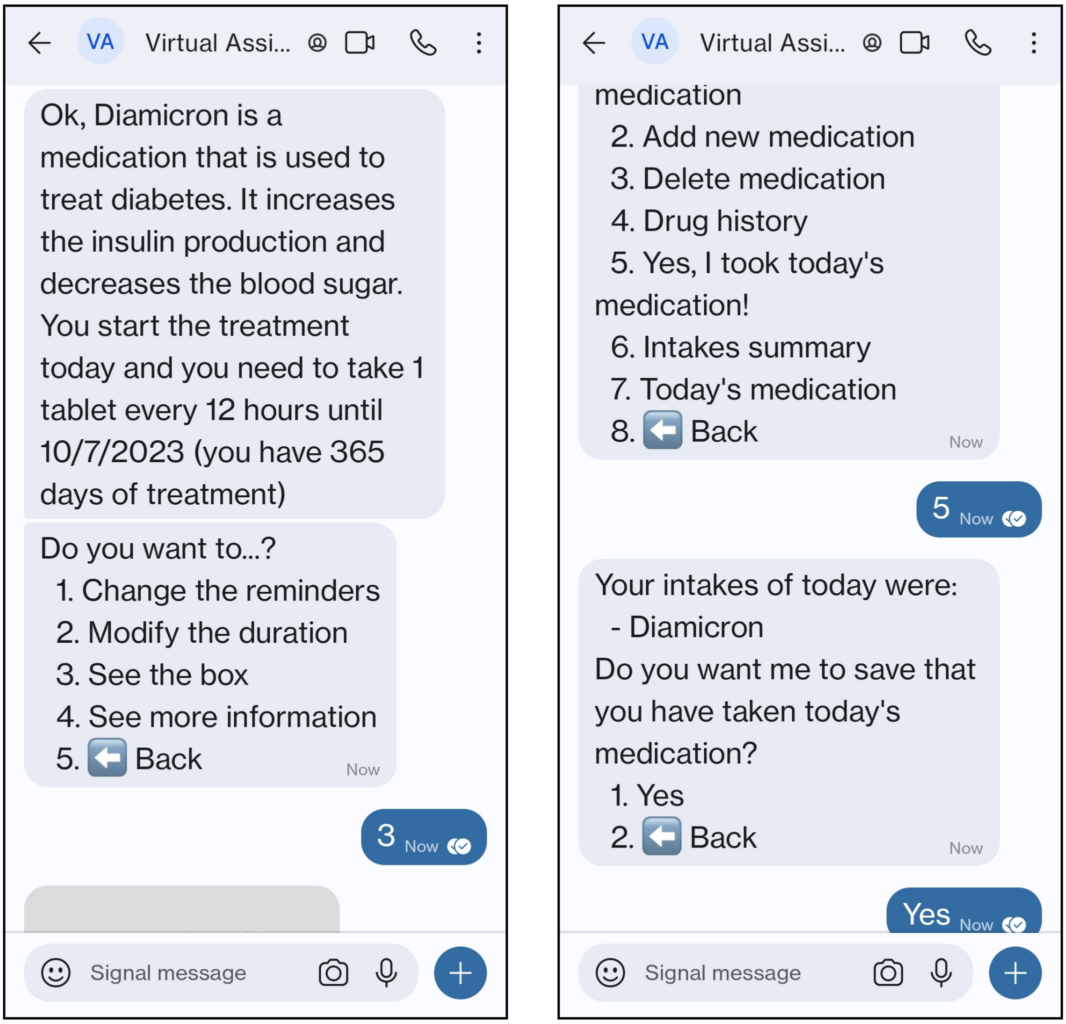
The prototype developed in this section for type 2 diabetes mellitus and depressive disorder patients is one example of the possible use-case scenarios where the proposed microservice-based virtual assistant architecture can be used. The virtual assistant aims to support and provide different tools such as managing medication, setting reminders, and data visualization to its users.

The functionalities for this medication adherence study were discussed and designed in collaboration with the nurses from the primary healthcare center Las Fuentes Norte in Zaragoza, Spain. Comorbid type 2 diabetes mellitus and depressive disorder diseases are addressed through the *Medication* and *Tutorials* microservices (shown in Table 5.3) and a *Reminders* microservice modified from the generic prototype. In order to adapt and integrate the text messaging to different scenarios where the virtual assistant interacts with the users, the text messages have been modeled using sequence diagrams and discussed with the nurses.

Microservice	**Tasks**
Medication	1. Show the current medication 2. Add new medication 3. Delete medication 4. Show the drug history 5. Inform the virtual assistant that the patient has taken the day's medication 6. Show the medication intake summary 7. Show the remaining medication to be taken during the day 8. Set the reminders for taking the medication 9. Show the medication information (box, information pamphlet, etc.)
Reminders	1. Send a message at the reminder time 2. Send three reminders, one every 10 minutes 3. Provide the medication box image if selected
Tutorials	List the tutorials available on YouTube

Table 5.3: Prototype: Type 2 diabetes mellitus and depressive disorder virtual assistant.

Medication management involves tasks such as add new medication, show the drug history and show the remaining medication to be taken that day. Medication management also performs tasks like show the medication information. The user can ask, as shown in Figure 5.4a, the image of the medication box, to help her/him to take the correct medication. An example of how a patient informs that the medication was taken can be seen in Figure 5.4b.



(a) Medication information. **(b)** Medication intake.

Figure 5.4: Real interaction with the virtual assistant developed.

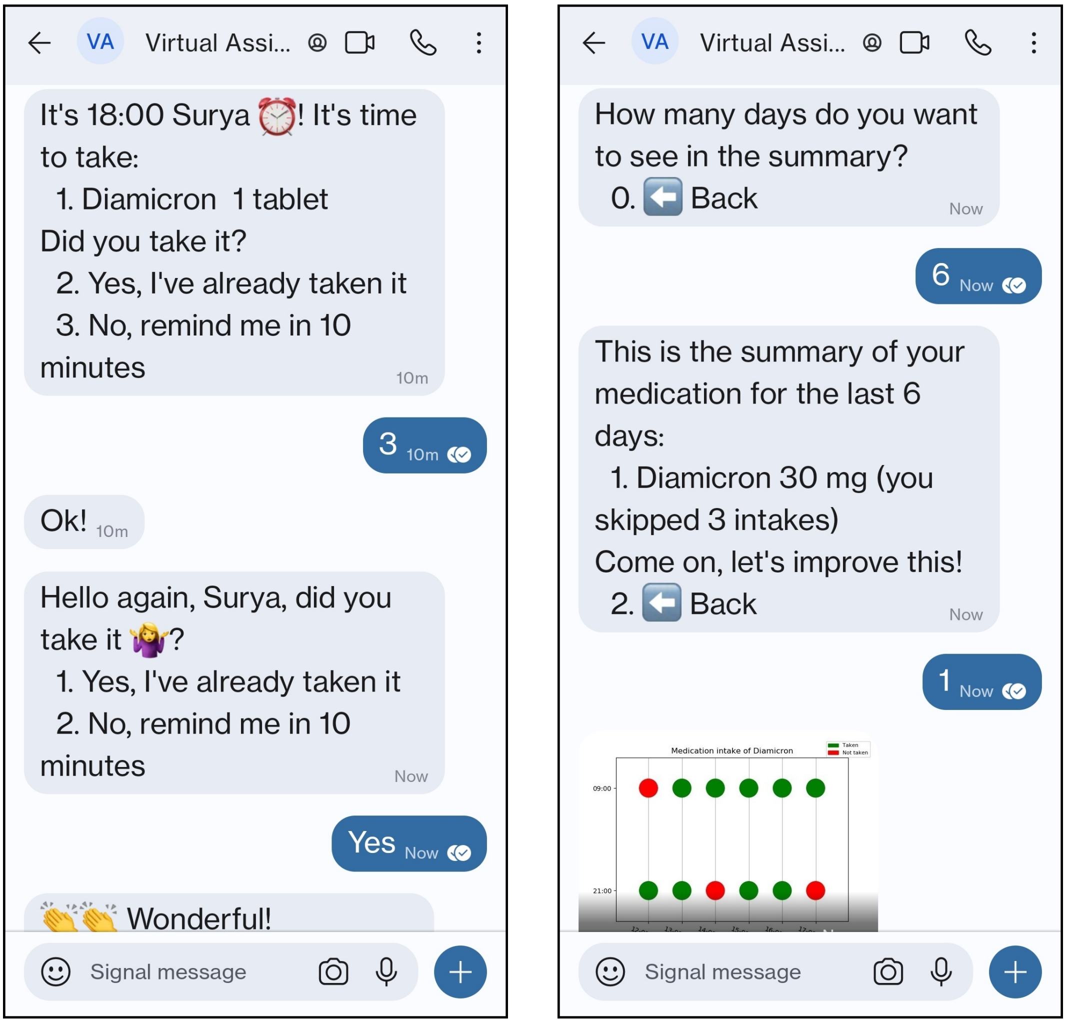
Reminders are messages sent to the users at the scheduled time when they need to take a medication. The usage of reminders can help patients improve their adherence to medication. Reminders can be set by patients, caregivers or healthcare professionals (patients' care group) when a new medication is added. These reminders are specific for reminding the drug dosage and the quantity. In response to these reminders, the user can specify if the medication has been taken or not. Medication reminders are sent a maximum of three times per programmed intake. If the patient does not respond to the reminder, another reminder is scheduled with a delay of 10 minutes. If the patient has not answered after three reminders, the virtual assistant stops reminding the patient and marks the intake as unanswered. Nevertheless, the virtual assistant includes a way for the patient to indicate that despite not answering the reminders, the medication was taken. This is a very useful safeguard because it could be a common situation for the patient not to hear the reminders (thus not answering on time), but taking the medication nevertheless. The patient can inform the virtual assistant that the medication was taken by two options: either by answering the reminders or by telling the virtual assistant using the medication options menu. An example of the conversation that takes place when a user receives a new reminder of a drug administration is shown in Figure 5.5a.

The response (or absence of) to the reminders is used to estimate the adherence to medication and to create a summary of drugs intake. Furthermore, the proposed system allows the flexibility to the patients' care group (patients, healthcare professionals, and informal caregivers) to receive the summaries of medication periodically (e.g. once every three days) or to receive a summary of the last days whenever the user asks for it (see Figure 5.5b for an example).

In addition, the virtual assistant shows the weather forecast every morning and sends links to tutorials on YouTube that explain how to use the virtual assistant. It was included an initial tutorial where the basic usage of the virtual assistant was explained.

The interaction developed in this prototype is a menu-based interaction with AIML. Nevertheless, the medication management functionality sometimes may need a more complex implementation to understand the dosage and the different configurations of the medication. Therefore, in this prototype, we integrated an AIML medication microservice with a menu-based interaction and an additional NLP-based medication microservice with a human-like interaction, both providing the same tools. That way, the nurses can decide which interaction fits better for them.

The structure of the proposed NLP-based system was described in Section 3.1.2, with the internal structure shown in Figure 3.2. The joint NLU component uses the SlotRefine model trained with short slots and short sentences, chosen based on the results from Section 4.2.2 for the medication domain. The model is trained to obtain the medication, the dosage, and the frequency of the intake. NLU, dialog manager and generation manager are modified for medication management domain from the demo described in [125] for a task-oriented dialogue system working on the cooking domain.



(a) Reminder example. **(b)** Summary example.

Figure 5.5: Real interaction with the virtual assistant developed.

The interaction with the joint NLU component through Signal is shown in Figure 5.6. The example shows a user introducing the medication and the virtual assistant asking for the posology (using the NLP medication interaction).

The input of the medication data can be done in different ways. One possible option is to do it manually by entering the medication name or national code. With this code, it is possible to search for detailed information about this drug in the Medicine Online Information Center website from the Spanish Medication Agency. After that, the virtual assistant asks further information related to that drug, such as dosage or intake duration. Another interesting possible option is to use a photo of the medical prescription. We have developed an Optical Character Recognition (OCR) module trained with the Aragón health department prescription template, able to extract the medication details (including dosage, timespan, etc.) from the photo. Although the procedure is very accurate, possible errors may occur due to the luminosity or the angle of the page in the photograph. To guarantee that only correct data is introduced, the users are requested to check the accuracy of the extracted data in a step by step process, verify the data and change the data if it is not accurate. Nevertheless, there is an inherent limitation with OCR recognition. The OCR that we have developed only works with the prescription from our Spanish region (Aragón) due to the customization needed in the OCR module. If the system is used with other prescriptions models, the OCR recognition needs to be adapted accordingly.

Figure 5.6: Example of medication interaction with the virtual assistant using a smartphone.

Once the medication data is introduced by the user, the data is stored in the system following the medical standard HL7 FHIR. The FHIR resources added in this scenario are "Person" to store the actors of the system and "Medication", "Medication Request", and "Medication Administration" to store the medical data.

5.3 Psoriasis profile

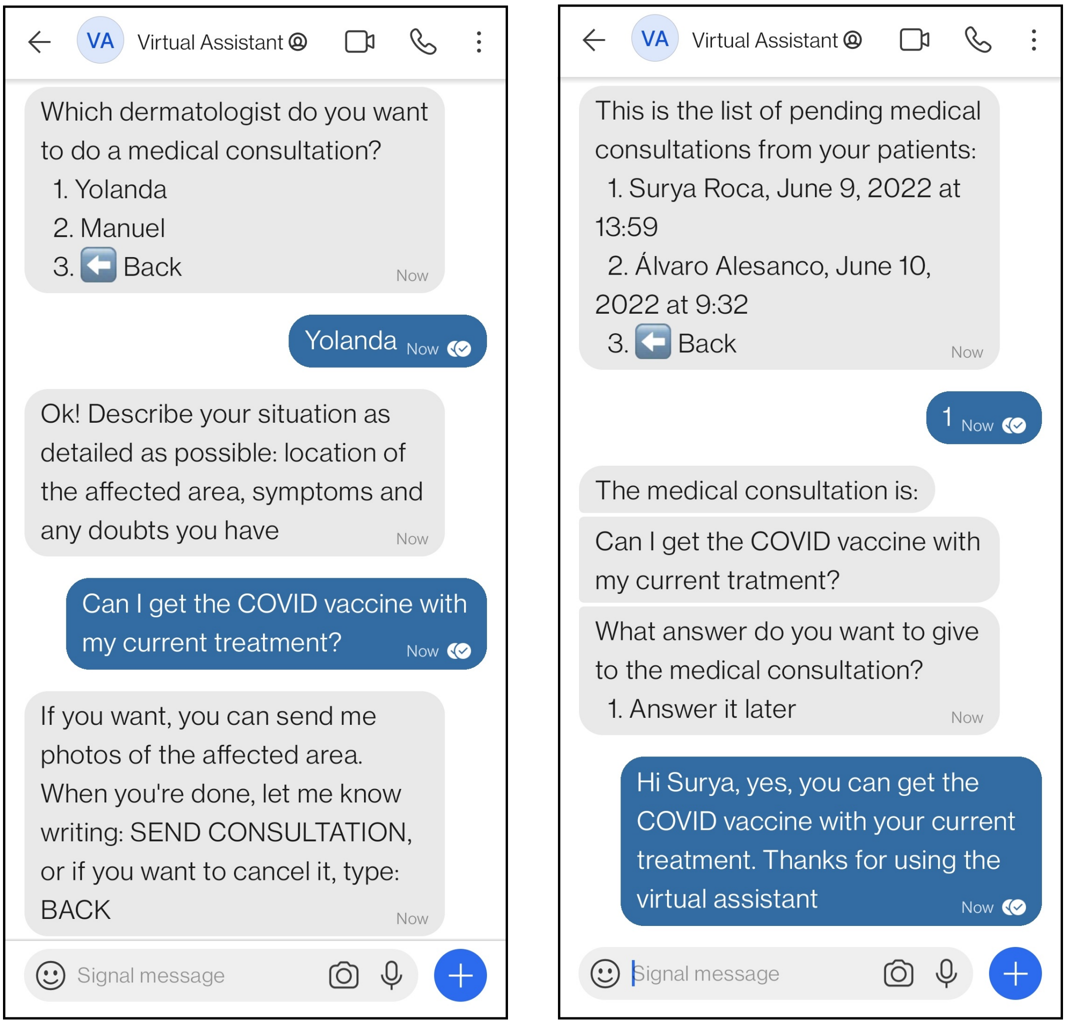
The psoriasis prototype implemented in this section aims to demonstrate the feasibility of another possible use-case scenario for the proposed microservice-based virtual assistant architecture. The new virtual assistant seeks to support and help chronic psoriatic patients by providing specific tools for psoriasis monitoring. Patients have the possibility of saving images of the body surface area through the virtual assistant which is then able to show the saved images over time. Thanks to these functionalities, patients, caregivers and healthcare professionals are able to see the evolution of the affected skin area and how it reacts to the treatment.

The functionalities are designed with the collaboration of a group of dermatologists, which are given their needs and opinion about what they want to have in the virtual assistant to give support to psoriatic patients. In order to obtain the specific functionalities needed in a psoriasis scenario, we have followed two steps. First, we have asked what are the parameters that the dermatologists usually monitor in their medical consultations, and we have obtained the questionnaires that they use to ask periodically to their patients. Then, with this information, we have designed a group of functionalities and discussed them with the dermatologists, obtaining the final functionalities specially designed for the psoriasis scenario. The microservices that we have added to the generic virtual assistant are shown in Table 5.4. We have added new monitoring tools that helps with tasks which were very difficult to perform before, the image storage of the affected areas. The virtual assistant provides an improvement in the facility to store images and watch them in a timeline, in order to see the evolution of the affected area. Questions about the patient illness can be sent to the dermatologists through the virtual assistant and the dermatologist can answer them. Furthermore, dermatologists can save clinical values made in the medical appointment with the patient and send informative messages through the virtual assistant. Essentially, the psoriasis disease is addressed through the *Broadcast*, *Medical consultation*, *Scales*, *Send photos*, and *Record* microservices. A usage example of this microservices is when a patient saves photographs every day during a flare-up in the *Send photos* functionality and later on sees the progress of the symptoms over time thanks to the *Record* functionality. The FHIR resource "Media" has been added into the scenario, to store the images generated from the users. In addition, the resource "Observation" has been added to store the queries with their responses and the patient scales.

Microservice	**Tasks**
Broadcast	Send informative messages to the patients
Medical consultation	1. Redirect the questions that patients have about their illnesses to the dermatologists 2. Save the answers of the dermatologists to the queries 3. Show the answers to the patients 4. Modify the answers to medical consultations 5. Delete medical consultations
Scales	Store clinical values measured in the medical appointment by the dermatologist
Send photos	1. Save patient's photographs 2. Explain how to take a good photograph
Record	Display the images that are saved in the virtual assistant

Table 5.4: Prototype: psoriasis virtual assistant.

