INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Developing an AI Virtual Assistant for Higher Education

Patrícia Marquês Bacelar

Master in Business Analytics

Supervisor:
Doctor Raul Manuel Silva Laureano, Associate Professor, ISCTE Business School

October, 2025

Department of Quantitative Methods for Management and Economics

Developing an AI Virtual Assistant for Higher Education

Patrícia Marquês Bacelar

Master in Business Analytics

Supervisor:
Doctor Raul Manuel Silva Laureano, Associate Professor, ISCTE Business School

October, 2025

DEDICATION

I dedicate this thesis to my father, who taught me never to give up, even when life became difficult. His strength and love have always inspired me, and it was with that inspiration that I found the courage to keep going and reach this moment. Wherever he is, I hope he is proud of me, as I'm of how far I've come.

ACKNOWLEDGEMENTS

To ISCTE,

For the opportunity to develop this project and for all the support provided throughout my academic journey.

To Professor Doctor Raul Laureano,

For his invaluable guidance, encouragement, and availability, which were essential to the completion of this work.

To all the students who participated in the study,

For their collaboration and valuable contribution to the evaluation of the IBS Virtual Assistant.

To my mother, Cristina Bacelar,

For her unconditional love, patience, and unwavering support, which have shaped me into the person I am today.

To my brother, Tiago Bacelar,

For being my constant source of inspiration and motivation to always strive for more.

To my grandparents,

For their affection, wisdom, and life lessons, which have deeply marked me throughout my life.

To my family, especially my godfather, André,

For his constant support, strength, and motivation throughout my life.

To my friends,

For their friendship, laughter, and encouragement, especially during the most challenging moments.

To my classmates,

For the shared knowledge, teamwork, and mutual support along this journey.

ABSTRACT

This research investigates the design, development, and evaluation of an AI-powered educational chatbot, the IBS Virtual Assistant, created to support master's students at ISCTE Business School. Motivated by the growing demand for scalable digital solutions to enhance communication and reduce repetitive administrative tasks, the study follows the Design Science Research Methodology through iterative design and testing cycles. The chatbot's knowledge base was structured around ten validated informational categories derived from an institutional questionnaire with 542 student-generated questions. Implemented in Microsoft Copilot Studio, two prototypes were developed and evaluated. Quantitative results revealed significant improvements between versions, with accuracy rising from 0.86 to 1.00 and completeness from 0.71 to 0.86, while maintaining perfect clarity and tone. A complementary user evaluation of eleven master's students, based on the Unified Theory of Acceptance and Use of Technology model, showed high mean scores across all constructs, confirming strong perceived usefulness, ease of use, and acceptance. These findings validate the methodological robustness of the Design Science Research approach and demonstrate the potential of conversational artificial intelligence to enhance informational accessibility and institutional support. The study contributes theoretically by integrating the Design Science Research and Unified Theory of Acceptance and Use of Technology frameworks, and practically by delivering a replicable and scalable solution for higher education.

Keywords: AI chatbot, Design Science Research, higher education, student support.

JEL Classification System: I23 (Higher Education), O33 (Technological Innovation).

RESUMO

Esta investigação analisa o processo de conceção, desenvolvimento e avaliação de um *chatbot* educacional com base em inteligência artificial, o IBS Virtual Assistant, criado para apoiar estudantes de mestrado da ISCTE Business School. Motivado pela crescente necessidade de soluções digitais escaláveis que melhorem a comunicação e reduzam tarefas administrativas repetitivas, o estudo segue a Design Science Research Methodology, através de ciclos iterativos de conceção e teste. A base de conhecimento do chatbot foi estruturada em dez categorias informacionais validadas, derivadas de um questionário institucional com 542 questões formuladas por estudantes. Implementado na plataforma Microsoft Copilot Studio, foram desenvolvidos e avaliados dois protótipos. Os resultados quantitativos revelaram melhorias significativas entre versões, com a precisão a aumentar de 0,86 para 1,00 e a completude de 0,71 para 0,86, mantendo pontuação perfeita em clareza e tom. Uma avaliação complementar com utilizadores, envolvendo onze estudantes de mestrado e baseada no modelo da Unified Theory of Acceptance and Use of Technology, apresentou médias elevadas em todas as dimensões, confirmando uma forte perceção de utilidade, facilidade de uso e aceitação. Estes resultados validam a robustez metodológica da abordagem Design Science Research e demonstram o potencial da inteligência artificial conversacional para melhorar o acesso à informação e o apoio institucional no ensino superior. O estudo contribui teoricamente, ao integrar as abordagens Design Science Research e Unified Theory of Acceptance and Use of Technology, e praticamente, ao oferecer uma solução replicável e escalável para o ensino superior.