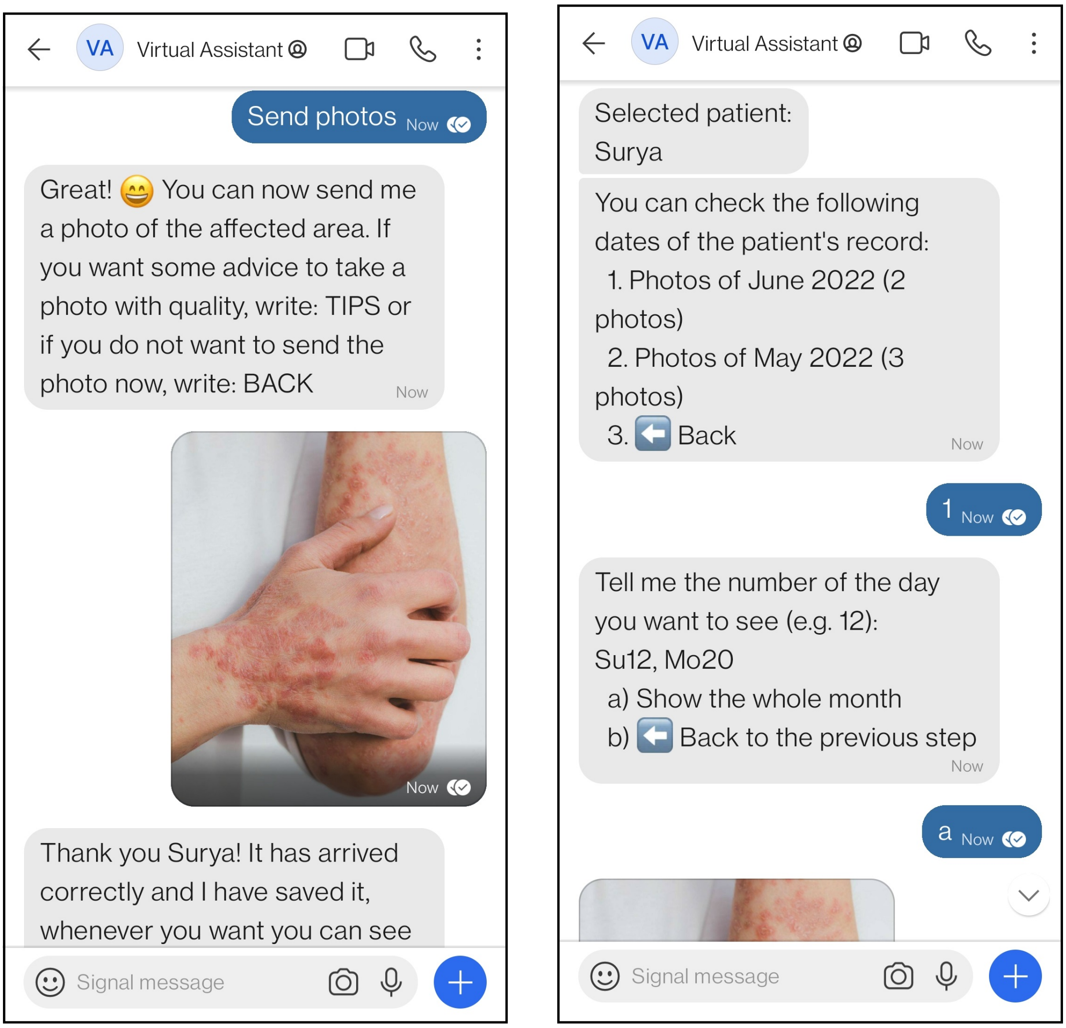
As an example of how the virtual assistant works, Figure 5.7 shows the conversation between a patient and the virtual assistant while the patient is doing a medical consultation (Figure 5.7a) and the dermatologist answering the query through the virtual assistant (Figure 5.7b). Another example of how the system works is shown in Figure 5.8, where a patient is saving a photograph (Figure 5.8a), and the dermatologist is obtaining the saved photos of a patient through the virtual assistant (Figure 5.8b).



(a) Patient interaction.

(b) Healthcare professional interaction.

Figure 5.7: Example of real interactions with the virtual assistant using *Medical consultation* functionality.



(a) Patient interaction using *Send photos* functionality.

(b) Healthcare professional interaction using *Record* functionality.

Figure 5.8: Example of real interactions with the virtual assistant.

5.4 Conclusions

In this chapter, we presented the development of a generic virtual assistant along with the extension for two concrete scenarios. The generic virtual assistant provides generic functionalities with the aim of caring for different chronic diseases and makes some additional contributions in terms of automation to translate an FHIR resource into an AIML interaction file, making the conversations more personal.

Moreover, this chapter included the implementation of two specific prototypes for different scenarios. The first specific scenario is tailored for the type 2 diabetes mellitus and depressive disorder profile, which provides help in the management of medication administration and intake (sending reminders, notifications, and summary of medication intakes). This prototype uses menu-based interaction in general and both menu-based and human-like interaction for medication management. The second specific scenario is developed for the psoriasis profile, which provides a photographic record for skin area tracking and a teleconsultation feature.

Chapter 6

Clinical Evaluation

Previous chapters focused on the technical evaluation of the proposed research. Nevertheless, following the idea described in the thesis approach regarding how our system can help in home-based caring scenarios, this thesis includes two clinical evaluations with the objective of studying the real impact on the daily life of patients and healthcare professionals. Therefore, this chapter aims to validate the effectiveness of the proposed healthcare virtual assistant in the AIML version, integrated within messaging platforms, with the aim of improving medication adherence in patients with comorbid type 2 diabetes mellitus and depressive disorder (scenario 1) and investigating the impact on the quality of life when the virtual assistant is used to connect patients with psoriasis and their dermatologists (scenario 2).

6.1 Patients with comorbid type 2 diabetes mellitus and depressive disorder

This section focuses on a virtual assistant for patients with type 2 diabetes mellitus and depressive disorder, with the aim of improving medication adherence through medication reminders (prototype developed in Section 5.2). The study examined medical variables (the level of glycosylated hemoglobin (HbA1c), the patient health questionnaire (PHQ-9), the medication adherence, and the number of medical appointments), as well as the acceptance and real use of the virtual assistant to validate our approach. Specifically, from all the functionalities offered, the nurses included in this study decided to use and evaluate *Medication*, *Medical appointments*, *Reminders*, *Suggestions*, and *Tutorials*. Moreover, the type of interaction used in this study was menu-based because the nurses found it easier to use and thought the patients would feel more comfortable with this type of interaction.

6.1.1 Materials and methods

6.1.1.1 Initial setup

We designed and implemented a nine-month pilot study with patients with comorbid type 2 diabetes mellitus and depressive disorder to test a virtual assistant developed to improve medication adherence. Participants received detailed information about the study, the procedure, the virtual assistant and their privacy and anonymity. A document containing all the information and contact details of the research team was given to the participants. Once they acknowledged that they had understood the information, a written and signed informed consent was obtained from all the participants of the study (patients and healthcare professionals). When the participants interacted with the virtual assistant for the very first time, they needed to confirm that they had signed the written consent to participate in this study.

The research team explained to the nurses how to use and configure the virtual assistant. The nurses then explained and configured the virtual assistant for their patients during their medical appointments. The initial configuration consisted of assistance in downloading the Signal app, explaining the first interaction with the virtual assistant, the registration of the patient in the platform, and the configuration of the medication and the reminder functions.

6.1.1.2 Participants

Eligible participants were patients with type 2 diabetes and depressive disorder who were 18 years old or older. The reference code of the diseases listed in the *International Classification of Primary Care* (2nd Edition) are T90 (type 2 diabetes) and P76 (depressive disorder). The participants in the experiment were recruited from the primary healthcare center Las Fuentes Norte in Zaragoza, Spain. Potential patients were recruited by the nurses working in the primary healthcare center. They were asked in their regular visits to the healthcare center to confirm their availability and willingness to participate. The experimental research design was a pre/post design, i.e., a comparison of outcomes in the same group of patients before and after the planned intervention of the virtual assistant.

The inclusion criteria of the patients were the following:

- Patients must have regular appointments with the nurses.
- Patients need to take medication every day.
- Patients have poor medication adherence. Poor adherence is measured with the level of the medication possession ratio (MPR). MPR is the division between the number of drug units prescribed for a specific period divided by the number of days [126]. The MPR value is capped at 100%. A presence/absence of medication adherence is calculated with a binary variable. When the MPR value is below 80%, the medication adherence is considered as absence [127].
- Patients can read and understand Spanish.
- Patients have the ability to write a message in a messaging platform.
- Patients need to have a smartphone with Android or iOS, and they need to have access to the Internet in their smartphones.
- Excluded from the study were subjects with cognitive, visual, or physical impairments that would interfere with the use of the virtual assistant.

The inclusion criteria of the healthcare professionals were the following:

- Healthcare professionals have a clinical interview experience.
- Healthcare professionals receive regular visits from patients with type 2 diabetes and depressive disorder.
- Healthcare professionals can read and understand Spanish.
- Healthcare professionals have the ability to write a message in a messaging platform.

- Healthcare professionals need to have a smartphone with Android or iOS, and they need to have access to the Internet in their smartphones.
- Excluded from the study were subjects with cognitive, visual, or physical impairments that would interfere with the use of the virtual assistant.

6.1.1.3 Medical outcomes measures

The medical effectiveness of the virtual assistant was measured by the level of glycosylated hemoglobin (HbA1c) in patients, which gives the mean level of blood glucose for the previous three months. The information provided by HbA1c allows the progression of the patients in the management of their diabetes to be evaluated. The higher the HbA1c values, the higher the risk of having complications related to diabetes. HbA1c measurements were obtained from a blood test done in a laboratory. The depression was measured using the patient health questionnaire (PHQ-9) [128], which monitors the severity of depression and the response to a medical treatment. Higher values in the PHQ-9 score indicate more severe depressive symptoms. The medication adherence was evaluated using the MPR value, with 80% as the threshold. The impact on healthcare resources was evaluated with the number of medical appointments per month for each patient. All the medical outcomes measurements were obtained at baseline and after 9 months.

6.1.1.4 Use outcomes measures

The patients' use and acceptance of the virtual assistant were measured as follows: (a) the use of the virtual assistant was measured by the number of interactions every day. The type of interaction was studied in order to ascertain which kind of interaction the patients prefer (numeric or text-based interaction). (b) The use of tools was measured with the number of functionalities used every day. (c) The use of reminders was measured by the number of reminders answered. (d) The acceptance was measured by the number of patients that did not uninstall Signal. (e) The usefulness or otherwise of the assistant was measured by the number of times the patient was not understood by the virtual assistant. All use outcome measurements were taken 9 months after the start of the study.

6.1.1.5 Participant interviews

Two sets of interviews were conducted during the study. The first interviews were conducted to obtain the patients' opinions after three months of interaction. The second interviews were held to obtain the results and the final opinion of the participants.

In the first set of interviews, patients were asked for their opinions three months after the outset of the study. The interviews were designed to obtain the general impressions of the patients as well as to identify any issues and the suggestions for the improvement of the virtual assistant. An important objective was to gather ideas for the content of the YouTube tutorials to help patients in the use of the virtual assistant. These guided interviews took place in the primary healthcare center during the appointments of the patients with the nurses involved in the study.

The second set of interviews were conducted at the end of the study in order to obtain information about the HbA1c values, the PHQ-9 scores, the number of medical appointments and the opinion of the participants. During these post-study interviews, patients and healthcare professionals were asked their overall opinions about the study. The answers with multiple options for the patients were weighted with the following scale: always (5), almost always (4), sometimes (3), rarely (2), and never (1). The answers with multiple options for the healthcare professionals were weighted with the following scale: all the patients (5), almost all the patients (4), some patients (3), almost no patient (2) no patient (1). The patient questionnaire was focused on obtaining the differences in medication intake before and after the use of the virtual assistant, and the usefulness of the virtual assistant in their lives. This questionnaire was completed via telephone calls made by the nurses involved in the study. The healthcare professionals' questionnaire included a question about their perception of the medication intake improvement of their patients. Interviews with healthcare professionals were self-administered.

The intermediate and post-study interview questions are provided in Appendices B and C.

6.1.1.6 Ethical aspects

The study protocol was approved and registered by the Comité de Ética de la Investigación de la Comunidad Autónoma de Aragón (CEICA) [129] on 13 March, 2019 (minutes nº 05/2019). The CEICA committee acts in accordance with the Declaration of Helsinki (last modified in 2013) and with the Good Clinical Practice (GCP) standard. The study complied with both national data protection law LO 03/2018 and European GDPR, providing the required measures of privacy and users' rights.

6.1.1.7 Statistical analysis

In order to describe continuous variables, mean and standard deviation (SD) were calculated, and to describe categorical variables, frequency and percentage were used. To compare the outcomes before and after the usage of the virtual assistant, the McNemar's test was estimated for dichotomous variables (MPR score) and the Wilcoxon signed-rank test for continuous variables (HbA1c values, PHQ-9 score, medical appointments per month and post-study interview). Variables measured before and after were considered significantly different when the *p*-value was less than 0.05. All statistical analyses were conducted using R software, version 4.0.1.

6.1.2 Results

6.1.2.1 Participants

In total, 15 patients completed a baseline assessment for eligibility (as shown in Figure 6.1). One patient needed to be excluded from the pilot study because the smartphone did not have enough space to install new applications. After selecting and registering the patients for the study, one patient did not interact with the virtual assistant. Finally, a total of 5 healthcare professionals and 13 patients with type 2 diabetes mellitus and depressive disorder participated and completed the study from 9 May, 2019 to 9 February, 2020. The average age of patients was 63.8 years (SD 9.1) (range 44–83 years) and 69% of them were females (9/13). The patients needed to take an average of five different drugs per day. They also had periodical medical appointments every two or three months with their nurses if needed.

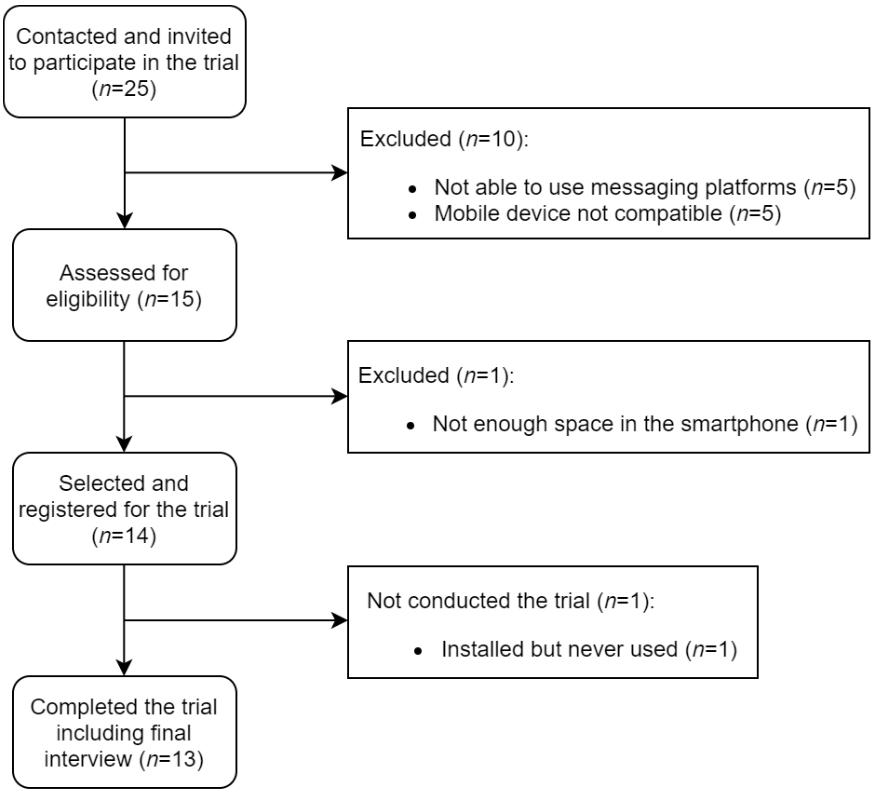


Figure 6.1: Flow diagram of patient selection and completion of the pilot study.

6.1.2.2 Evaluation outcomes

Medical outcomes The results related to medical outcomes are shown in Tables 6.1 and 6.2. The average level of HbA1c improved from 7.6 (SD 0.7) to 7.3 (SD 0.8) (results significantly different). The average PHQ-9 scores improved from 13.2 (SD 6.0) to 8.6 (3.6) (results significantly different). Furthermore, the average number of medical appointments per month was reduced from 2.0 (SD 2.6) to 1.3 (SD 1.5). We obtained the MPR value dividing the number of drugs taken by the number of days for which the medication was scheduled. Every patient presents a different number of days in the MPR estimation since each patient adapted the use of the virtual assistant to his/her own medication and needs. The MPR value was capped at 100%, and the threshold for positive medication adherence was 80%. A total of 76.9% (10/13) of the patients demonstrated positive medication adherence (note that at the beginning of the study none of them showed positive medication adherence). Medication adherence of patients before and after the study was significantly different ($p = 0.004$).

Patient	HbA1c[1] values (%) ($p = 0.02$)		PHQ-9[2] score ($p = 0.002$)	
	Before	After	Before	After
1 *	7.6	7.4	11	9
2	7.4	6.9	5	4
3	7.5	7.2	18	10
4	6.9	6.2	6	4
5	7.8	7.6	9	5
6	7.8	6.9	22	12
7 *	7.9	7.9	14	8
8	9.3	9.2	5	5
9	6.9	6.9	16	7
10	7.9	6.9	20	14
11 *	7.8	8.0	21	15
12	6.7	6.8	13	10
13	7.5	7.2	12	9

* Patients who had uninstalled Signal: results calculated using the period they had Signal installed.
[1] Level of glycosylated hemoglobin; [2] Patient health questionnaire.

Table 6.1: Level of glycosylated hemoglobin and PHQ-9 for each patient before and after using the virtual assistant.

Patient	**Number of medical appointments per month (p = 0.10)**		**MPR[1] %**
	Before	**After**	
1 *	2	2	100.0
2	10	6	86.3
3	1	1	19.8
4	1	1	100.0
5	4	1	100.0
6	1	1	83.0
7 *	1	1	86.7
8	1	0	91.7
9	1	1	27.1
10	2	1	92.9
11 *	0.3	0.3	16.5
12	1	1	100.0
13	1	1	100.0

* Patients who had uninstalled Signal: results calculated using the period they had Signal installed.
[1] Medication possession ratio.

Table 6.2: Medical appointments per month for each patient before and after using the virtual assistant and MPR values after using the virtual assistant.

Use outcomes The number of messages sent every day from the user to the virtual assistant (normalized by the number of active patients) is shown in Figure 6.2. The average number of messages sent per day by a patient was 2.7 (SD 2.9). A significantly high number of interactions can be observed on the first day of the study compared with the rest of the days. This is because all the participants had to take part in an initial tutorial where the basic use was explained. The rest of the peaks during the study mean that one or more patients were exploring the different options of the virtual assistant. In general, most of the patients interacted with the virtual assistant uniformly. A total of 88.5% of the interactions (4278/4835) were made using numeric-based messages (i.e., from the virtual assistant-offered choices, the user selects the choice by writing the number that identifies the choice), in contrast with 11.5% of text-based messages (i.e., by typing the text of the choice) sent (557/4835).

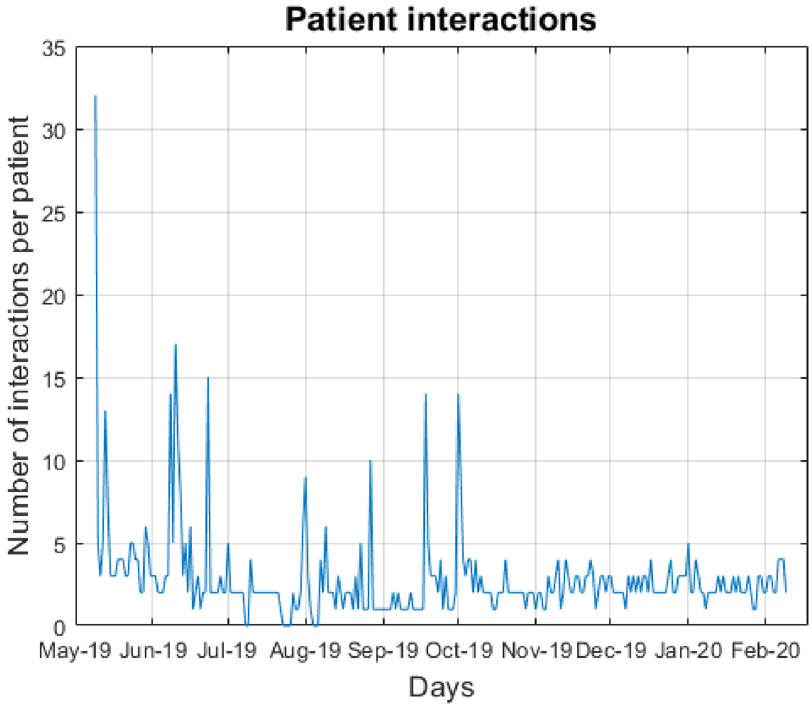


Figure 6.2: Patient interactions with the virtual assistant over time (normalized by the number of active patients).

The messages sent by the patients to the virtual assistant, classified by functionality in percentage terms, are presented in Table 6.3. The messages sent by the virtual assistant were not included because some reminders were not answered by the patient, meaning that the functionality was not used. It can be observed that the functionalities *Reminders* and *Medication* were the most used during the study. The patients used other functionalities, such as suggestions or medical appointments, but the prevailing usage of the virtual assistant was responding to reminders.