Palavras-chave: Chatbot com Inteligência Artificial, Design Science Research, ensino superior, apoio ao estudante.

Sistema de Classificação JEL: I23 (Educação Superior), O33 (Inovação Tecnológica).

LIST OF ACRONYMS AND ABBREVIATIONS

AI	Artificial Intelligence
BA	Business Analytics
BERT	Bidirectional Encoder Representations from Transformers
BLEU	Bilingual Evaluation Understudy
DSR	Design Science Research
DSRM	Design Science Research Methodology
e.g.	For example
EN	English
et al.	And others
FEDS	Framework for Evaluation in Design Science
GPT	Generative Pre-trained Transformer
IBS	ISCTE Business School
ISCTE	ISCTE – Instituto Universitário de Lisboa
LLMs	Large Language Models
MBA	Master's in Business Analytics
ML	Machine Learning
NLP	Natural Language Processing
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PT	Portuguese
Q&A	Question-and-answer
RNNs	Recurrent Neural Networks
RPA	Robotic Process Automation
SPSS	Statistical Package for the Social Sciences
UTAUT	Unified Theory of Acceptance and Use of Technology
WoS	Web of Science

INDEX

INDEX OF TABLE

INDEX OF FIGURES

1 INTRODUCTION

This chapter provides a comprehensive overview of the topic under investigation, highlighting its relevance in the current context and its alignment with digital transformation in higher education institutions. It outlines the research problem, objectives, and key contributions, both academic and practical. Furthermore, the adopted methodological approach is presented, establishing a solid foundation for subsequent analyses. Finally, the structure of the study is detailed, offering a clear and organised overview of its chapters.

1.1 Theme and its Importance

The growing integration of digital technologies into higher education has led institutions to explore innovative ways of improving communication and administrative efficiency. Artificial intelligence (AI), particularly in the form of conversational systems, has emerged as a key enabler of this transformation, offering the potential to support students and academic staff through automated and scalable information services (Jezler et al., 2025; Chen, 2024). Higher education institutions face increasing administrative workloads due to rising volumes of student inquiries related to admissions, academic structures, and assessment processes. This challenge is particularly evident during peak academic periods, when communication channels become overloaded, leading to delayed responses and reduced efficiency in academic support (Thinnyun et al., 2021).

AI technologies are increasingly applied to enhance operational processes, reduce repetitive administrative tasks, and improve access to institutional information (Dwivedi et al., 2023). The expansion across sectors, driven by advances in automation, data analysis, and natural language processing (NLP), reflects a global trend toward the optimization of services that traditionally relied on human mediation (Haenlein & Kaplan, 2019). In the educational context, this transformation was accelerated by the COVID-19 pandemic, which highlighted the importance of digital tools in maintaining continuity of communication and academic support (Jezler et al., 2025). Within this technological evolution, chatbots have gained prominence as one of the most practical applications of AI in educational institutions (Okonkwo & Ade-Ibijola, 2021). These conversational agents are designed to simulate human dialogue using NLP and machine learning (ML), enabling real-time interaction between students and institutional systems (Shawar & Atwell, 2007; Thinnyun et al., 2021). Research indicates that such systems can provide efficient responses to frequently asked questions, thereby reducing repetitive workloads and facilitating continuous information access for students and staff (Anjulo Lambebo & Chen, 2024; Akinwalere & Ivanov, 2022; MA et al., 2024; Ravey, 2024). Additionally, chatbots enhance learning environments by providing immediate and consistent feedback, optimizing institutional resources, and improving student retention rates (Peyton et al., 2025). Chatbots offer continuous availability, providing support 24/7 and improving overall user experience, as highlighted by Guszcza et al. (2018), who emphasize their cost-effectiveness and efficiency in service delivery. In the educational context, Shawar and Atwell (2007, p. 57) note that "chatbots offer a practical solution to optimize communication across sectors, including educational institutions, by enabling fast and effective interactions". Similarly, interviews conducted by Jezler et al. (2025) revealed recurring concerns among academic staff about the overload of student requests, particularly during critical periods of the semester. Respondents reported that chatbots could alleviate repetitive administrative tasks and enhance responsiveness, noting that students are increasingly accustomed to technology-mediated communication and expect immediate, accessible support.

However, despite their advantages, the integration of AI-powered chatbots in higher education remains dependent on factors such as content accuracy, reliability, and user trust. Concerns persist regarding chatbot adaptability to complex queries, user acceptance, and integration with existing academic platforms (Yeoh, 2022). As Chen (2024) points out, educational question-and-answer (Q&A) systems require domain expertise and structured datasets to ensure factual correctness and contextual relevance. Similarly, Thinnyun et al. (2021) highlight that, in large-scale academic environments, the volume of student inquiries often surpasses available human resources, particularly during exam or enrolment periods, reinforcing the potential value of automated conversational tools. While there is a general consensus that automation could enhance productivity, concerns about depersonalization remain (Chassignol et al., 2018). These findings reinforce the need for intelligent automation tools, such as chatbots, that streamline administrative communication while preserving essential human interaction (Jenneboer et al., 2022).

Recent developments in large language models (LLMs) have further expanded the capabilities of chatbots, allowing them to generate contextually appropriate and personalized responses. Nevertheless, their application in formal education requires careful control, as these systems may still produce factual inaccuracies or incomplete answers if not constrained by institutional data sources (Chen, 2024; Dwivedi et al., 2023). In this context, the use of AI-powered chatbots in higher education represents a relevant and timely research field. By exploring their implementation in an institutional setting such as ISCTE Business School (IBS), this study contributes to understanding how conversational AI can complement traditional academic support mechanisms, potentially improving communication efficiency while maintaining accuracy, accessibility, and alignment with institutional standards.

1.2 Research Problem and Question

Although higher education institutions increasingly adopt digital solutions to enhance communication, many still rely heavily on human-mediated processes to address students' academic and administrative inquiries (Chatterjee & Bhattacharjee, 2020). This dependence often results in information bottlenecks and response delays, particularly during peak academic periods, when the volume of student requests surpasses available staff capacity (Thinnyun et al., 2021). As highlighted by Jezler et al. (2025), academic personnel frequently report difficulty managing repetitive questions, acknowledging that chatbots could automate routine interactions and alleviate workloads. Similarly, Shawar and Atwell (2007) emphasize that conversational systems can optimize communication by providing fast and effective interactions, while maintaining institutional consistency. However, the effective integration of AI-powered chatbots in higher education still requires research on their reliability, contextual accuracy, and perceived usefulness (McGrath et al., 2024). Accordingly, this study addresses the central question: "How can an AI-powered chatbot be effectively and efficiently used to support users in higher education contexts?". By exploring this question, the research seeks to determine how conversational automation can improve both the accuracy and speed of information delivery, reduce repetitive administrative workload, and ultimately enhance communication and user experience within IBS.

1.3 Objectives and Contributions

The main goal of this research is to develop and analyse the impact of an AI-powered educational chatbot designed to support master's students in higher education. In alignment with the research question (see Chapter 1.2), the following specific objectives were defined: (i) to identify and analyse students' most frequent informational needs; (ii) to design and develop the virtual assistant; and (iii) to assess user perceptions and acceptance.

By achieving these objectives, this study is expected to contribute both to academic knowledge and to professional practice. From a theoretical perspective, it advances the understanding of how conversational AI and Design Science Research (DSR) can be combined to develop and evaluate educational technologies in higher education contexts. From a practical standpoint, it provides insights and methodological guidance for institutions seeking to implement AI-powered virtual assistants to improve communication efficiency, reduce administrative workload, and enhance student support services.

1.4 Methodological Approach

This study follows the Design Science Research Methodology (DSRM), a structured approach commonly used in information systems research to develop and assess technological innovations (Hevner et al., 2004; Peffers et al., 2007). The methodology guided the project through iterative cycles of design, testing, and evaluation, allowing the artefact to be progressively refined based on empirical evidence.

The research was conducted at IBS, a faculty of ISCTE – Instituto Universitário de Lisboa (ISCTE), and focused on developing an AI-powered educational chatbot to support master's students. The process began with identifying students' informational needs through a questionnaire that collected 542 individual questions from 65 respondents. These data were analysed and classified into ten validated informational domains, which served as the foundation for structuring the chatbot's knowledge base. The virtual assistant was then built in Microsoft Copilot Studio, using a low-code approach aligned with ISCTE's institutional ecosystem. Two design iterations were performed: an initial prototype to test feasibility and a final prototype incorporating improvements in accuracy, completeness, and interaction design. Evaluation combined technical and user-centred perspectives. The system's performance was assessed using four metrics, accuracy, completeness, clarity, and tone, while user perception and acceptance were assessed through a questionnaire grounded in the Unified Theory of Acceptance and Use of Technology (UTAUT). This survey, applied to eleven master's students who interacted with the assistant via Microsoft Teams, provided complementary insights into the chatbot's effectiveness, usability, and acceptance within an academic environment.

1.5 Structure of the Dissertation

To explore how the implementation of an AI-powered educational chatbot can support administrative and informational processes within the academic context, this research is structured into five chapters, each corresponding to a specific stage of the study.