Functionality	**Messages sent by the patients by functionality, % (n) ($N = 4069$)**
Reminders	59.2 (2408)
Medication	37.9 (1543)
Suggestions	1.5 (60)
Medical appointments	1.1 (43)
Tutorials	0.4 (15)

Table 6.3: Analysis of the functionalities.

The average number of medication reminders configured in the virtual assistant was 3.1.

The use of reminders is presented in Table 6.4, where the number and the percentage of the reminders answered by each patient can be observed. A total of 74.4% (2184/2936) of the reminders were answered by the patients; 23.1% of users (3/13) answered less than 50% of their reminders. Three of these patients uninstalled Signal so were unable to receive any of the reminders on their phones.

Patient	**Reminders answered "Yes" or "No"), % (n/N)**	**Reminders not answered, % (n/N)**	**Number of times the medication was taken (by "Yes" or manually introduced), % (n/N)**
1 *	89.8 (150/167)	10.2 (17/167)	90.4 (151/167)
2	85.5 (194/227)	14.5 (33/227)	85.9 (195/227)
3	17.9 (15/84)	82.1 (69/84)	20.2 (17/84)
4	94.0 (638/679)	6.0 (41/679)	95.9 (651/679)
5	88.3 (218/247)	11.7 (29/247)	87.9 (217/247)
6	80.5 (140/174)	19.5 (34/174)	81.6 (142/174)
7 *	82.8 (24/29)	17.2 (5/29)	89.7 (26/29)
8	91.0 (132/145)	9.0 (13/145)	91.0 (132/145)
9	13.5 (26/192)	86.5 (166/192)	13.5 (26/192)
10	79.9 (147/184)	20.1 (37/184)	92.9 (171/184)
11 *	11.0 (10/91)	89.0 (81/91)	16.5 (15/91)
12	90.9 (261/287)	9.1 (26/287)	91.6 (263/287)
13	53.3 (229/430)	46.7 (201/430)	83.3 (358/430)

* Patients who had uninstalled Signal: results calculated using the period they had Signal installed.

Table 6.4: Results of reminders and medication activity.

The total medication taken by the patients is also presented in Table 6.4. Indeed, because the medication intakes can be reported by answering the reminders and also through the medication menu, we observe that patients 10 and 13 took the medication even though they did not answer the reminders. The HbA1c, the PHQ-9 and the MPR values of these two patients improved (see Tables 6.1 and 6.2), showing that besides the reminders, being aware of the possibility of informing the virtual assistant about the medication intakes helps in improving the adherence.

The times the patient was not understood and the functionalities usage are shown in Table 6.5. A total of 125 messages (2.6% (125/4835) of the total) sent by patients were not understood by the virtual assistant. In addition, the mean of the functionalities used by patients was 2.5 (SD 0.9) (69% of the patients (9/13) only used the functionalities *Medication* and *Reminders*, indicating that patients prefer to have as simple a virtual assistant as possible). Finally, around 23% of the patients (3/13) uninstalled Signal during the study.

Patient	**Number of times the patient was not understood**	**Total number of functionalities used**	**Nº of functionalities used every day, mean (SD[1])**
1 *	5	2	1.3 (0.4)
2	31	2	0.9 (0.6)
3	1	2	0.5 (0.7)
4	4	4	1.1 (0.5)
5	0	2	1.0 (0.4)
6	0	2	0.8 (0.5)
7 *	2	2	0.9 (0.6)
8	0	2	0.8 (0.6)
9	30	4	0.4 (0.6)
10	2	2	1.1 (0.5)
11 *	0	3	0.2 (0.4)
12	15	2	1.1 (0.3)
13	35	4	1.7 (0.7)

* Patients who had uninstalled Signal: results calculated using the period they had Signal installed.
[1] Standard deviation.

Table 6.5: Results of interaction and functionalities information.

6.1.2.3 Participant interviews

Intermediate opinion interview The opinion interviews took place during October 2019. All patients participating in the study agreed that the virtual assistant is useful, and they also reported that it will be useful with patients older than themselves. They all used the virtual assistant daily and it covered their medication reminder needs. Around

38% of the patients (5/13) found difficulties in learning how to use the virtual assistant, such as not being comfortable with the menu system (based on options with numbers). A total of 15.4% of the patients (2/13) reported that the virtual assistant was not able to understand what they said.

All the patients agreed that the language and the vocabulary used by the virtual assistant were appropriate. However, around 15% of the patients (2/13) expressed the opinion that the clarity of the functionalities used in the virtual assistant was poor. One patient reported a problem of not knowing how to answer the reminders. The answers of the patients showed that they did not have any extra necessities apart from those provided by the virtual assistant. Additionally, one patient noted that the virtual assistant provided company in daily life and another patient said that the word size should be bigger.

Based on the answers given by the patients, we concluded that no technical changes in the virtual assistant were needed, but tutorials explaining the interaction and the interface (e.g., explaining how to make the size bigger in Signal) were needed.

Post-study interview The outcomes related to the patients' answers are shown in Tables 6.6 and 6.7. We can observe an overall improvement in medication management. Statistically significant outcomes related to taking the medication at the set time and taking the prescribed dosages were observed. All of the patients were already attending all of the medical appointments before the study, so no improvement was obtained. A statistically significant improvement was observed in the accommodation of medication schedules to their daily lives (from 3.6 to 4.1), whereas the outcomes related to the supervision of their family or friends are not significantly different. There was a notable improvement from 4.2 to 5 (all patients answered always) for completing the medical treatment without supervision and a decrease from 2.8 to 2.1 in the difficulty of remembering to take the medicines. Both outcomes were significantly different as it is shown in Table 6.6.

Statement/Question	**Before, mean (SD[1])**	**After, mean (SD[1])**	***p*-value**
You take the medications at the set time	4.2 (0.7)	4.8 (0.4)	0.01
You take all the indicated dosages	4.5 (0.5)	4.9 (0.3)	0.02
You attend your medical appointments	5.0 (0.0)	5.0 (0.0)	-
You accommodate your medication schedules to your daily life activities	3.6 (1.0)	4.1 (0.6)	0.02
Your family or friends are involved in your care	2.2 (1.4)	2.5 (1.3)	0.15
You complete the medical treatment without supervision of your family or friends	4.2 (0.8)	5.0 (0.0)	0.02
How often is it difficult to remember that you should take all your medicines?	2.8 (1.3)	2.1 (1.3)	0.04

[1] Standard deviation.

Table 6.6: Comparison between before and after using the virtual assistant.

Question	**Yes, % (*n*) (*N* = 13)**	**No, % (*n*) (*N* = 13)**
Do you ever forget to take medications to treat your disease?	23.1 (3)	76.9 (10)
When you are well, do you stop taking your medication?	0.0	100.0
If you ever feel bad, do you stop taking it?	15.4 (2)	84.6 (11)
Did you find it easy to use the virtual assistant?	92.3 (12)	7.7 (1)
Do you think it is useful for you?	92.3 (12)	7.7 (1)
Do you think the virtual assistant improves your medication adherence?	92.3 (12)	7.7 (1)
Do you think you will continue using the virtual assistant after the project?	69.2 (9)	30.8 (4)
From a certain moment, the virtual assistant began to give the weather forecast. Has this helped you to use the virtual assistant more frequently?	100.0	0.0

Table 6.7: Patients' responses.

The outcomes related to the healthcare professionals' answers are shown in Tables 6.8

and 6.9. In the comments, the healthcare professionals mentioned that the virtual assistant sometimes failed, creating uncertainty and mistrust. One nurse also noted that the messages about the weather forecast were repetitive and created a lack of interest. Finally, one of the nurses mentioned that some patients found the process of saying that they had taken the medication of the day even if they had not answered the reminder to be too long.

Question	**Yes, % (n) ($N = 5$)**	**No, % (n) ($N = 5$)**
Did you find the virtual assistant easy to use?	80.0 (4)	20.0 (1)
Do you think it is useful for you?	80.0 (4)	20.0 (1)
Do you think it is useful for the patients?	100.0	0.0
Do you think the virtual assistant improves the medication adherence of the patients?	100.0	0.0
Did you use the functionalities to monitor the medication intakes of your patients?	60.0 (3)	40.0 (2)
Do you think you will continue using the virtual assistant after the project?	80.0 (4)	20.0 (1)

Table 6.8: Healthcare professionals' responses.

Question	**Mean (SD[1])**
In the medical appointment, did you observe motivation in using the virtual assistant on the part of the patients?	4.2 (1.3)
In the medical appointment, did you observe dissatisfaction with the virtual assistant on the part of the patients?	1.8 (0.8)
In the medical appointment, did you observe an improvement in the patients' medication intake?	4.2 (0.8)

[1] Standard deviation.

Table 6.9: Healthcare professionals' responses.

6.1.3 Discussion

6.1.3.1 Principal findings

The medical outcomes of the study show a moderate effect in the improvement of HbA1c, with a difference in the mean of -0.3%. This result is similar to the finding obtained by Cui et al. [71], where a difference in the HbA1c mean of −0.4% is observed. Other studies related to tele-assistance systems observed a difference in the HbA1c mean of −0.22% [130] and −0.4% [131]. Most of the patients who were more active in the use of the virtual assistant showed more improvements in the HbA1c measure. Moreover, the PHQ-9 score and the number of medical appointments per month were significantly reduced, with a mean difference in PHQ-9 of −4.6 and in the medical appointments per month of −0.7. Other works that used smartphone applications to provide interactive skills training or assess treatment progress to patients with depression were evaluated by Mohr et al. [132] and McCue et al. [133], which obtained an improvement in the PHQ-9 mean difference of −6.1 and −9.5, respectively. These studies observed an improvement in PHQ-9 results using digital health tools, which is consistent with our outcomes. A significant improvement was also obtained in medication adherence, with 76.9% of the patients (10/13) improving their performance after the study. Trials showed that higher medication adherence was associated with fewer emergency department visits, improved glycemic control, and lower medical costs [134]. Furthermore, the complexity of the clinical cases, such as multiple health conditions, is one of the factors that may influence medication errors, being the alert systems one of the next steps in person-centered care [135].

6.1.3.2 Patient acceptance

The experience with the virtual assistant seems to have been very positive, as observed in the patients' opinions and interactions throughout the study. Almost 70% of the patients (9/13) agreed with the idea of continuing using the virtual assistant after the study. This finding is consistent with the results reported by Nadarzynski et al. [136], where the acceptability of virtual assistants in healthcare was 67%. While observing the acceptance rate, the human factors and usability issues need to be considered, since the average age of the patients was 63.8 years and the oldest patient participating in the study was 83 years old. Most of the patients that stopped using the virtual assistant did so at the beginning. They found the menu system of the dialogues inconvenient and had difficulties in being understood by the virtual assistant. An easier interaction (e.g., based on buttons instead of text) could be used in order to try to improve the acceptance for these patients who face difficulties in the interaction.

After studying the answers in the intermediate opinion interview, we created a YouTube channel [137] where we uploaded tutorials solving all the patients' doubts. We noticed that the number of views of our videos are sufficiently significant (with an average of 14 views (SD 6.5)) when compared with the number of participants, suggesting a positive acceptance of the videos.

We observed different behaviors in the patients. More than half of them were able to communicate fluently with the virtual assistant. Nevertheless, a few patients found various limitations while interacting with it. An analysis of the patients with difficulties who interacted with thirty messages or more showed that one patient had problems in correctly answering the reminders. However, after the second reminder ten minutes later, the patient was able to answer the reminder correctly the second time. We noted from Table 6.4 that patient 9 was not answering the reminders. This was because the virtual assistant was not able to understand the patient's messages (we observed in Table 6.5 that 30 interactions were not understood). However, patient 13 had 34 interactions not understood, but this patient was able to indicate manually that the medication had been taken.

Some patients showed that they were keen to answer the virtual assistant, which may increase the attention and the responsibility of patients in taking their medication. Other patients who usually forgot to take their medication finally remembered to take it thanks to the alert on their mobile phones.

Three of the patients yielded low MPR values, as shown in Table 6.2. The main reason for this result is because two of them found difficulties while interacting with the virtual assistant, and the third one was only active for a few days during some months and, finally, uninstalled Signal.

Before the study, all of the patients were attending all the medical appointments, so no improvement related to the attendance was observed. Nevertheless, a reduction in the number of medical appointments per month was found in 30.8% of the patients (4/13), who went less frequently to the consultation when compared to the beginning of the study. This reduction in medical appointments implies a reduction in associated healthcare resources.

6.1.3.3 Healthcare professional acceptance

The healthcare professionals who participated in this study found the virtual assistant useful for the patients. As observed in Table 6.8, a total of 80% of them (4/5) agreed that the virtual assistant is useful and easy to use. In addition, all of them believed that the virtual assistant improves the medication adherence of the patients. The main use of the virtual assistant by the healthcare professionals was to register the patients and to check if their medication was correctly configured. Moreover, 60% of them (3/5) were interested in monitoring the medication intake of the patients, using the functionalities offered by the virtual assistant. Since their use of the virtual assistant was more sporadic and they did not need to use it every day, some of the professionals found the weather forecast messages repetitive and wanted to disable them.

6.1.3.4 Limitations

This study has several limitations. The virtual assistant is designed to ask the patient about the total period of the medication, but we observed that some patients did not update the information after the initial configuration for several reasons, such as forgetting to do so, being unaware of needing to do so, etc. In such cases, the virtual assistant stops sending reminders about the medication and consequently patients usually forget to use the virtual assistant. In addition, the patients need to have digital literacy or have people around them that can help if they find difficulties in configuring the messaging platform, or if they accidentally disable the sound of the reminders.

6.2 Patients with psoriasis

This section aims to investigate the impact on the quality of life of patients with psoriasis using a virtual assistant, which includes specific functionalities such as remote medical consultations with dermatologists and storage and viewing of patient photographs (prototype presented in Section 5.3). This work also examines the usability and acceptance of the virtual assistant as well as the quality-of-life questionnaires and medication adherence scales to validate our virtual-assistant-integrated system. Specifically, the healthcare professionals decided to evaluate the following functionalities in detail: *Questionnaires*, *Medical consultation*, *Send photos*, and *Record*.

6.2.1 Materials and methods

6.2.1.1 Study design

We designed and implemented a one-year prospective study with psoriatic patients and dermatologists to test the usage of a virtual assistant and its impact on the patient's quality of life. The study was conducted from 22 April 2021 to 22 April 2022. The virtual assistant used in the system has two-fold usage: (1) as a tool to connect patients with healthcare professionals and vice versa (teledermatology); (2) as a monitoring tool for disease management (questionnaires, medication administration, and image records). We analyzed whether a virtual assistant could improve patients' quality of life by using a virtual assistant that connects patients with healthcare professionals through online medical consultations.

Participants who met the inclusion criteria received detailed information about the study, their privacy, and anonymity, and were invited to participate in the study. A written and signed informed consent was obtained from all the participants (patients and healthcare professionals). The healthcare professionals did face-to-face personal interviews with the patients to include them in the study. In face-to-face medical consultations, the healthcare professionals explained to the patients how the virtual assistant works, helped them to download and install the messaging platform, and registered them. They also explained how a medical consultation can be performed through the virtual assistant.

6.2.1.2 Participants

Eligible participants were healthcare professionals and patients with psoriasis from the Miguel Servet University Hospital in Zaragoza, Spain. Healthcare professionals were 18 years old or older. The sample size included in the study was based on the availability of healthcare professionals and patients who met the requirements for inclusion criteria. Specifically, patients were recruited with the following inclusion criteria:

- Moderate–severe psoriasis under follow-up in the monographic psoriasis consultation (a consultation in which only one pathology is attended, in this case, patients with psoriasis), defined as one or more of the following points:
  - PASI (Psoriasis Area Severity Index) > 10.
  - BSA (Body Surface Area) > 10.
  - IGA (Investigator Global Assessment) scale levels 3 or 4.
  - DLQI (Dermatology Life Quality Index) [138] > 5.
  - Classic systemic treatment.
  - Biological systemic treatment.
- Age between 18 and 65 years old.
- Patients can read and understand Spanish.
- Own a smartphone with Android or iOS and internet access.
- Patients have not used other eHealth apps or chat-based care platforms before.
- The patient demonstrates a good level of technological knowledge and handling of smartphones (i.e., the patient needs to know how to use a smartphone and its apps with ease). The technological knowledge was assessed based on the impression of healthcare professionals.
- Excluded from the study were subjects with cognitive, visual, or physical impairments that would interfere with the use of the virtual assistant and patients without a smartphone.

6.2.1.3 Outcomes measures

One relevant parameter was the total number of medical consultations done. Moreover, another relevant parameter was the average number of photos per patient (who stored photographs in the system), which was calculated using the total number of photos sent and stored and the number of patients who used *Send photos* functionality to do so.

The answers to clinical questionnaires from all patients were evaluated before (at the beginning of the study) and after using the virtual assistant. The selected questionnaires were the standard and widely validated ones used in all pivotal trials of psoriasis drugs and other psoriasis studies. The percentage of questionnaires answered was also measured. Additionally, the number of patients active in answering the questionnaires was also obtained (a patient is considered active when the percentage of questionnaires answered is higher or equal to 50%). The selected questionnaires were:

- Psoriasis Quality of Life (PSOLIFE) [139]: PSOLIFE is a psoriasis quality of life questionnaire consisting of 20-item responses with a range from 20 to 100 points. Higher values of PSOLIFE score mean a better quality of life related to health (or less impact on the quality of life) [140].

- Dermatology Life Quality Index (DLQI): DLQI is a dermatological quality of life questionnaire consisting of 10-item responses with a range from 0 to 30 points. 0 means the patient does not have any problem and 30 means the illness in the patient has a severe impact.

- Treatment Satisfaction Questionnaire for Medication (TSQM) [141]: TSQM is a measure widely used to assess treatment satisfaction. TSQM scores on four different scales: effectiveness, side effects, convenience, and global satisfaction, each from 0 to 100. Higher values of TSQM scores mean higher satisfaction.

- Eight-item Morisky Medication Adherence Scale (MMAS-8) [142]: MMAS-8 is a widely used questionnaire that measures medication-taking behavior and consists of 8-item responses with a range from 0 to 8. It is measured using the following criteria: Items 1, 2, 3, 4, 6, and 7: Yes is 0 and No is 1. Item 5: Yes is 1 and No is 0. Item 8: Never/Rarely is 1, From time to time is 0.75, Sometimes is 0.5, Normally is 0.25 and Always is 0. A score of 8 reflects high adherence, values of 7 or 6 reflect medium adherence, and scores lower than 6 reflect low adherence.

In addition, some follow-up questionnaires were asked to observe the progress of their quality of life. These questionnaires were related to sleep, alcohol, and positiveness. Furthermore, to analyze the participants' use of the virtual assistant, the number of messages sent to each functionality was measured.

Finally, a satisfaction survey was conducted with all participants in the study to obtain their opinion and satisfaction with the virtual assistant. We used the Spanish version [143] of the System Usability Scale (SUS) [144] complemented with some questions from the mHealth App Usability Questionnaire (MAUQ) [145] (a few items adapted from MAUQ Usefulness (MAUQ_U)). The SUS score range is from 0 to 100. The results are considered above average when the SUS score is above 68, and results below 68 are below average. The final version of the surveys can be accessed in Appendices D and E. At the end of the study, the participants received the link to the satisfaction survey, and after a few days, the healthcare professionals made phone calls to remind them about the importance of answering the satisfaction survey. The surveys were anonymized once they were obtained to increase honesty and decrease bias. The answers with multiple options for the participants' satisfaction survey were weighted with the following scale: totally agree (7), quite agree (6), somewhat agree (5), neutral (4), somewhat disagree (3), quite disagree (2), and totally disagree (1); too easy (5), easy (4), somewhat easy (3), difficult (2) and too difficult (1); always (5), almost always (4), sometimes (3), rarely (2), and never (1).

6.2.1.4 Ethical aspects

The study provides the required measures of privacy and users' rights by complying with both national data protection law LO 03/2018 and European GDPR. The study protocol was approved and registered by the CEICA committee on 7 October 2020 (minutes nº 19/2020). The CEICA committee acts in accordance with the Declaration of Helsinki (last modified in 2013) and with the GCP standard.

6.2.1.5 Statistical analysis

We used frequency and percentage to describe categorical variables while for the continuous variables, we used mean and standard deviation (SD). The effect size was calculated using Cohen's D. To check the normality of the data, the Shapiro–Wilk normality test was used. The normality can be assumed when the p-value is more than 0.05. To compare the outcomes at the beginning and after using the virtual assistant, the Paired Samples T-test and Wilcoxon signed-rank test were estimated where appropriate for continuous variables (PSOLIFE, DLQI, TSQM, and MMAS-8 questionnaires, and satisfaction survey). Variables measured at the beginning and after were considered significantly different when the p-value was less than 0.05. All statistical analyses were conducted using R software, version 4.0.3.

6.2.2 Results

6.2.2.1 Participants

During the study, four professionals from the department of dermatology used the virtual assistant (three dermatologists and one nurse). Throughout the year, the healthcare professionals contacted 40 patients and invited them to use the virtual assistant (as shown in Figure 6.3). Three patients did not finish the process of installing the messaging platform and registering themselves with the virtual assistant. After the selection and registration, three of them did not continue using the virtual assistant after the first day. A total of 34 psoriatic patients used the virtual assistant (with at least 4 months of usage for the last patient registered in the virtual assistant). Finally, 30 patients completed the final satisfaction survey.

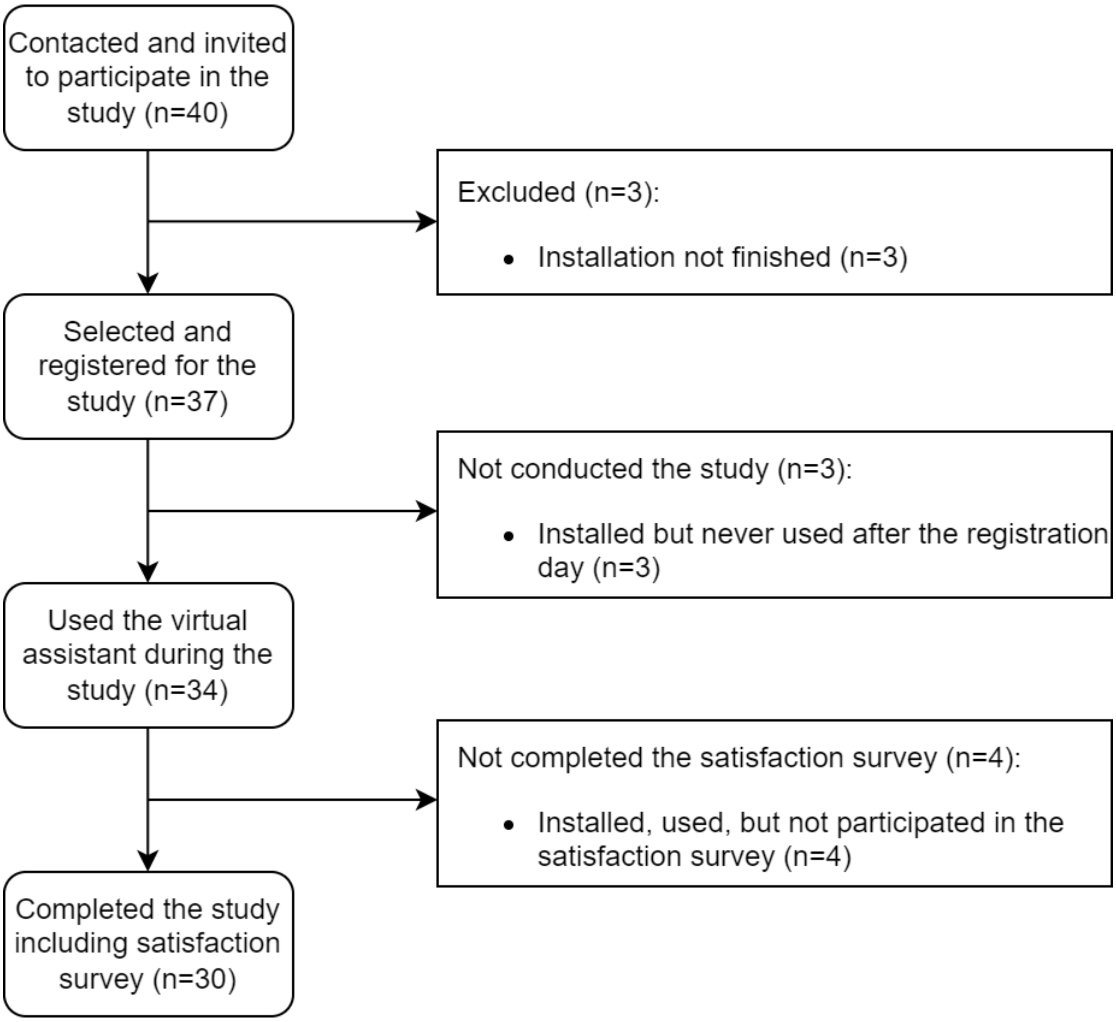


Figure 6.3: Patient selection, exclusion and completion of the study.

The composition of patients was 36.7% females (11/30) and 63.3% males (19/30). The average age of patients was 36.0 years old (SD 10.0) with a range from 18 to 58 years old. The healthcare professionals were 50% females (2/4) and 50% males (2/4). The average age of healthcare professionals was 35.3 years old (SD 7.0) with a range from 27 to 44 years old.

6.2.2.2 Evaluation outcomes

Patients used the virtual assistant for a total of 52 remote medical consultations with healthcare professionals. Moreover, patients stored 29 photos in the virtual assistant. The photos were stored by 13.3% of the patients (4/30) involved in the study, with an average of 7.3 photos (SD 5.4) by the patient who saved photographs.

The results of the questionnaire are shown in Table 6.10 (results are not significantly different). A total of 51.1% (284/556) of the questionnaires were answered by the patients. A total of 60.0% of patients (18/30) were active in answering the questionnaires. Average scores improved for PSOLIFE (from 63.8 to 64.8), DLQI (from 4.4 to 2.8), TSQM: Convenience (from 74.8 to 79.1), TSQM: Global satisfaction (from 63.7 to 68.5) and MMAS-8 (from 6.9 to 7.2). The first MMAS-8 questionnaire was asked on 20 October 2021, 6 months after the start of the study. The last MMAS-8 questionnaire was asked on 18 April 2022. Nevertheless, TSQM: Effectiveness (from 61.8 to 61.4) and TSQM: Side effects (from 25.7 to 18.4) show a decrease in the scores after patients used the virtual assistant.

The follow-up questionnaire results are shown in Table 6.11. The percentage of follow-up questionnaires answered was 59.9% (358/598). For the follow-up questionnaires, a total of 66.7% of patients (20/30) were actively answering them. We did not collect a baseline measurement for sleep, alcohol, or positiveness. The question to follow up on the sleep was "Did you sleep well today?" and 70.1% of the patients answered "Yes". Furthermore, the answer was "Nothing" in 74.6% of answers regarding alcohol intake. The average for the positiveness questionnaire was 3.8 (SD 1.0) (with the question "How do you think your day will go today? Range from 0 (very bad) to 5 (very good)").

Questionnaire [range]	**First time filled, average (SD[5])**	**Last time filled, average (SD[5])**	***p*-value**	**Effect size**
PSOLIFE[1] [20–100]	63.8 (16.9)	64.8 (15.4)	0.66	0.06
DLQI[2] [0–30]	4.4 (4.9)	2.8 (5.1)	0.04	0.31
TSQM[3]: Effectiveness [0–100]	61.8 (29.2)	61.4 (26.2)	0.95	0.01
TSQM[3]: Side effects [0–100]	25.7 (33.4)	18.4 (30.1)	0.31	0.23
TSQM[3]: Convenience [0–100]	74.8 (24.7)	79.1 (20.4)	0.43	0.19
TSQM[3]: Global satisfaction [0–100]	63.7 (24.4)	68.5 (18.6)	0.66	0.22
MMAS-8[4] [0–8]	6.9 (1.5)	7.2 (0.7)	0.57	0.28

[1] Psoriasis Quality of Life; [2] Dermatology Life Quality Index; [3] Treatment Satisfaction Questionnaire for Medication; [4] Eight-item Morisky Medication Adherence Scale; [5] Standard deviation.

Table 6.10: Average and SD of the first and last PSOLIFE, DLQI, TSQM, and MMAS-8 answered questionnaires.

Follow-up questionnaire	**Question**	**Answers**
Sleep	"Did you sleep well today?"	70.1% (117/167) of answers were "Yes"
Alcohol	"How much alcohol did you drink today?"	74.6% (47/63) of answers were "Nothing"
Positiveness	"How do you think your day will go today? Range from 0 (very bad) to 5 (very good)"	Average of 3.8 (SD[1] 1.0)

[1] Standard deviation.

Table 6.11: Results related to the follow-up questionnaires (sleep, alcohol, and positiveness).

The analysis of the functionalities used by participants is shown in Table 6.12. The table shows the percentage of messages sent by participants to the virtual assistant to observe the real usage of functionalities and avoid the notification messages related to questionnaires that are not answered (and therefore not used). We normalized the data by dividing the messages sent to each functionality by the total of messages sent by a participant. After that, we did the average of the normalized messages of all the participants. The results show that *Questionnaires* functionality is the main functionality used (78.0%), followed by *Medical consultation* functionality with 19.8%.

Functionality	**Normalized number of messages sent by participants, % (n) ($N = 34$)**
Questionnaires	78.0 (26.5)
Medical consultation	19.8 (6.7)
Record	1.4 (0.5)
Send photos	0.7 (0.2)

Table 6.12: Participants' usage analysis of the functionalities.

6.2.2.3 Participant satisfaction survey

The first outcome observed in the satisfaction survey is the SUS score of the participants of the study. The average SUS score is 70.1 (SD 15.2) (range of SUS score from 0 to 100). The outcomes from the satisfaction survey divided by patients and healthcare professionals are detailed in the following subsections.

Patients' survey outcomes The outcomes related to the patients' answers are shown in Tables 6.13 and 6.14. More than half of the patients (18/30) agree that *Medical consultation* functionality is the option they like most. *Send photos* functionality was chosen by 10% (3/30) of the patients. The rest of the patients chose other functionalities related to medication and appointments. One patient liked all the functionalities. Furthermore, patients agreed that it was easy (with an average of 4.3 (SD 1.0)) to answer questionnaires through the virtual assistant. Moreover, the face-to-face medical appointments attendance before and after using the virtual assistant decreased from 4.7 (SD 0.8) to 4.5 (SD 1.1), with an effect size of 0.21 (the result is not significantly different with $p = 0.27$).

The main reasons why patients left questionnaires unanswered were: I read the notification but then I forgot to answer (11/30), I don't have time (7/30), too long/too many/too monotonous (6/30), and I was not aware of the mobile notification (4/30).

Affirmations [range from 1 to 7]	**Mean (SD[1])**
The language used by the virtual assistant was appropriate.	6.0 (1.3)
The virtual assistant was useful to improve my quality of life.	4.8 (1.7)
The virtual assistant made it convenient for me to communicate with my health care provider.	5.5 (1.7)
I felt confident that any information I sent to my provider using the virtual assistant would be received.	5.4 (1.5)
I felt comfortable communicating with my health care provider using the virtual assistant.	5.4 (1.6)

[1] Standard deviation.

Table 6.13: Patients' responses regarding the virtual assistant utility.

Question	**Yes, % (n/N)**	**No, % (n/N)**
Did you do medical consultations with your dermatologist through the virtual assistant?	50.0 (15/30)	50.0 (15/30)
If you did medical consultations, were they resolved satisfactorily?	80.0 (12/15)	20.0 (3/15)
Has having a tool with which to be able to contact your dermatologist directly provided you with security/peace of mind?	83.3 (25/30)	16.7 (5/30)
Do you think the virtual assistant improves your treatment adherence?	66.7 (20/30)	33.3 (10/30)
Do you stop using the virtual assistant when you feel better with your symptoms?	23.3 (7/30)	76.7 (23/30)
Have you reduced the number of face-to-face medical consultations after using the virtual assistant?	16.7 (5/30)	83.3 (25/30)
Do you think you will continue using the virtual assistant after the project?	73.3 (22/30)	26.7 (8/30)

Table 6.14: Patients' responses regarding general use and perception of the virtual assistant.

Finally, some comments noted from patients were: "I use it frequently if I have a problem with my biological therapy, I find it extraordinary to be attended to at any time.", "Right now my psoriasis is under control but it will come in handy when I have a flare-up.", "If my psoriasis gets worse and I need a close follow-up, I would use the virtual assistant more.", and "I believe that it is a very necessary tool. The psoriasis condition can vary a lot from one medical consultation to another and it gives a lot of security to have it and to be able to consult the doctor or nurse if there is any mishap, doubt, or flare-up at that time without having to wait until the appointment arrives. It helps to have the disease more controlled. I would recommend it 100%.”.

Healthcare professionals’ survey outcomes The outcomes related to the healthcare professionals’ answers are shown in Tables 6.15 and 6.16. The functionality the healthcare professionals agree that they like the most is *Medical consultation*. Furthermore, the average score regarding how easy or difficult they find the usage of the *Questionnaires* functionality (generating or assigning questionnaires) was 3.8 (SD 0.5).

Affirmations [range from 1 to 7]	**Mean (SD[1])**
The language used by the virtual assistant was appropriate.	5.8 (0.5)
The virtual assistant was useful to improve the quality of life of my patients.	6.0 (0.8)
The virtual assistant made it convenient for me to communicate with my patients.	6.3 (1.0)
I felt confident that any information I sent to my patients using the virtual assistant would be received.	5.5 (0.6)
I felt comfortable communicating with my patients using the virtual assistant.	5.5 (1.0)

[1] Standard deviation.

Table 6.15: Healthcare professionals’ responses regarding the virtual assistant utility.

Question	**Yes, % (*n*) (*N* = 4)**	**No, % (*n*) (*N* = 4)**
Do you think the virtual assistant improves the patients' treatment adherence?	75.0 (3)	25.0 (1)
Was the quality of the images sufficient to make diagnoses?	100.0 (4)	0.0 (0)
Was it easy for you to answer patients' queries?	75.0 (3)	25.0 (1)
Have you seen the number of face-to-face medical consultations of patients reduced after using the virtual assistant?	100.0 (4)	0.0 (0)
Do you think you will continue using the virtual assistant after the project?	100.0 (4)	0.0 (0)

Table 6.16: Healthcare professionals' responses regarding general use and perception of the virtual assistant.

6.2.3 Discussion

6.2.3.1 Principal results

New messaging-based mobile phone systems for chronic diseases have been deployed, but such systems lack evaluation. Further research on evaluating mHealth interventions and their user acceptance should be addressed [146]. This study attempted to address the gap by conducting a one-year prospective study to evaluate the use of a virtual assistant in teledermatology. The virtual assistant is a complex and comprehensive system that facilitates daily psoriasis monitoring and patient-healthcare professional medical consultations.

The key findings highlighted the improvement in the patient's quality of life and the good usability of the virtual assistant. The improvement in the quality of life is demonstrated by higher PSOLIFE scores (with a mean difference of 1) and lower DLQI scores (mean difference of 1.6). These results agree with the findings of a study conducted by Kornmehl et al. [147], which found that DLQI scores increased by 4.1 using online management for atopic dermatitis after 12 months. The adherence also improved slightly (with a mean difference of 0.3), which may be due to the fact that patients pay attention to a tool related to their disease, and, therefore, they may pay more attention to the treatment. The availability of a physician anytime they require could have also helped.

This result is consistent with the study by Rhee et al. [148], which observed improvement in the awareness of symptoms and treatment adherence and sense of control with the use of their system. The result is also consistent with the answer in the survey related to treatment adherence, where 66.7% of patients (20/30) think the virtual assistant improved their adherence. Additionally, the above average (with a value of 70.1, above 68) SUS score obtained for the usability of the virtual assistant is consistent with the scores obtained for similar systems described in Ref. [149–151].

Only half of the patients (15/30) did consultations with their healthcare professionals through the virtual assistant. This result may be due to the fact that psoriasis is a disease based on flare-ups that are not regular and can or cannot appear in the patients during the one-year study period. Furthermore, patients stored 29 photos in the virtual assistant to save their progress (they sent photos of their psoriasis plaques, their nails, their hair, etc.). During the consultation, physicians could see the date on which the photographs were taken and assess the disease evolution.

6.2.3.2 Patient acceptance

The patients seemed to agree to the use of the virtual assistant as a total of 73.3% of patients (22/30) agreed to the idea of using a virtual assistant even after the project (Table 6.14). This outcome is consistent with the findings observed by Nadarzynski et al. [136], where the acceptability of virtual assistants in healthcare was 67%. Another key point to emphasize is that 83.3% of the patients (25/30) found having a tool to contact their dermatologist provides them with a sense of security or peace of mind as the virtual assistant enables them to consult the doctor any time between in-person consultations.

A total of 66.7% of patients (20/30) were active in answering the follow-up questionnaires, in contrast to 60.0% of patients (18/30) active in answering the clinical questionnaires (PSOLIFE, DLQI, etc.). The follow-up questionnaires were shorter than clinical questionnaires, which may explain why patients were more involved in answering them (indeed, 6 out of 30 patients agreed with the idea that the questionnaires were too long, too many, or too monotonous).

We observed that face-to-face medical consultations were not reduced in general (only 16.7% of patients (5/30) reduced the number of consultations after using the virtual assistant). This result is consistent with the idea discussed by Corbett et al. [152], which views telemedicine as an adjunct to in-person consultations, unable to replace them.

6.2.3.3 Healthcare professional acceptance

All the healthcare professionals agree with the idea of using the virtual assistant after the project. Moreover, a total of 75% of healthcare professionals agree with the idea that it is easy to answer patient queries. Answering patient queries is crucial because if a physician cannot answer their questions not knowing how to use it, the patient could feel frustrated, start feeling mistrust, and stop using the virtual assistant.

Another essential factor is the quality of the images. When an image is sent through a messaging platform, it is compressed, reducing the original quality of the image. The images should have enough quality to allow dermatologists to observe the disease conditions and make decisions accordingly. In the survey, all the healthcare professionals answered that the quality of the images was sufficient to make diagnoses. Thus, virtual assistant-integrated systems could also be used for other diseases such as dermatological affections, which need image support.

6.2.3.4 Limitations

This prospective study has several limitations. The reduced number of dermatologists participating in the study does not allow us to obtain statistically significant differences in the results. Therefore, further studies with more dermatologists are needed to generalize the results. In addition, the sample size was small, which should be extended to cover more samples in the future.

Sometimes patients need a dermatological examination that includes palpation, dermatoscopy, microbiology studies, Wood lamp examination, or biopsy (for example, when the disease or new conditions are diagnosed). Such an examination is not currently possible with the virtual assistant. A face-to-face consultation is required in such cases.

Additionally, patients not providing enough photographs to dermatologist during remote medical consultation has risks of misdiagnosis (e.g., other sites affected, atypical presentation, or clinical lookalikes).

The integration of new methods for healthcare professionals has burnout risks [153]. So, training of such methods should be conducted within their working hours. Our system allows healthcare professionals to set the virtual assistant notifications in silent mode while they are not working so that medical consultations are managed only while at work.

Furthermore, the patients' usability regarding medical consultations is closely related to the psoriasis flare-ups. As observed in the comments made by patients, some did not have outbursts during the study period, but, notwithstanding, they consider it will be helpful when they have a flare-up in the future. Moreover, the participants need to have digital literacy to use messaging platforms or have people who can help them if they find difficulties.

6.3 Conclusions

This chapter provided an overall perspective of how virtual assistants can affect patients' medication adherence, and the improvements and limitations that arose while using the virtual assistant under study. The findings of the study suggest that the use of virtual assistants can be useful and effective for improving patient medication adherence (patients answered 74.4% of the reminders received, the HbA1c mean improved 0.3%, and the PHQ-9 mean improved 4.6). The mean of the medical appointments per month decreased by 0.7 appointments per month, which supports the potential use of virtual assistants for reducing associated healthcare resources. Furthermore, our findings suggest that virtual assistants can provide a tool for healthcare professionals to help patients improve their adherence by configuring the medication themselves, or checking if the medication is correctly configured.

Furthermore, this chapter evaluated the use of virtual assistants in teledermatology, specifically, to connect patients with dermatologists remotely using a messaging platform. During the evaluation, the study provided an overall perspective of how the patient's quality of life is affected by the use of virtual assistants. The general usability of the virtual assistant was above average (with a SUS score of 70.1), and 26 out of 34 participants agreed to continue using the virtual assistant after the study. The patient's quality of life improved (with a mean difference of 1 and 1.6 for PSOLIFE and DLQI, respectively). Furthermore, our results suggest that virtual assistants can provide a tool to improve patient's treatment adherence (agreed by 66.7% of the patients). The use of virtual assistant also provided security or peace of mind (83.3% of the patients agreed with it) to patients as they could directly contact dermatologists.

Chapter 7

Conclusion and Future Directions

In this chapter, the research objectives achieved and the thesis' conclusions are reported. Moreover, future works are highlighted in the key areas where further extensions and solutions can be focused on in the future.

7.1 Research objectives achieved

Chapter 1 presented evidence of the main benefits that the new advances in technology can provide to eHealth and medical telemonitoring. More specifically, the first chapter of this thesis described the relevance of virtual assistants in eHealth scenarios. Section 1.1 introduced virtual assistants, with special detail on the concepts of software architectures, communication channels, and artificial intelligence. Challenges and critical issues were also presented in this section. In particular, this thesis established and addressed the main research question *"how a new virtual assistant architecture based on microservices can be helpful in home-based caring, with special consideration for chronic diseases"*, identified in Section 1.2. Additionally, this thesis has met all the research objectives outlined in Section 1.2, given as follows.

- The review of the state of the art of the main concepts researched in this thesis is addressed in Chapter 2.
- The design of a new microservice-based virtual assistant architecture is presented in Chapter 3.

- The studies in order to provide a specific joint NLU model for the medication management domain in Spanish are presented in Chapter 4.
- The development of a generic and two specialized prototypes are addressed in Chapter 5.
- The evaluations of the proposed virtual assistant architecture for improving medication adherence and quality of life are addressed in Chapter 6.

The conclusions and objectives achieved in each chapter are presented in Sections 2.5, 3.2, 4.3, 5.4, and 6.3.

7.2 Contributions and conclusions

This thesis aimed to make contributions in the field of eHealth scenarios by aligning technology with remote healthcare. Specifically, this thesis aimed to investigate virtual assistants and their possible contribution to eHealth and mHealth scenarios. The overall objectives proposed in the first chapter have been fulfilled. Indeed, this thesis provides a complete, functional, and validated virtual assistant for healthcare purposes as its major contribution. The alignment with users' daily lives is also achieved using a messaging platform as the communication channel (in this case, Signal). The potential application of the virtual assistant was validated by studying its application for improving medication adherence and patients' quality of life.

Regarding the technical approach, a microservice-based architecture was designed and proposed as architecture for the virtual assistant internal structure. The virtual assistant not only processes the user information but also performs automatic tasks. Microservices allowed for building scalable services for each disease without changing the core virtual assistant. In addition, modularity, standardization, and security have been covered. Furthermore, two types of interactions (menu-based and human-like interaction) were included in the design. This work derived the following published scientific contribution.

Roca, S., Sancho, J., García, J., & Alesanco, Á. (2020). Microservice chatbot architecture for chronic patient support. Journal of biomedical informatics, 102, 103305.

Regarding human-like interaction, and more specifically, Natural Language Understanding, three contributions have been presented in this thesis. First, it has been proven that general corpora yielded better results than specific corpora in most of the embeddings in a Spanish medication domain using joint NLU. Second, it has been proven that sentence length slightly affected sentence-level and intent detection and did not affect slot filling outcomes. In contrast, slot type affected slot filling but did not affect intent detection and sentence-level outcomes. Third, four generated datasets were made available in a publicly accessible repository. More specifically, the datasets generated in this study are openly available at FigShare [154]. This work derived the following published scientific contributions.

Roca, S., Rosset, S., García, J., & Alesanco, Á. (2020, November). Evaluation of Embeddings in Medication Domain for Spanish Language Using Joint Natural Language Understanding. In European Medical and Biological Engineering Conference (pp. 510-517). Springer, Cham.

Roca, S., Rosset, S., García, J., & Alesanco, Á. (2022). A study on the impacts of slot types and training data on Joint Natural Language Understanding in a Spanish medication management assistant scenario. Sensors, 22, 2364.

Furthermore, a generic virtual assistant prototype was developed to obtain a virtual assistant suitable for the healthcare management of different chronic diseases. In addition, two specific prototypes were developed, which provided innovative applications for eHealth scenarios. The first specific prototype focused on medication management, integrating tools and services for type 2 diabetes mellitus and depressive disorder profiles. The second specific prototype aimed to provide support to image-based chronic diseases such as psoriasis with the integration of teleconsultation and image storage. This work derived the following published scientific contribution.

Roca, S., Hernández, M., Sancho, J., García, J., & Alesanco, Á. (2019, September). Virtual assistant prototype for managing medication using messaging platforms. In Mediterranean Conference on Medical and Biological Engineering and Computing (pp. 954-961). Springer, Cham.

Finally, the hypothesis has been evaluated in two clinical scenarios. A study with comorbid type 2 diabetes mellitus and depressive disorder patients and healthcare professionals was conducted for nine months. The work has concluded that the virtual assistant helped on improving patients' medication adherence. Furthermore, it has been shown that the virtual assistant improved psoriatic patients' quality of life. This conclusion was obtained from a one-year study conducted with psoriatic patients and healthcare professionals. This work derived the following published scientific contributions.

Roca, S., Lozano, M. L., García, J., & Alesanco, Á. (2021). Validation of a virtual assistant for improving medication adherence in patients with comorbid type 2 diabetes mellitus and depressive disorder. International Journal of Environmental Research and Public Health, 18, 12056.

Roca, S., Almenara, M., Gilaberte, Y., Gracia-Cazaña, T., Morales Callaghan, A.M., Murciano, D., García, J., & Alesanco, Á. (2022). When virtual assistants meet teledermatology: validation of a virtual assistant to improve the quality of life of psoriatic patients. International Journal of Environmental Research and Public Health, 19, 14527.

7.3 Future work

The objectives proposed in this thesis have been fulfilled and promising results have been observed in the studies performed. Nevertheless, new challenges for future research opportunities have been identified.

- Develop some possible functionalities for other diseases, such as a food tracker with a specific tool that counts the calories in obesity scenarios or a sleep monitor. All these new functionalities could be added following the needs proposed by the physicians of each chronic disease.

- Integrate the virtual assistant architecture storage with the hospital's electronic medical record.

- Add advanced and automated image processing-based advice services for skin care or other image-required diseases with the objective to provide comfort and convenience when using medical services since the tools can be used anywhere. This, in turn, can reduce the number of face-to-face medical consultations.

- Include an extension to the microservice architecture with speech recognition for new user-virtual assistant interaction.

- In Chapter 4, we have investigated the impacts of slot types and training data in Spanish for the medication management domain, but future research can focus on other languages and domains.

- In Chapter 6, we have validated the virtual assistant with comorbid type 2 diabetes mellitus and depressive disorder, and psoriasis. Still, other chronic diseases such as cancer or obesity could use the virtual assistant to improve their conditions. Research works can be conducted to study the improvement achieved by the virtual assistant in such cases.

- Further research can focus on the human-like virtual assistant prototype (presented in Section 5.2). An analysis of the usage of choice-based and NLP-based virtual assistants can be performed as a further comparison between systems.

- Conduct detailed evaluations of long-term effects on healthcare professionals, such as the possible burnout created by the incorporation of virtual assistants into their working time.

Bibliography

[1] Ahmed Fadhil. Beyond patient monitoring: Conversational agents role in telemedicine & healthcare support for home-living elderly individuals. *CoRR*, abs/1803.06000, 2018. URL: http://arxiv.org/abs/1803.06000, arXiv:1803.06000.

[2] Benjamin Chaix, Jean-Emmanuel Bibault, Arthur Pienkowski, Guillaume Delamon, Arthur Guillemassé, Pierre Nectoux, and Benoît Brouard. When chatbots meet patients: One-year prospective study of conversations between patients with breast cancer and a chatbot. *JMIR Cancer*, 5(1):e12856, May 2019. doi:10.2196/12856.

[3] Symptomate. Accessed on 21 January 2019. URL: https://symptomate.com/.

[4] Florence. Accessed on 18 February 2019. URL: https://www.florence.chat/.

[5] Forksy. Accessed on 18 February 2019. URL: https://getforksy.com/.

[6] Adam Palanica, Peter Flaschner, Anirudh Thommandram, Michael Li, and Yan Fossat. Physicians' perceptions of chatbots in health care: Cross-sectional web-based survey. *Journal of Medical Internet Research*, 21(4):e12887, Apr 2019. doi:10.2196/12887.

[7] Emma Wallace, Chris Salisbury, Bruce Guthrie, Cliona Lewis, Tom Fahey, and Susan M. Smith. Managing patients with multimorbidity in primary care. *BMJ*, 350(jan20 2):h176–h176, Jan 2015. doi:10.1136/bmj.h176.

[8] Juanan Pereira and Óscar Díaz. Using health chatbots for behavior change: A mapping study. *Journal of Medical Systems*, 43(5), Apr 2019. doi:10.1007/s10916-019-1237-1.

[9] Carlos Marcos, Arturo González-Ferrer, Mor Peleg, and Carlos Cavero. Solving the interoperability challenge of a distributed complex patient guidance system: a data integrator based on HL7's virtual medical record standard. *Journal of the American Medical Informatics Association*, 22(3):587–599, Apr 2015. doi:10.1093/jamia/ocv003.

[10] HL7 FHIR. Accessed on 18 March 2019. URL: http://www.hl7.org/FHIR/.

[11] Jacek Ruminski, Adam Bujnowski, Tomasz Kocejko, Aliaksei Andrushevich, Martin Biallas, and Rolf Kistler. The data exchange between smart glasses and healthcare information systems using the HL7 FHIR standard. In *2016 9th International Conference on Human System Interactions (HSI)*. IEEE, Jul 2016. doi:10.1109/hsi.2016.7529684.

[12] Number of mobile phone messaging app users worldwide from 2018 to 2025 (in billions). Accessed on 14 September 2022. URL: https://www.statista.com/statistics/483255/number-of-mobile-messaging-users-worldwide/.

[13] Álvaro Alesanco, Jorge Sancho, Yolanda Gilaberte, Edgar Abarca, and Jose García. Bots in messaging platforms, a new paradigm in healthcare delivery: application to custom prescription in dermatology. In *EMBEC & NBC 2017. IFMBE Proceedings*, pages 185–188. Springer Singapore, Jun 2017. doi:10.1007/978-981-10-5122-7_47.

[14] Richard S. Wallace. The elements of aiml style. *Alice AI Foundation*, 139, 2003.

[15] Anda Stoica, Tibor Kadar, Camelia Lemnaru, Rodica Potolea, and Mihaela Dinsoreanu. The impact of data challenges on intent detection and slot filling for the home assistant scenario. In *2019 IEEE 15th International Conference on Intelligent Computer Communication and Processing (ICCP)*. IEEE, Sep 2019. doi:10.1109/iccp48234.2019.8959642.

[16] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013. URL: https://arxiv.org/abs/1301.3781, arXiv:1301.3781.

[17] Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, 5:135–146, Dec 2017. doi:10.1162/tacl_a_00051.

[18] Matthew E. Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep contextualized word representations. *CoRR*, abs/1802.05365, 2018. URL: http://arxiv.org/abs/1802.05365, arXiv:1802.05365.

[19] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018. URL: http://arxiv.org/abs/1810.04805, arXiv:1810.04805.

[20] Kerstin Denecke, Sandra Lutz Hochreutener, Annkathrin Pöpel, and Richard May. Talking to ana: A mobile self-anamnesis application with conversational user interface. In *Proceedings of the 2018 International Conference on Digital Health*. ACM, Apr 2018. doi:10.1145/3194658.3194670.

[21] Lucia Vaira, Mario A. Bochicchio, Matteo Conte, Francesco Margiotta Casaluci, and Antonio Melpignano. MamaBot: a system based on ML and NLP for supporting women and families during pregnancy. In *Proceedings of the 22nd International Database Engineering & Applications Symposium on - IDEAS 2018*, pages 273–277. ACM Press, 2018. doi:10.1145/3216122.3216173.

[22] Telegram homepage. Accessed on 18 March 2019. URL: https://telegram.org/.

[23] Chin-Yuan Huang, Ming-Chin Yang, Chin-Yu Huang, Yu-Jui Chen, Meng-Lin Wu, and Kai-Wen Chen. A chatbot-supported smart wireless interactive healthcare system for weight control and health promotion. In *2018 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*. IEEE, Dec 2018. doi:10.1109/ieem.2018.8607399.

[24] Slack homepage. Accessed on 27 March 2019. URL: https://slack.com.

[25] Mauro Tschanz, Tim Lucas Dorner, Jurgen Holm, and Kerstin Denecke. Using eMMA to manage medication. *Computer*, 51(8):18–25, Aug 2018. doi:10.1109/mc.2018.3191254.

[26] Ahmed Fadhil. A conversational interface to improve medication adherence: Towards AI support in patient's treatment. *CoRR*, abs/1803.09844, 2018. URL: http://arxiv.org/abs/1803.09844, arXiv:1803.09844.

[27] Luca Catarinucci, Danilo de Donno, Luca Mainetti, Luca Palano, Luigi Patrono, Maria Laura Stefanizzi, and Luciano Tarricone. An IoT-aware architecture for smart healthcare systems. *IEEE Internet of Things Journal*, 2(6):515–526, Dec 2015. doi:10.1109/jiot.2015.2417684.

[28] Richard Hill, Dharmendra Shadija, and Mo Rezai. Enabling community health care with microservices. In *2017 IEEE International Symposium on Parallel and Distributed Processing with Applications and 2017 IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC)*. IEEE, Dec 2017. doi:10.1109/ispa/iucc.2017.00220.

[29] Orla O'Brien and Ruairi Donagh O'Reilly. Beats-per-minute (BPM): A microservice-based platform for the monitoring of health related data via activity trackers. In *2018 IEEE 20th International Conference on e-Health Networking, Applications and Services (Healthcom)*. IEEE, Sep 2018. doi:10.1109/healthcom.2018.8531169.

[30] Sajjad Ali, Muhammad Golam Kibria, Muhammad Aslam Jarwar, Sunil Kumar, and Ilyoung Chong. Microservices model in WoO based IoT platform for depressive disorder assistance. In *2017 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, Oct 2017. doi:10.1109/ictc.2017.8190800.

[31] Your.MD. Accessed on 25 September 2018. URL: http://www.your.md.

[32] Ada. Accessed on 21 January 2019. URL: https://ada.com/.

[33] Mediktor. Accessed on 21 January 2019. URL: https://www.mediktor.com/en.

[34] HealthTap. Accessed on 21 January 2019. URL: https://www.healthtap.com/.

[35] Ahmed Fadhil and Silvia Gabrielli. Addressing challenges in promoting healthy lifestyles: the al-chatbot approach. In *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. ACM, May 2017. doi:10.1145/3154862.3154914.

[36] HealthBot. Accessed on 25 September 2018. URL: https://healthbot.in.

[37] Agnese Augello, Mario Scriminaci, Salvatore Gaglio, and Giovanni Pilato. A modular framework for versatile conversational agent building. In *2011 International*

Conference on Complex, Intelligent, and Software Intensive Systems. IEEE, Jun 2011. doi:10.1109/cisis.2011.95.

[38] Mengting Yan, Paul Castro, Perry Cheng, and Vatche Ishakian. Building a chatbot with serverless computing. In *MOTA '16: Proceedings of the 1st International Workshop on Mashups of Things and APIs*. ACM, Dec 2016. doi:10.1145/3007203.3007217.

[39] A M Rahman, Abdullah Al Mamun, and Alma Islam. Programming challenges of chatbot: Current and future prospective. In *2017 IEEE Region 10 Humanitarian Technology Conference (R10-HTC)*. IEEE, Dec 2017. doi:10.1109/r10-htc.2017.8288910.

[40] Yanshan Wang, Sijia Liu, Naveed Afzal, Majid Rastegar-Mojarad, Liwei Wang, Feichen Shen, Paul Kingsbury, and Hongfang Liu. A comparison of word embeddings for the biomedical natural language processing. *Journal of Biomedical Informatics*, 87:12–20, Nov 2018. doi:10.1016/j.jbi.2018.09.008.

[41] Antoine Neuraz, Vincent Looten, Bastien Rance, Nicolas Daniel, Nicolas Garcelon, Leonardo Campillos Llanos, Anita Burgun, and Sophie Rosset. Do you need embeddings trained on a massive specialized corpus for your clinical natural language processing task? *Studies in health technology and informatics*, 264:1558–1559, 2019. doi:10.3233/SHTI190533.

[42] Sahar Ghannay, Antoine Neuraz, and Sophie Rosset. What is best for spoken language understanding: small but task-dependant embeddings or huge but out-of-domain embeddings? In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8114–8118. IEEE, May 2020. doi:10.1109/icassp40776.2020.9053278.

[43] Madhumita Sushil, Simon Suster, and Walter Daelemans. Are we there yet? exploring clinical domain knowledge of BERT models. In *Proceedings of the 20th Workshop on Biomedical Language Processing*, pages 41–53. Association for Computational Linguistics, Jun 2021. doi:10.18653/v1/2021.bionlp-1.5.

[44] Isabel Segura-Bedmar and Paloma Martínez. Simplifying drug package leaflets written in Spanish by using word embedding. *Journal of Biomedical Semantics*, 8(1), Sep 2017. doi:10.1186/s13326-017-0156-7.

[45] Felipe Soares, Marta Villegas, Aitor Gonzalez-Agirre, Martin Krallinger, and Jordi Armengol-Estapé. Medical word embeddings for Spanish: Development and evaluation. In *Proceedings of the 2nd Clinical Natural Language Processing Workshop*. Association for Computational Linguistics, 2019. doi:10.18653/v1/w19-1916.

[46] Yu Wang, Abhishek Patel, Yilin Shen, and Hongxia Jin. A deep reinforcement learning based multimodal coaching model (DCM) for slot filling in spoken language understanding(SLU). In *Interspeech 2018*, pages 3444–3448. ISCA, Sep 2018. doi:10.21437/interspeech.2018-1379.

[47] Yuka Kobayashi, Takami Yoshida, Kenji Iwata, and Hiroshi Fujimura. Slot filling with weighted multi-encoders for out-of-domain values. In *Interspeech 2019*, pages 854–858. ISCA, Sep 2019. doi:10.21437/interspeech.2019-1226.

[48] Heike Adel and Hinrich Schütze. Impact of coreference resolution on slot filling. *CoRR*, abs/1710.09753, 2017. URL: http://arxiv.org/abs/1710.09753, arXiv:1710.09753.

[49] Joo-Kyung Kim, Gokhan Tur, Asli Celikyilmaz, Bin Cao, and Ye-Yi Wang. Intent detection using semantically enriched word embeddings. In *2016 IEEE Spoken Language Technology Workshop (SLT)*. IEEE, Dec 2016. doi:10.1109/slt.2016.7846297.

[50] Wei Wang, Saghar Hosseini, Ahmed Hassan Awadallah, Paul N. Bennett, and Chris Quirk. Context-aware intent identification in email conversations. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, Jul 2019. doi:10.1145/3331184.3331260.

[51] Chih-Wen Goo, Guang Gao, Yun-Kai Hsu, Chih-Li Huo, Tsung-Chieh Chen, Keng-Wei Hsu, and Yun-Nung Chen. Slot-gated modeling for joint slot filling and intent prediction. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 753–757. Association for Computational Linguistics, Jun 2018. doi:10.18653/v1/n18-2118.

[52] Bing Liu and Ian R. Lane. Attention-based recurrent neural network models for joint intent detection and slot filling. *CoRR*, abs/1609.01454, 2016. URL: http://arxiv.org/abs/1609.01454, arXiv:1609.01454.

[53] Congrui Wang, Zhen Huang, and Minghao Hu. SASGBC: Improving sequence labeling performance for joint learning of slot filling and intent detection. In *Proceedings of 2020 the 6th International Conference on Computing and Data Engineering*, pages 29–33. ACM, Jan 2020. doi:10.1145/3379247.3379266.

[54] Mladen Dimovski, Claudiu Musat, Vladimir Ilievski, Andreea Hossmann, and Michael Baeriswyl. Submodularity-inspired data selection for goal-oriented chatbot training based on sentence embeddings. *CoRR*, abs/1802.00757, 2018. URL: http://arxiv.org/abs/1802.00757, arXiv:1802.00757.

[55] International diabetes federation: Diabetes atlas. 9th edition. https://diabetesatlas.org/en/resources/. Accessed on 17 August 2020.

[56] Julián Rodríguez-Almagro, Álvaro García-Manzanares, Alfredo J. Lucendo, and Antonio Hernández-Martínez. Health-related quality of life in diabetes mellitus and its social, demographic and clinical determinants: A nationwide cross-sectional survey. *Journal of Clinical Nursing*, 27(21-22):4212–4223, Jul 2018. doi:10.1111/jocn.14624.

[57] Andres Pichon-Riviere, Vilma Irazola, Andrea Beratarrechea, Andrea Alcaraz, and Carolina Carrara. Quality of life in type 2 diabetes mellitus patients requiring insulin treatment in buenos aires, argentina: a cross-sectional study. *International Journal of Health Policy and Management*, 4(7):475–480, Apr 2015. doi:10.15171/ijhpm.2015.80.

[58] SV Bădescu, C Tătaru, L Kobylinska, EL Georgescu, DM Zahiu, AM Zăgrean, and L Zăgrean. The association between diabetes mellitus and depression. *Journal of medicine and life*, 9(2):120–125, 2016. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4863499/.

[59] Calum D Moulton, John C Pickup, and Khalida Ismail. The link between depression and diabetes: the search for shared mechanisms. *The Lancet Diabetes & Endocrinology*, 3(6):461–471, Jun 2015. doi:10.1016/s2213-8587(15)00134-5.

[60] Abdallah Alzoubi, Rnad Abunaser, Adi Khassawneh, Mahmoud Alfaqih, Aws Khasawneh, and Nour Abdo. The bidirectional relationship between diabetes and depression: A literature review. *Korean Journal of Family Medicine*, 39(3):137–146, May 2018. doi:10.4082/kjfm.2018.39.3.137.

[61] Spencer L James, Degu Abate, Kalkidan Hassen Abate, Solomon M Abay, Cristiana Abbafati, Nooshin Abbasi, Hedayat Abbastabar, Foad Abd-Allah, Jemal Abdela, Ahmed Abdelalim, et al. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017. *The Lancet*, 392(10159):1789–1858, Nov 2018. doi:10.1016/s0140-6736(18)32279-7.

[62] Jeffrey L. Birk, Ian M. Kronish, Nathalie Moise, Louise Falzon, Sunmoo Yoon, and Karina W. Davidson. Depression and multimorbidity: Considering temporal characteristics of the associations between depression and multiple chronic diseases. *Health Psychology*, 38(9):802–811, Sep 2019. doi:10.1037/hea0000737.

[63] Christoph Jüngst, Stefan Gräber, Sven Simons, Heiner Wedemeyer, and Frank Lammert. Medication adherence among patients with chronic diseases: a survey-based study in pharmacies. *QJM: An International Journal of Medicine*, 112(7):505–512, Mar 2019. doi:10.1093/qjmed/hcz058.

[64] Stephanie Greer, Danielle Ramo, Yin-Juei Chang, Michael Fu, Judith Moskowitz, and Jana Haritatos. Use of the chatbot "Vivibot" to deliver positive psychology skills and promote well-being among young people after cancer treatment: Randomized controlled feasibility trial. *JMIR mHealth and uHealth*, 7(10):e15018, oct 2019. doi:10.2196/15018.

[65] Sandra Hauser-Ulrich, Hansjörg Künzli, Danielle Meier-Peterhans, and Tobias Kowatsch. A smartphone-based health care chatbot to promote self-management of chronic pain (SELMA): Pilot randomized controlled trial. *JMIR mHealth and uHealth*, 8(4):e15806, Apr 2020. doi:10.2196/15806.

[66] Clara K. Chow, Julie Redfern, Graham S. Hillis, Jay Thakkar, Karla Santo, Maree L. Hackett, Stephen Jan, Nicholas Graves, Laura de Keizer, Tony Barry, Severine Bompoint, Sandrine Stepien, Robyn Whittaker, Anthony Rodgers, and Aravinda Thiagalingam. Effect of lifestyle-focused text messaging on risk factor modification in patients with coronary heart disease: a randomized clinical trial. *JAMA*, 314(12):1255–1263, Sep 2015. doi:10.1001/jama.2015.10945.

[67] Benjamin Chaix, Jean-Emmanuel Bibault, Arthur Pienkowski, Guillaume Delamon, Arthur Guillemassé, Pierre Nectoux, and Benoît Brouard. When chatbots meet patients: One-year prospective study of conversations between patients with breast cancer and a chatbot. *JMIR Cancer*, 5(1):e12856, May 2019. doi:10.2196/12856.

[68] Jacques Quilici, Lionel Fugon, Shirley Beguin, Pierre Emmanuel Morange, Jean-Louis Bonnet, Marie-Christine Alessi, Patrizia Carrieri, and Thomas Cuisset. Effect of motivational mobile phone short message service on aspirin adherence after coronary stenting for acute coronary syndrome. *International Journal of Cardiology*, 168(1):568–569, Sep 2013. doi:10.1016/j.ijcard.2013.01.252.

[69] Linda G. Park, Jill Howie-Esquivel, and Kathleen Dracup. A quantitative systematic review of the efficacy of mobile phone interventions to improve medication adherence. *Journal of Advanced Nursing*, 70(9):1932–1953, Apr 2014. doi:10.1111/jan.12400.

[70] Jay Thakkar, Rahul Kurup, Tracey-Lea Laba, Karla Santo, Aravinda Thiagalingam, Anthony Rodgers, Mark Woodward, Julie Redfern, and Clara K. Chow. Mobile telephone text messaging for medication adherence in chronic disease. *JAMA Internal Medicine*, 176(3):340–349, Mar 2016. doi:10.1001/jamainternmed.2015.7667.

[71] Mingxuan Cui, Xueyan Wu, Jiangfeng Mao, Xi Wang, and Min Nie. T2DM self-management via smartphone applications: A systematic review and meta-analysis. *PLOS ONE*, 11(11):e0166718, Nov 2016. doi:10.1371/journal.pone.0166718.

[72] Kathleen Kara Fitzpatrick, Alison Darcy, and Molly Vierhile. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using

a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, 4(2):e19, Jun 2017. doi:10.2196/mental.7785.

[73] Russell Fulmer, Angela Joerin, Breanna Gentile, Lysanne Lakerink, and Michiel Rauws. Using psychological artificial intelligence (tess) to relieve symptoms of depression and anxiety: Randomized controlled trial. *JMIR Mental Health*, 5(4):e64, Dec 2018. doi:10.2196/mental.9782.

[74] Archana Sarda, Suresh Munuswamy, Shubhankar Sarda, and Vinod Subramanian. Using passive smartphone sensing for improved risk stratification of patients with depression and diabetes: Cross-sectional observational study. *JMIR mHealth and uHealth*, 7(1):e11041, Jan 2019. doi:10.2196/11041.

[75] Hillary R Bogner, Knashawn H Morales, Heather F de Vries, and Anne R Cappola. Integrated management of type 2 diabetes mellitus and depression treatment to improve medication adherence: A randomized controlled trial. *The Annals of Family Medicine*, 10(1):15–22, Jan 2012. doi:10.1370/afm.1344.

[76] National health council: National psoriasis foundation. Accessed on 11 May 2022. URL: https://nationalhealthcouncil.org/member/national-psoriasis-foundation/.

[77] National psoriasis foundation: Psoriasis statistics. Accessed on 11 May 2022. URL: https://www.psoriasis.org/psoriasis-statistics/.

[78] Psoriasis association: About psoriasis. Accessed on 11 May 2022. URL: https://www.psoriasis-association.org.uk/about-psoriasis.

[79] Robert S. Stern, Tamar Nijsten, Steven R. Feldman, David J. Margolis, and Tara Rolstad. Psoriasis is common, carries a substantial burden even when not extensive, and is associated with widespread treatment dissatisfaction. *Journal of Investigative Dermatology Symposium Proceedings*, 9(2):136–139, Mar 2004. doi:10.1046/j.1087-0024.2003.09102.x.

[80] Charles G. Helmick, Hyewon Lee-Han, Shawn C. Hirsch, Tiffany L. Baird, and Christopher L. Bartlett. Prevalence of psoriasis among adults in the US: 2003–2006 and 2009–2010 national health and nutrition examination surveys. *American Journal of Preventive Medicine*, 47(1):37–45, Jul 2014. doi:10.1016/j.amepre.2014.02.012.

[81] Joel M. Gelfand, Steven R. Feldman, Robert S. Stern, John Thomas, Tara Rolstad, and David J. Margolis. Determinants of quality of life in patients with psoriasis: A study from the US population. *Journal of the American Academy of Dermatology*, 51(5):704–708, Nov 2004. doi:10.1016/j.jaad.2004.04.014.

[82] Alison Havelin and Philip Hampton. Telemedicine and e-Health in the management of psoriasis: Improving patient outcomes – a narrative review. *Psoriasis: Targets and Therapy*, 12:15–24, Mar 2022. doi:10.2147/ptt.s323471.

[83] Jacob Beer, Edward Hadeler, Alejo Calume, Howard Gitlow, and Keyvan Nouri. Teledermatology: current indications and considerations for future use. *Archives of Dermatological Research*, 313(1):11–15, Oct 2020. doi:10.1007/s00403-020-02145-3.

[84] SymTrac™ psoriasis. Accessed on 1 June 2022. URL: https://play.google.com/store/apps/details?id=com.novartis.pharma.ukie.symtracpso.ext.

[85] AI psoriasis app: Manage and care. Accessed on 1 June 2022. URL: https://play.google.com/store/apps/details?id=com.beat.psoriasis.

[86] Imagine skin condition tracker. Accessed on 1 June 2022. URL: https://play.google.com/store/apps/details?id=com.leo_pharma.imagine.

[87] Kopa. Accessed on 1 June 2022. URL: https://kopa.com/.

[88] Claro. Accessed on 1 June 2022. URL: https://www.clarobyalmirall.com/.

[89] MiPsoriasis. Accessed on 1 June 2022. URL: https://play.google.com/store/apps/details?id=com.novartis.psocare&gl=US.

[90] Monique A Sager, Aditya M Kashyap, Mila Tamminga, Sadhana Ravoori, Christopher Callison-Burch, and Jules B Lipoff. Identifying and responding to health misinformation on Reddit dermatology forums with artificially intelligent bots using natural language processing: Design and evaluation study. *JMIR Dermatology*, 4(2):e20975, Sep 2021. doi:10.2196/20975.

[91] Laura Catherine Armitage, Aikaterini Kassavou, and Stephen Sutton. Do mobile device apps designed to support medication adherence demonstrate efficacy? a systematic review of randomised controlled trials, with meta-analysis. *BMJ Open*, 10(1):e032045, Jan 2020. doi:10.1136/bmjopen-2019-032045.

[92] M.T. Svendsen, F. Andersen, K.H. Andersen, A. Pottegård, H. Johannessen, S. Möller, B. August, S.R. Feldman, and K.E. Andersen. A smartphone application supporting patients with psoriasis improves adherence to topical treatment: a randomized controlled trial. *British Journal of Dermatology*, 179(5):1062–1071, jul 2018. doi:10.1111/bjd.16667.

[93] Lena Domogalla, Alena Beck, Theresa Schulze-Hagen, Raphael Herr, Johannes Benecke, and Astrid Schmieder. Impact of an eHealth smartphone app on the mental health of patients with psoriasis: Prospective randomized controlled intervention study. *JMIR mHealth and uHealth*, 9(10):e28149, Oct 2021. doi:10.2196/28149.

[94] April W. Armstrong, Cindy J. Chambers, Emanual Maverakis, Michelle Y. Cheng, Cory A. Dunnick, Mary-Margaret Chren, Joel M. Gelfand, David J. Wong, Brittany M. Gibbons, Caitlin M. Gibbons, et al. Effectiveness of online vs in-person care for adults with psoriasis: a randomized clinical trial. *JAMA Network Open*, 1(6):e183062–e183062, Oct 2018. doi:10.1001/jamanetworkopen.2018.3062.

[95] April W. Armstrong, Adam R. Ford, Cindy J. Chambers, Emanual Maverakis, Cory A. Dunnick, Mary-Margaret Chren, Joel M. Gelfand, Caitlin M. Gibbons, Brittany M. Gibbons, and Christianne J. Lane. Online care versus in-person care for improving quality of life in psoriasis: A randomized controlled equivalency trial. *Journal of Investigative Dermatology*, 139(5):1037–1044, May 2019. doi:10.1016/j.jid.2018.09.039.

[96] Signal homepage. Accessed on 18 March 2019. URL: https://signal.org/.

[97] Md. Shahriare Satu, Md. Hasnat Parvez, and Shamim-Al-Mamun. Review of integrated applications with AIML based chatbot. In *2015 International Conference on Computer and Information Engineering (ICCIE)*. IEEE, Nov 2015. doi:10.1109/ccie.2015.7399324.

[98] Chris Richardson and Floyd Earl Smith. Microservices - From Design to Deployment. *NGINX Inc*, 2016.

[99] PsoriasisCalc. Accessed on 30 January 2019. URL: https://play.google.com/store/apps/details?id=org.dipler.psoriasiscalc.android.

[100] Diabetes diagnostics. Accessed on 30 January 2019. URL: https://play.google.com/store/apps/details?id=uk.ac.exeter.DiabetesDiagnostics.

[101] Haolin Wang, Qingpeng Zhang, Mary Ip, and Joseph Tak Fai Lau. Social media–based conversational agents for health management and interventions. *Computer*, 51(8):26–33, Aug 2018. doi:10.1109/mc.2018.3191249.

[102] Paula M. Gardiner, Kelly D. McCue, Lily M. Negash, Teresa Cheng, Laura F. White, Leanne Yinusa-Nyahkoon, Brian W. Jack, and Timothy W. Bickmore. Engaging women with an embodied conversational agent to deliver mindfulness and lifestyle recommendations: A feasibility randomized control trial. *Patient Education and Counseling*, 100(9):1720–1729, sep 2017. doi:10.1016/j.pec.2017.04.015.

[103] multiLSTM. Accessed on 27 January 2020. URL: https://github.com/SNUDerek/multiLSTM.

[104] Hugo Boulanger. Évaluation systématique d'une méthode commune de génération. In *Proceedings of the 6e conférence conjointe Journées d'Études sur la Parole (JEP, 33e édition), Traitement Automatique des Langues Naturelles (TALN, 27e édition), Rencontre des Étudiants Chercheurs en Informatique pour le Traitement Automatique des Langues (RÉCITAL, 22e édition). Volume 3: Rencontre des Étudiants Chercheurs en Informatique pour le TAL*, pages 43–56. ATALA; AFCP, Nancy, France, June 2020. Accessed on 15 March 2022. URL: https://hal.archives-ouvertes.fr/hal-02786185/.

[105] CIMA - centro de inforamción de medicamentos. Accessed on 14 January 2020. URL: https://cima.aemps.es/cima/publico/home.html.

[106] Nikita Kitaev, Steven Cao, and Dan Klein. Multilingual constituency parsing with self-attention and pre-training. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, 2019. doi:10.18653/v1/p19-1340.

[107] Yifan Peng, Shankai Yan, and Zhiyong Lu. Transfer learning in biomedical natural language processing: An evaluation of BERT and ELMo on ten benchmarking datasets. In *Proceedings of the 18th BioNLP Workshop and Shared Task*. Association for Computational Linguistics, 2019. doi:10.18653/v1/w19-5006.

[108] Yuqi Si, Jingqi Wang, Hua Xu, and Kirk Roberts. Enhancing clinical concept extraction with contextual embeddings. *Journal of the American Medical Informatics Association*, 26(11):1297–1304, Jul 2019. doi:10.1093/jamia/ocz096.

[109] Jennifer Hill, W. Randolph Ford, and Ingrid G. Farreras. Real conversations with artificial intelligence: A comparison between human–human online conversations and human–chatbot conversations. *Computers in Human Behavior*, 49:245–250, Aug 2015. doi:10.1016/j.chb.2015.02.026.

[110] Di Wu, Liang Ding, Fan Lu, and Jian Xie. SlotRefine: A fast non-autoregressive model for joint intent detection and slot filling. *CoRR*, abs/2010.02693, 2020. URL: https://arxiv.org/abs/2010.02693, arXiv:2010.02693.

[111] Libo Qin, Wanxiang Che, Yangming Li, Haoyang Wen, and Ting Liu. A stack-propagation framework with token-level intent detection for spoken language understanding. *CoRR*, abs/1909.02188, 2019. URL: http://arxiv.org/abs/1909.02188, arXiv:1909.02188.

[112] Haihong E, Peiqing Niu, Zhongfu Chen, and Meina Song. A novel bi-directional interrelated model for joint intent detection and slot filling. *CoRR*, abs/1907.00390, 2019. URL: http://arxiv.org/abs/1907.00390, arXiv:1907.00390.

[113] Chenwei Zhang, Yaliang Li, Nan Du, Wei Fan, and Philip S. Yu. Joint slot filling and intent detection via capsule neural networks. *CoRR*, abs/1812.09471, 2018. URL: http://arxiv.org/abs/1812.09471, arXiv:1812.09471.

[114] Bing Liu and Ian R. Lane. Joint online spoken language understanding and language modeling with recurrent neural networks. *CoRR*, abs/1609.01462, 2016. URL: http://arxiv.org/abs/1609.01462, arXiv:1609.01462.

[115] R: The R project for statistical computing. https://www.r-project.org/. Accessed on 15 December 2021.

[116] N. Lasierra, A. Alesanco, S. Guillén, and J. García. A three stage ontology-driven solution to provide personalized care to chronic patients at home. *Journal of Biomedical Informatics*, 46(3):516–529, Jun 2013. doi:10.1016/j.jbi.2013.03.006.

[117] A facade to make using the libsignal-service easy. Accessed on 11 January 2019. URL: https://github.com/Turakar/signal4j.

[118] A java library for communicating via Signal. Accessed on 11 January 2019. URL: https://github.com/Turasa/libsignal-service-java.

[119] Signal oficial repository. Accessed on 11 January 2019. URL: https://github.com/signalapp.

[120] A bot for signal. Accessed on 11 January 2019. URL: https://github.com/nerdclub-tfg/signal-bot.

[121] Program AB. ALICE A.I. foundation. Accessed on 17 March 2019. URL: https://code.google.com/archive/p/program-ab/.

[122] Kibana. Accessed on 18 February 2019. URL: https://www.elastic.co/products/kibana.

[123] Constitutional law 3/2018 of the 5th december concerning protection of personal data and guarantee of digital rights. Accessed on 20 March 2019. URL: https://www.boe.es/buscar/doc.php?id=BOE-A-2018-16673.

[124] Regulation (EU) 2016/679 of the european parliament and of the council of 27 April 2016. Accessed on 20 March 2019. URL: https://eur-lex.europa.eu/eli/reg/2016/679/oj.

[125] Eneko Agirre, Arantxa Otegi, Camille Pradel, Sophie Rosset, Anselmo Penas, and Mark Cieliebak. LIHLITH: learning to interact with humans by lifelong interaction with humans. *Procesamiento del Lenguaje Natural*, 63:147–150, 2019. URL: http://journal.sepln.org/sepln/ojs/ojs/index.php/pln/article/view/6106.

[126] N. M. Vink, O. H. Klungel, R. P. Stolk, and P. Denig. Comparison of various measures for assessing medication refill adherence using prescription data. *Pharmacoepidemiology and Drug Safety*, 18(2):159–165, Dec 2008. doi:10.1002/pds.1698.

[127] Ines Krass, P Schieback, and Teerapon Dhippayom. Adherence to diabetes medication: a systematic review. *Diabetic Medicine*, 32(6):725–737, Jan 2015. doi:10.1111/dme.12651.

[128] Kurt Kroenke, Robert L. Spitzer, and Janet B. W. Williams. The PHQ-9: validity of a brief depression severity measure. *Journal of General Internal Medicine*, 16(9):606–613, Sep 2001. doi:10.1046/j.1525-1497.2001.016009606.x.

[129] Comité de ética de la investigación de la comunidad autónoma de aragón: CEICA. http://www.iacs.es/investigacion/comite-de-etica-de-la-investigacion-de-aragon-ceica/. Accessed on 3 April 2020.

[130] María I. Rodríguez-Idígoras, Jesús Sepúlveda-Muñoz, Ramón Sánchez-Garrido-Escudero, José L. Martínez-González, José L. Escolar-Castelló, Isabel M. Paniagua-Gómez, Rosa Bernal-López, María V. Fuentes-Simón, and Daniel Garófano-Serrano. Telemedicine influence on the follow-up of type 2 diabetes patients. *Diabetes Technology & Therapeutics*, 11(7):431–437, Jul 2009. doi:10.1089/dia.2008.0114.

[131] Anna-Leena Orsama, Jaakko Lähteenmäki, Kari Harno, Minna Kulju, Eva Wintergerst, Holly Schachner, Pat Stenger, Juha Leppänen, Hannu Kaijanranta, Ville Salaspuro, and William A. Fisher. Active assistance technology reduces glycosylated hemoglobin and weight in individuals with type 2 diabetes: Results of a theory-based randomized trial. *Diabetes Technology & Therapeutics*, 15(8):662–669, Aug 2013. doi:10.1089/dia.2013.0056.

[132] David C Mohr, Kathryn Noth Tomasino, Emily G Lattie, Hannah L Palac, Mary J Kwasny, Kenneth Weingardt, Chris J Karr, Susan M Kaiser, Rebecca C Rossom, Leland R Bardsley, Lauren Caccamo, Colleen Stiles-Shields, and Stephen M Schueller. IntelliCare: An eclectic, skills-based app suite for the treatment of depression and anxiety. *Journal of Medical Internet Research*, 19(1):e10, Jan 2017. doi:10.2196/jmir.6645.

[133] Maggie McCue, Christopher Blair, Ben Fehnert, James King, Francesca Cormack, Sara Sarkey, Anna Eramo, Christopher Kabir, Rasha Khatib, and David Kemp. Mobile app to enhance patient activation and patient-provider communication in major depressive disorder management: Collaborative, randomized controlled pilot study. *JMIR Formative Research*, 6(10):e34923, Oct 2022. doi:10.2196/34923.

[134] Kam Capoccia, Peggy S. Odegard, and Nancy Letassy. Medication adherence with diabetes medication: a systematic review of the literature. *The Diabetes Educator*, 42(1):34–71, Dec 2015. doi:10.1177/0145721715619038.

[135] World health organization. medication errors. (2016). https://apps.who.int/iris/bitstream/handle/10665/252274/9789241511643-eng.pdf. Accessed on 26 November 2020.

[136] Tom Nadarzynski, Oliver Miles, Aimee Cowie, and Damien Ridge. Acceptability of artificial intelligence (AI)-led chatbot services in healthcare: A mixed-methods study. *DIGITAL HEALTH*, 5:205520761987180, Jan 2019. doi:10.1177/2055207619871808.

[137] eHealthZ YouTube channel. https://www.youtube.com/channel/UCYSyheAygDwqG7Cxoe7nvTw/videos. Accessed on 21 April 2020.

[138] M.K.A. Basra, R. Fenech, R.M. Gatt, M.S. Salek, and A.Y. Finlay. The dermatology life quality index 1994-2007: a comprehensive review of validation data and clinical results. *British Journal of Dermatology*, 159(5):997–1035, Sep 2008. doi:10.1111/j.1365-2133.2008.08832.x.

[139] JL Sánchez-Carazo, E Daudén, E Herrera, J Toribio, L Puig, FJ Sabater, N Perulero, and X Badia. Development of a new measure for assessing health-related quality of life in patients with psoriasis: PSO-LIFE. *Value in Health*, 12(7):A458, Oct 2009. doi:10.1016/s1098-3015(10)75270-0.

[140] Esteban Dauden, Enrique Herrera, Lluis Puig, José Luis Sánchez-Carazo, Jaime Toribio, Ma Teresa Caloto, Gonzalo Nocea, Montse Roset, and Nuria Lara. Validation of a new tool to assess health-related quality of life in psoriasis: the PSO-LIFE questionnaire. *Health and Quality of Life Outcomes*, 10(1):1–10, May 2012. doi:10.1186/1477-7525-10-56.

[141] Mark J Atkinson, Anusha Sinha, Steven L Hass, Shoshana S Colman, Ritesh N Kumar, Meryl Brod, and Clayton R Rowland. Validation of a general measure of treatment satisfaction, the treatment satisfaction questionnaire for medication (TSQM), using a national panel study of chronic disease. *Health and Quality of Life Outcomes*, 2(1):1–13, 2004. doi:10.1186/1477-7525-2-12.

[142] Donald E. Morisky, Alfonso Ang, Marie Krousel-Wood, and Harry J. Ward. Predictive validity of a medication adherence measure in an outpatient setting. *The Journal of Clinical Hypertension*, 10(5):348–354, May 2008. doi:10.1111/j.1751-7176.2008.07572.x.

[143] Magdalena Del Rocio Sevilla-Gonzalez, Lizbeth Moreno Loaeza, Laura Sofia Lazaro-Carrera, Brigette Bourguet Ramirez, Anabel Vázquez Rodríguez, María Luisa Peralta-Pedrero, and Paloma Almeda-Valdes. Spanish version of the system usability scale for the assessment of electronic tools: Development and validation. *JMIR Human Factors*, 7(4):e21161, Dec 2020. doi:10.2196/21161.

[144] Simone Borsci, Stefano Federici, and Marco Lauriola. On the dimensionality of the system usability scale: a test of alternative measurement models. *Cognitive Processing*, 10(3):193–197, Jun 2009. doi:10.1007/s10339-009-0268-9.

[145] Leming Zhou, Jie Bao, I Made Agus Setiawan, Andi Saptono, and Bambang Parmanto. The mHealth app usability questionnaire (MAUQ): Development and validation study. *JMIR mHealth and uHealth*, 7(4):e11500, Apr 2019. doi:10.2196/11500.

[146] Thyra de Jongh, Ipek Gurol-Urganci, Vlasta Vodopivec-Jamsek, Josip Car, and Rifat Atun. Mobile phone messaging for facilitating self-management of long-term illnesses. *Cochrane Database of Systematic Reviews*, Dec 2012. doi:10.1002/14651858.cd007459.pub2.

[147] Heather Kornmehl, Sanminder Singh, Mary Ann Johnson, and April W. Armstrong. Direct-access online care for the management of atopic dermatitis: A randomized clinical trial examining patient quality of life. *Telemedicine and e-Health*, 23(9):726–732, Sep 2017. doi:10.1089/tmj.2016.0249.

[148] Hyekyun Rhee, James Allen, Jennifer Mammen, and Mary Swift. Mobile phone-based asthma self-management aid for adolescents (mASMAA): a feasibility study. *Patient Preference and Adherence*, page 63, Jan 2014. doi:10.2147/ppa.s53504.

[149] Tzu Tsun Luk, Judy Hiu Tung Lui, and Man Ping Wang. Efficacy, usability, and acceptability of a chatbot for promoting COVID-19 vaccination in unvaccinated or booster-hesitant young adults: Pre-post pilot study. *Journal of Medical Internet Research*, 24(10):e39063, Oct 2022. doi:10.2196/39063.

[150] Jooyoung Oh, Sooah Jang, Hyunji Kim, and Jae-Jin Kim. Efficacy of mobile app-based interactive cognitive behavioral therapy using a chatbot for panic disorder. *International Journal of Medical Informatics*, 140:104171, Aug 2020. doi:10.1016/j.ijmedinf.2020.104171.

[151] Johannes Knitza, Felix Muehlensiepen, Yuriy Ignatyev, Franziska Fuchs, Jacob Mohn, David Simon, Arnd Kleyer, Filippo Fagni, Sebastian Boeltz, Harriet Morf, et al. Patient's perception of digital symptom assessment technologies in rheumatology: Results from a multicentre study. *Frontiers in public health*, 10:844669, Feb 2022. doi:10.3389/fpubh.2022.844669.

[152] Jeffrey A. Corbett, Janice M. Opladen, and John D. Bisognano. Telemedicine can revolutionize the treatment of chronic disease. *International Journal of Cardiology Hypertension*, 7:100051, Dec 2020. doi:10.1016/j.ijchy.2020.100051.

[153] Pooja Shah, Deborah N Dorrell, Steven R Feldman, and William W Huang. The impact of the coronavirus disease 2019 pandemic on dermatologist burnout: a survey study. *Dermatology Online Journal*, 27(6), Jul 2021. doi:10.5070/d327654050.

[154] Datasets in Spanish for short/long slots and short/long sentences. Posted on 15 March 2022. doi:10.6084/m9.figshare.19361825.

[155] M. Rijsbergen, T. Niemeyer van der Kolk, R. Rijneveld, J.H.F.M. Pinckaers, I. Meshcheriakov, J.N. Bouwes Bavinck, M.B.A. van Doorn, G. Hogendoorn, G. Feiss, A.F. Cohen, J. Burggraaf, M.I.E. van Poelgeest, and R. Rissmann. Mobile e-diary application facilitates the monitoring of patient-reported outcomes and a high treatment adherence for clinical trials in dermatology. *Journal of the European Academy of Dermatology and Venereology*, 34(3):633–639, Sep 2019. doi:10.1111/jdv.15872.

[156] A.E. Arafa, F. Anzengruber, A.M. Mostafa, and A.A. Navarini. Perspectives of online surveys in dermatology. *Journal of the European Academy of Dermatology and Venereology*, 33(3):511–520, dec 2018. doi:10.1111/jdv.15283.

[157] Mayra B.C. Maymone, Samantha Venkatesh, Eric Secemsky, Kavitha Reddy, and Neelam A. Vashi. Research techniques made simple: Web-based survey research

in dermatology: Conduct and applications. *Journal of Investigative Dermatology*, 138(7):1456–1462, Jul 2018. doi:10.1016/j.jid.2018.02.032.

Appendix A

Additional Prototypes

This appendix presents and discusses a multi-platform system architecture intending to conveniently collect and manage health information through questionnaires in chronic scenarios using a website and a messaging platform as user interfaces. The system is built to create simple and flexible tools and services to collect and manage medical data for more effective disease management. The system also offers help to its users in medication management by providing all the information and reminders about the patients' medication. An additional access control functionality is also included in the system architecture using JWT authentication and public keys to ensure the highest level of security and privacy.

Additionally, this appendix presents a multilingual microservice prototype. This prototype enables the use of different languages based on the phone number's prefix.

A.1 Multi-platform system prototype

The importance of retrieving information in healthcare scenarios, and more specifically in those in which the questionnaires reflect the disease evolution, is a key factor in the correct management of the patients. The questionnaires used in diseases such as psoriasis or diabetes, include specific questions related to the gravity of the disease, the quality of life, and the medication adherence. When comorbidities come into the picture, the quantity of questionnaires keeps increasing. Medicine administration is also a key factor.

Mobile applications have been demonstrated to be useful tools in improving medication adherence [155].

Surveys have been widely used as a method to collect data. There exist different methods to answer surveys, such as paper-based, telephone, online or mixed surveys. Arafa et al. [156] analyzed online surveys in dermatology and observed many pros such as better data quality and time-saving when compared to paper surveys. The online surveys were shown to have concerns regarding security and confidentiality, and vulnerabilities to hacking. Maymone et al. [157] summarized different web-based survey tools available and analyzed their usability in dermatologic research. Notwithstanding, these tools rely on third parties thereby compromising the sensitivity of the personal data shared in the surveys.

The objective of this work is to provide a simple interface for patients and doctors to interact, collect, and manage sensitive medical and personal data over multiple devices while still ensuring the highest level of security and privacy. We introduce a multi-platform system architecture for patients' monitoring in chronic scenarios grounded on three important aspects: ***security*** enabled by paying special attention to sensitive and private health and personal data, ***simplicity*** by providing easy-to-use interfaces across multiple platforms, and ***flexibility*** by using a multi-platform architecture based on scalable and flexible microservices. The architecture is expected to improve the effectiveness of disease management in chronic scenarios.

A.1.1 System overview

Online surveys are vulnerable to several issues of security, privacy, and confidentiality. Additionally, most of the surveys use third-party questionnaire tools thereby compromising the sensitivity of the collected personal data. Thus, to overcome such vulnerabilities while employing online surveys to collect health data, we implemented a secure and interactive multi-platform system to conveniently collect and manage patients' data and information. Our proposed multi-platform prototype provides two different interactive user interfaces - a website and a messaging platform, with an integrated virtual assistant. On the website, a user (patient or physician) can either use different management tools to manage the collection of the information or interact with the virtual assistant.

The overview of the scenario is shown in Figure A.1. In order to provide access to the tools and services, the system provides two different user interfaces, a website and a messaging platform where the virtual assistant is integrated. On the website, the user is able to use different management tools, as well as the virtual assistant that is integrated within the website. The data gathered on the website is encrypted and sent to the servers using HTTPS. The messaging platform used as a communication channel to interact with the virtual assistant is Signal. In addition, a user access control is present in the system to control access functionalities to the private data and the virtual assistant conversation through the website.

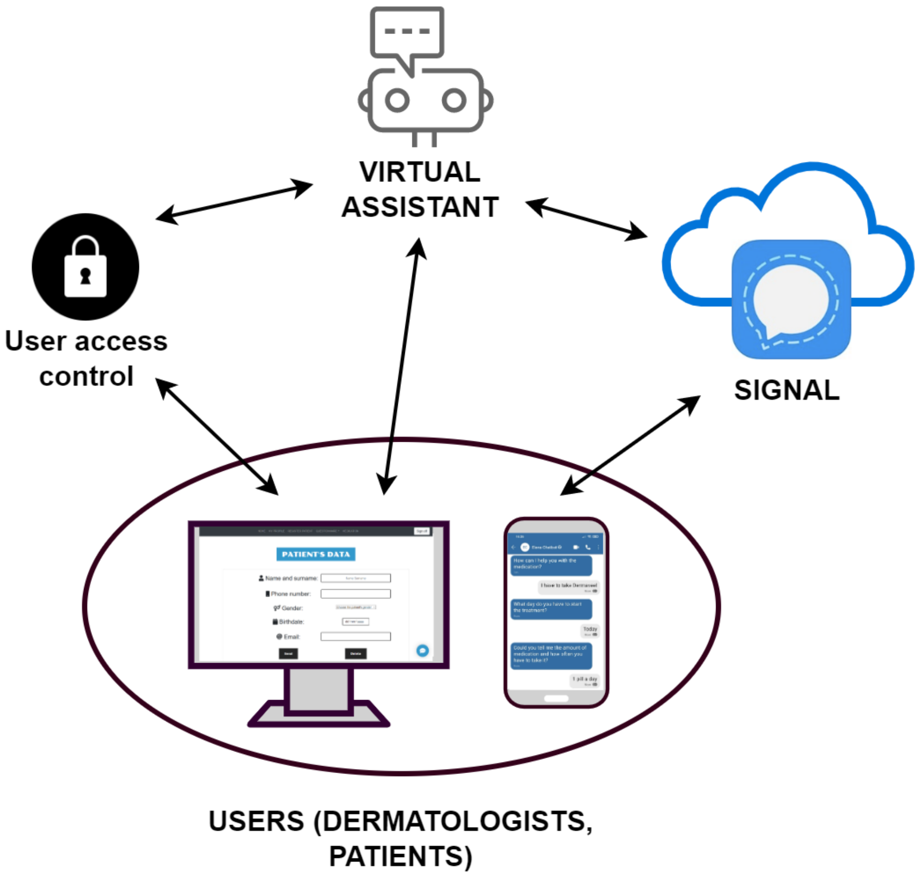


Figure A.1: Overview of the scenario.

The programming language used for developing the website is JavaScript with React. The main functionalities covered by the virtual assistant and the website are questionnaire management, medication management, and reminders.

A.1.2 Virtual assistant structure

The internal virtual assistant architecture (see Figure A.2) is based on the microservice architecture described in Section 3.1. In this multi-platform scenario, the virtual assistant has two proxies: one for receiving web petitions and the other one for communicating with the messaging platform. These proxies are in charge of translating the message received in a formatted JSON specific for the correct use of the architecture. All the petitions exchanged within the virtual assistant use HTTPS protocol and are authenticated with a specific access control for the services. The API gateway then receives the formatted message and decides which function should receive each petition. Another API gateway is added to the architecture with a user access control that handles the requests from the website. These requests can be used for obtaining users' information from the database or for using one of the functionalities developed for this scenario. All the information collected is standardized using HL7 FHIR for interoperability between different healthcare systems. The conversation with the virtual assistant is based on menu options, where the AIML implementation is used to understand the user.

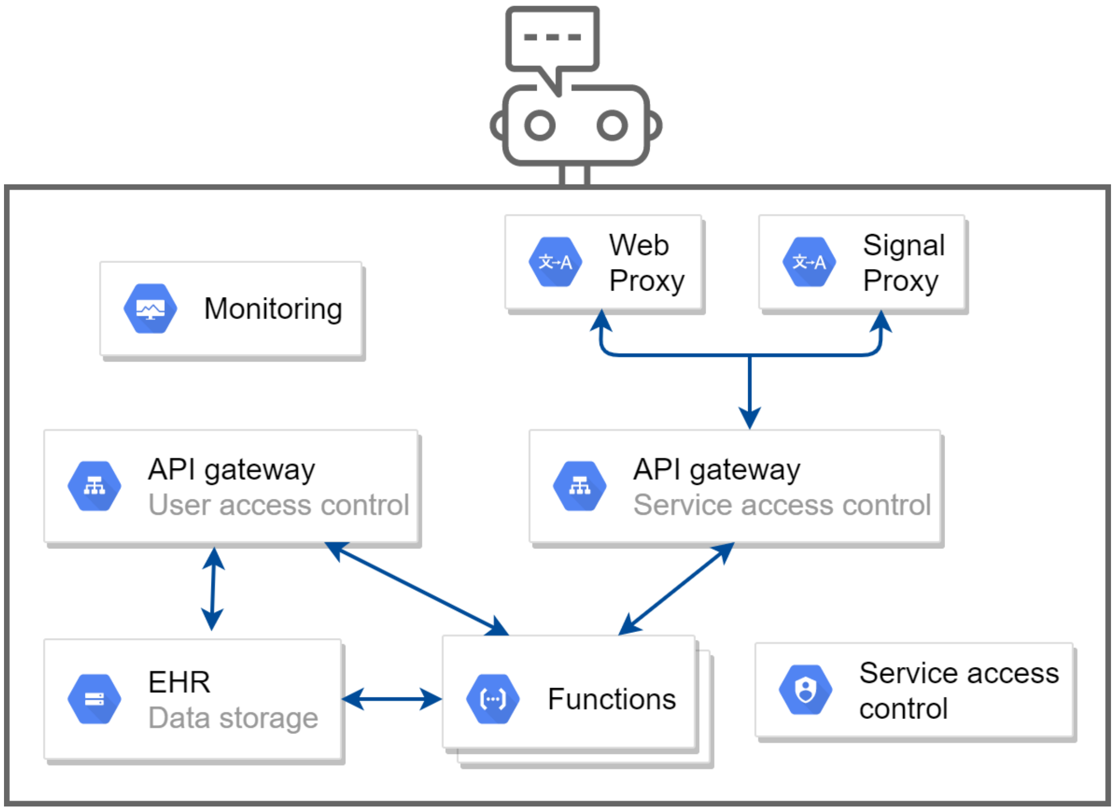


Figure A.2: Internal virtual assistant structure.

A.1.3 Web structure

The web structure components are the sign-up, the sign-in, the services offered based on the user's role and the virtual assistant integrated on the website. The web uses the web proxy of the virtual assistant structure to obtain the conversation interaction whenever the user uses the virtual assistant integrated on the website. The web uses specific functions from the virtual assistant architecture, thanks to the API gateway configured with the user access control. The services that the website uses are the following:

- **Sign-up microservice:** The microservice provides an API to register a user in the databases of the virtual assistant architecture.
- **Notification microservice:** With this API, the website is able to create or delete notifications related to answering questionnaires and medication management.
- **Medical information database:** The database has an API to query, create and update the medical information.

A.1.4 Services

The website offers the following services:

- **Questionnaire management:** The physicians are able to create, modify and assign questionnaires to patients. The patients are able to answer and see their progress in answering the questionnaires. Both are able to export the data of the questionnaires.
- **Medication management:** The physicians are able to assign and check the medication of the patients. Physicians and patients are able to update and check their own medication.
- **Register patients:** The physicians are able to register patients by their name, email, gender, birth date and phone number.

The virtual assistant offers all the website services, as well as the following service:

- **Reminders:** The patients receive reminders about questionnaires whenever they need to fill them. The patients also receive reminders about their medication after configuring them.

A.1.5 Access control

For the authentication in the system, we use three access control mechanisms. The first access control mechanism is to provide authentication between the microservices in the virtual assistant structure and it follows the standard JWT.

The second mechanism provides authentication to users while using the messaging platform. The mechanism used is the Signal's user authentication, which relies on a public key. When the user installs the application, the key pair is generated and the public keys are exchanged with other users for the first time they interact with each other.

The last mechanism is to provide authentication to users while using the website. A JWT authentication server is configured to provide tokens to users when they sign-in with their user and password. This server is called User access control in our system and allows data access based on the user's role.

The rules for the website are provided with the field scope in the token and it specifies the read, write or all-access for the different resources and services that the user is permitted to use. The rules for this scenario are described in Table A.1 and the detailed structure of the rules is the following:

FHIR resources:

Resource name / { id | patients | *} . { read | write | *}

API services:

Name of the API / { id | patients | *} . { read | write | *}

FHIR resources	
Physicians	**Patients**
Practitioner/id.read	Patient/id.read
Patient/patients.*	Questionnaire/id.read
Questionnaire/id.*	QuestionnaireResponse/id.*
QuestionnaireResponse/patients.read	CarePlan/id.read
List/id.*	Medication/id.*
CarePlan/patients.*	MedicationRequest/id.*
Organization/id.read	MedicationAdministration/id.*
Medication/patients.*	
MedicationRequest/patients.*	
MedicationAdministration/patients.*	
API services	
Physicians	**Patients**
Proxy/id.*	Proxy/id.*
Notifications/*.write	
ApiGateway/patients.write	

Table A.1: List of permissions based on the user's role.

A.1.6 User interfaces

Some examples of the application screens are shown in Figures A.3 and A.4 (the work has been done in Spanish, but examples are translated for ease of reading). A usage example of the website is shown in Figure A.3, where a physician is creating a questionnaire. The physician can create any type of questionnaire and assign it to the patients. Once it is assigned, the patients will receive a notification on their phones saying that they have available a new questionnaire. The physicians can specify the time and the frequency to remind the patients to do the questionnaire on their phones. The physicians are able to export the questionnaire data into an excel document for further data analysis. Moreover, the interaction with the virtual assistant within the website is shown in Figure A.3, where a button-based menu can be used. In addition, the interaction of a user answering a questionnaire using a smartphone is shown in Figure A.4.

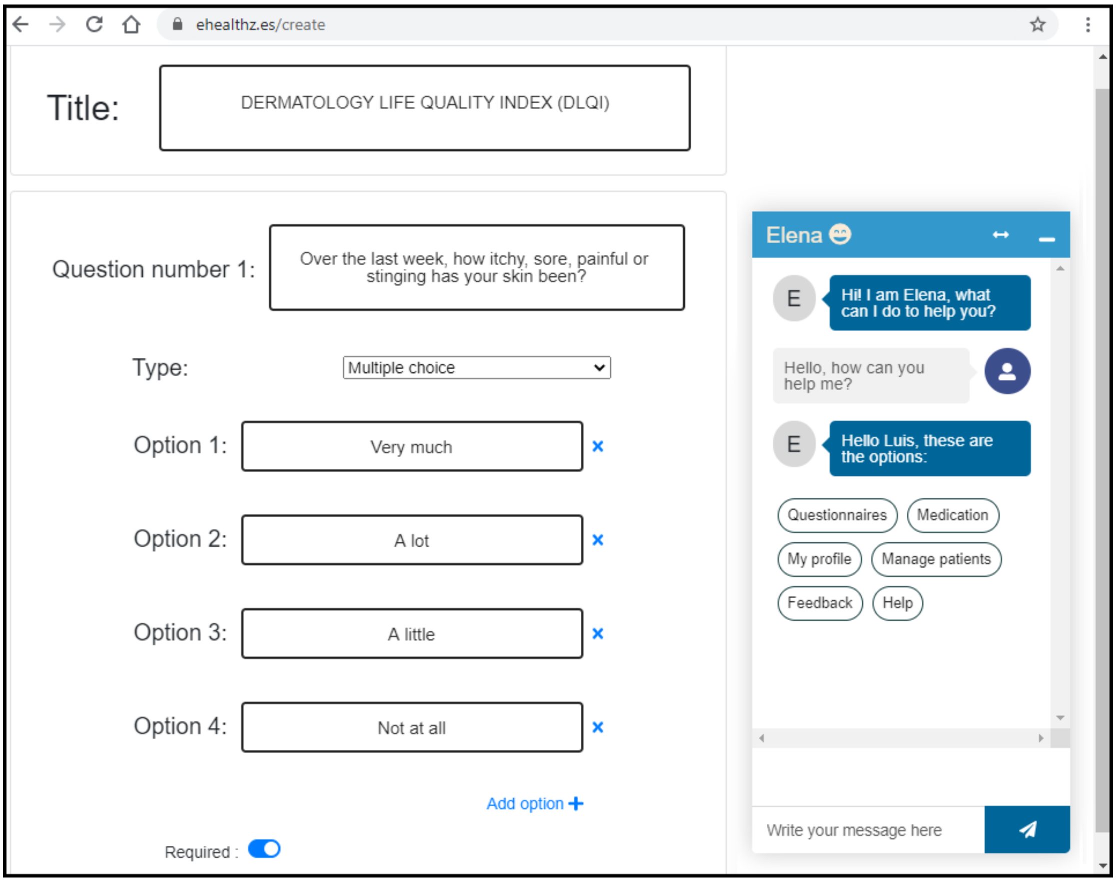


Figure A.3: Example of creating a questionnaire using the website.

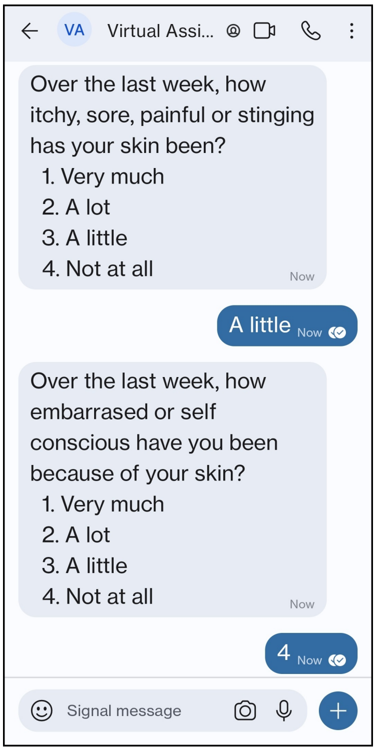


Figure A.4: Example of questionnaire interaction with the virtual assistant using a smartphone.

A.1.7 Discussion

Our system follows a secure technology and provides end-to-end encryption while using messaging platforms and TLS for the rest of the communications. Moreover, a detailed system component for access control is developed to provide access to the data based on the roles and access rights. The component also follows the consent of the users before making it accessible to others by asking the user if he/she consents to the use of the personal data within the virtual assistant. As such, all the information held and exchanged within our system is secure and resilient against any unauthorized usage.

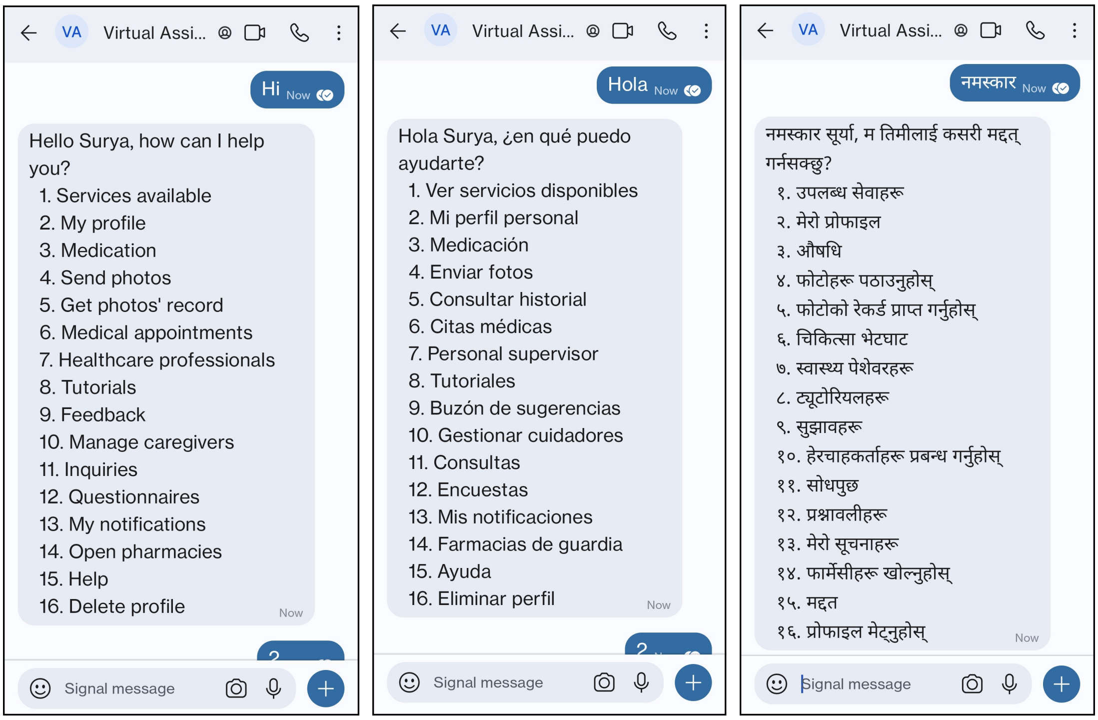
Our system offers a simple way to interact with the system by offering two different interfaces - the website and the virtual assistant with diversified usabilities. The website provides extra tools to make the questionnaire management more fluent and easy for the users, for example, at work. On the other hand, the virtual assistant in smartphones provides quick access to the data. The entire system architecture is based on microservices. Consequently, our system is flexible and scalable and new services can easily be added to the system as a new microservice component.

In our system, Signal is used as the messaging platform, which in general has some limitations. Signal is not one of the most common messaging platforms and it does not allow the users to interact with buttons, a feature that is available with other platforms like Telegram. Nevertheless, we expect our system to be useful to patients and physicians as the multi-platform system offers several useful features to collect and manage health data. The system can also be a central point to securely store the patients' data which can be easily accessed by physicians when required for any analysis as all the data is stored following international standards. The use of two different interfaces within the system is expected to reduce the limitation of technology illiteracy.

A.2 Multilingual microservice prototype

Designing a system that can be easily adapted to other languages is an important factor if telemonitoring systems want a global impact. We designed our microservice-based virtual assistant architecture in order to provide scalability and flexibility. One point in this scalability is to add new languages easily. This multilingual prototype was designed and developed to validate the scalability while integrating different languages.

The microservice has one new AIML file for each new language. The language is configured based on the phone number's prefix (the developed prototype is designed to work in a scenario with different countries). The multilingual microservice prototype proposed has been developed in three different languages: English (West Germanic language), Spanish (Romance language), and Nepali (Indo-Aryan language). An example of the interaction with the multilingual prototype is shown in Figure A.5.



(a) English example. **(b)** Spanish example. **(c)** Nepali example.

Figure A.5: Interaction example with the multilingual virtual assistant using a smartphone.

Appendix B

Intermediate Interview Questions

Evaluation guide

Evaluation guide to assess the patients' opinions about the usability of the virtual assistant: annotate all of the patient's comments.

1. General impression
   - Was the use and communication with the virtual assistant easy to learn?
   - Was the virtual assistant useful?
   - How frequently did you use the virtual assistant?
   - Did the virtual assistant cover the needs of the medication reminders?
   - Were the functionalities used in the virtual assistant easy to understand?
2. Problems observed
   - Was there any problem while using the virtual assistant?
   - Did the medication reminders work correctly?
   - Were the messages/the vocabulary of the virtual assistant appropriate?
3. Suggestions/improvements/opinions
   - Did any needs arise not covered by the virtual assistant?
   - Conclusions /final opinions

Appendix C

Post-Study Interview Questions

C.1 Patient's Interview

You will find a set of statements below. Please express exactly what you think in each case without worrying about whether other people would agree with you. Mark with an X the box that corresponds to your particular situation:

STATEMENTS	BEFORE THE VIRTUAL ASSISTANT					AFTER THE VIRTUAL ASSISTANT				
	Always	Almost Always	Sometimes	Rarely	Never	Always	Almost Always	Sometimes	Rarely	Never
1. You take the medications at the set time										
2. You take all the indicated dosages										
3. You attend your medical appointments										
4. You accommodate your medication schedules to your daily life activities										
5. Your family or friends are involved in your care										
6. You complete the medical treatment without supervision of your family or friends										
7. How often is it difficult to remember that you should take all your medicines?										

QUESTION	YES	NO
A—Do you ever forget to take medications to treat your disease?		
B—When you are well, do you stop taking your medication?		
C—If you ever feel bad, do you stop taking it?		
D—Did you find it easy to use the virtual assistant?		
E—Do you think it is useful for you?		
F—Do you think the virtual assistant improves your medication adherence?		
G—Do you think you will continue using the virtual assistant after the project?		
H—From a certain moment, the virtual assistant began to give the weather forecast. Has this helped you to use the virtual assistant more frequently?		

DATA	BEFORE	AFTER
HbA1c		
Number of medical appointments per month		

Thank you for your collaboration. We hope to improve our service with your help.

C.2 Healthcare Professional's Interview

You will find a set of questions below. Please express exactly what you think in each case without worrying about whether other people would agree with you. Mark with an X the box that corresponds to your particular situation:

QUESTION	YES	NO
A—Did you find the virtual assistant easy to use?		
B—Do you think it is useful for you?		
C—Do you think it is useful for the patients?		
D—Do you think the virtual assistant improves the medication adherence of the patients?		
E—Did you use the functionalities to monitor the medication intakes of your patients?		
F—Do you think you will continue using the virtual assistant after the project?		

QUESTION	All the Patients	Almost All the Patients	Some Patients	Almost No Patient	No Patient
1. In the medical appointment, did you observe motivation in using the virtual assistant on the part of the patients?					
2. In the medical appointment, did you observe dissatisfaction with the virtual assistant on the part of the patients?					
3. In the medical appointment, did you observe an improvement in the patients' medication intake?					

Other observations of interest:

Thank you for your collaboration. We hope to improve our service with your help.

Appendix D

Patient's satisfaction survey

Name:

Surname:

For each statement, please choose the best option that suits your experience with the virtual assistant.

Virtual assistant usability:

AFFIRMATIONS	**Strongly agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
I think that I would like to use this virtual assistant frequently					
I found the virtual assistant unnecessarily complex					
I thought the virtual assistant was easy to use					
I think that I would need the support of a technical person to be able to use this virtual assistant					
I found the various functions in this virtual assistant were well integrated					
I thought there was too much inconsistency in this virtual assistant					
I would imagine that most people would learn to use this virtual assistant very quickly					
I found the virtual assistant very cumbersome to use					
I felt very confident using the virtual assistant					
I needed to learn a lot of things before I could get going with this virtual assistant					

Virtual assistant utility:

AFFIRMATIONS	**Totally agree**	**Quite agree**	**Somewhat agree**	**Neutral**	**Somewhat disagree**	**Quite dis-agree**	**Totally dis-agree**
The language used by the virtual assistant was appropriate							
The virtual assistant was useful to improve my quality of life							
The virtual assistant made it convenient for me to communicate with my health care provider							
I felt confident that any information I sent to my provider using the virtual assistant would be received							
I felt comfortable communicating with my health care provider using the virtual assistant							

What option/feature do you like most about the virtual assistant?

❍ Consultations

❍ Questionnaires

❍ Medication

❍ Send photos

❍ Medical appointments

❍ Other:

Questionnaires:

QUESTION	**Too easy**	**Easy**	**Somewhat easy**	**Difficult**	**Too difficult**
How easy or difficult was it for you to answer the questionnaires in the virtual assistant?					

What were the main reasons why you left questionnaires unanswered?

- ❍ I don't have time
- ❍ Too monotonous
- ❍ Too long
- ❍ I did not consider that they were necessary for the psoriasis follow-up
- ❍ I read the notification but then I forgot to answer
- ❍ I was not aware of the mobile notification
- ❍ Other:

Face-to-face medical appointments:

QUESTION	**Always**	**Almost always**	**Sometimes**	**Rarely**	**Never**
Did you attend face-to-face medical appointments BEFORE using the virtual assistant?					
Do you attend face-to-face medical appointments AFTER using the virtual assistant?					

Did you do medical consultations with your dermatologist through the virtual assistant?

❍ Yes

❍ No

If you did medical consultations, were they resolved satisfactorily?

❍ Yes

❍ No

Has having a tool with which to be able to contact your dermatologist directly provided you with security/peace of mind?

❍ Yes

❍ No

Do you think the virtual assistant improves your treatment adherence?

❍ Yes

❍ No

Do you stop using the virtual assistant when you feel better with your symptoms?

❍ Yes

❍ No

Have you reduced the number of face-to-face medical consultations after using the virtual assistant?

❍ Yes

❍ No

Do you think you will continue using the virtual assistant after the project?

❍ Yes

❍ No

Do you have any suggestions to improve the virtual assistant?

What would encourage you to use the virtual assistant more often?

Any other comment

Thank you very much for participating! Your opinion is very important for us!

Appendix E

Healthcare professional's satisfaction survey

Name:

Surname:

For each statement please choose the best option that suits your experience with the virtual assistant.

Virtual assistant usability:

AFFIRMATIONS	**Strongly agree**	**Agree**	**Neutral**	**Disagree**	**Strongly disagree**
I think that I would like to use this virtual assistant frequently					
I found the virtual assistant unnecessarily complex					
I thought the virtual assistant was easy to use					
I think that I would need the support of a technical person to be able to use this virtual assistant					
I found the various functions in this virtual assistant were well integrated					
I thought there was too much inconsistency in this virtual assistant					
I would imagine that most people would learn to use this virtual assistant very quickly					
I found the virtual assistant very cumbersome to use					
I felt very confident using the virtual assistant					
I needed to learn a lot of things before I could get going with this virtual assistant					

Virtual assistant utility:

AFFIRMATIONS	**Totally agree**	**Quite agree**	**Somewhat agree**	**Neutral**	**Somewhat disagree**	**Quite dis-agree**	**Totally dis-agree**
The language used by the virtual assistant was appropriate							
The virtual assistant was useful to improve the quality of life of my patients							
The virtual assistant made it convenient for me to communicate with my patients							
I felt confident that any information I sent to my patients using the virtual assistant would be received							
I felt comfortable communicating with my patients using the virtual assistant							

Do you think the virtual assistant improves the patients' treatment adherence?

❍ Yes

❍ No

Was the quality of the images sufficient to make diagnoses?

❍ Yes

❍ No

What option/feature do you like most about the virtual assistant?

❍ Answer consultations

❍ Questionnaires

❍ Medication

❍ Send photos

❍ Medical appointments

❍ Send message

❍ Psoriasis scale

❍ Other:

Questionnaires:

QUESTION	**Too easy**	**Easy**	**Somewhat easy**	**Difficult**	**Too difficult**
If you generated or assigned questionnaires, how easy or difficult was it for you?					

Was it easy for you to answer patients' queries?

❍ Yes

❍ No

Have you seen the number of face-to-face medical consultations of patients reduced after using the virtual assistant?

❍ Yes

❍ No

Do you think you will continue using the virtual assistant after the project?

- ❍ Yes
- ❍ No

Do you have any suggestions to improve the virtual assistant?

What would encourage you to use the virtual assistant more often?

Any other comment

Thank you very much for participating! Your opinion is very important for us!