The first chapter introduces the research theme and its relevance, outlines the research problem, and presents the main objectives and expected contributions. It also briefly describes the methodological approach adopted throughout the study. The second chapter provides the literature review, beginning with the definition of key concepts related to AI and educational chatbots. It then presents the systematic literature review, detailing the research protocol, inclusion and exclusion criteria, and the critical synthesis of the selected studies. The chapter concludes with an evaluation of the reviewed articles and the identification of research gaps that guided this investigation. The third chapter focuses on the methodological framework, following the DSR approach. It describes the problem identification process, the analysis of institutional data and student questionnaires, and the definition of objectives for the solution. Furthermore, it details the design and development of the *IBS Virtual Assistant*, as well as the stages of demonstration, evaluation, and communication. Chapter four presents and discusses the results obtained. It analyses the prototypes developed, summarises the evaluation findings, and discusses the implications of the results for academic management and student support. Recommendations for improving future chatbot implementations are also provided. Finally, chapter five concludes the study by summarizing the main contributions of this research, outlining its limitations, and proposing directions for future work.

2 LITERATURE REVIEW

This chapter provides a review of research on AI-powered chatbots and virtual assistants, examining their implementation and real-world applications across different domains. It examines studies on AI-driven process automation and human-machine interaction, highlighting how chatbots contribute to optimizing communication.

Thus, a brief and clear definition of the main concepts involved follows, essential for framing this research. Subsequently, a critical analysis of the existing literature is conducted, exploring different applications of automation models, empirical results, adopted methodologies, key contributions, and limitations.

2.1 AI Systems: Key Concepts and Importance

To ensure a robust understanding of the foundations of this study and its objectives, it is essential to define the key concepts involved, namely AI, chatbots, and NLP. These technologies constitute the core of automation systems designed to enhance efficiency, accessibility, and communication within higher education institutions.

AI has become an indispensable element of modern society, permeating sectors ranging from business to education, and its accelerated evolution enables the automation of complex cognitive tasks traditionally performed by humans, improving efficiency, reducing costs, and delivering timely and data-driven insights (Jezler et al., 2025). AI systems, such as chatbots and LLMs, automate repetitive tasks, provide real-time feedback, and improve student engagement through adaptive learning strategies (Bilquise & Shaalan, 2022; Moșteanu, 2022). The use of predictive analytics allows institutions to identify at-risk students and implement early interventions (Sharma et al., 2023). During the post-pandemic digital transformation, AI systems became central to improving institutional operations, decision-making, and the overall student experience (Jezler et al., 2025).

Chatbots, conversational agents that simulate human interaction, have evolved substantially since the creation of *ELIZA* by Weizenbaum (1966). They can be broadly classified into *rule-based systems*, which rely on predefined scripts, and AI-powered systems, which employ machine learning and NLP to generate dynamic, context-aware responses (Adamopoulou & Moussiades, 2020). Within academia, chatbots have been adopted for student support, course information, and administrative assistance, effectively reducing the workload of staff and providing 24/7 availability (Hardi et al., 2022; Ilieva et al., 2023). Studies have shown that these tools are particularly valuable during high-demand periods, such as enrolment and examination seasons, when human resources are overextended (Thinnyun et al., 2021). By automating responses to frequently asked questions, chatbots improve response speed, institutional efficiency, and user satisfaction (Shawar & Atwell, 2007, p. 57; Jezler et al., 2025). Recent advancements in LLMs, such as ChatGPT (Goel et al., 2021), ChatGLM (Du et al., 2022), and LLaMA2-

Chat (Touvron et al., 2023), have enabled educational chatbots to generate personalized and contextually rich responses. These models, based on the Transformer architecture introduced by Vaswani et al. (2017), represent a milestone in intelligent automation, allowing systems to process and generate natural language with remarkable accuracy. However, while LLMs demonstrate high adaptability, their deployment in educational contexts requires rigorous validation to mitigate issues of reliability, bias, and misinformation (Chen, 2024; Dwivedi et al., 2023; Dalalah & Dalalah, 2023). As Chen (2024) highlights, educational Q&A systems depend heavily on the establishment of robust, domain-specific reference frameworks to ensure factual correctness and relevance.

NLP, a core subfield of AI, enables machines to interpret and generate human language. It is foundational to the operation of chatbots and LLMs, powering functionalities such as intent recognition, sentiment analysis, and automatic summarisation (Jurafsky & Martin, 2020; Diaz-De-Arcaya et al., 2024). In the academic domain, NLP has been applied to automated grading, textual feedback analysis, and intelligent tutoring systems (Adamopoulou & Moussiades, 2020). Beyond conversational AI, automation in higher education increasingly integrates Robotic Process Automation (RPA) and low-code AI tools such as Microsoft Power Automate and Microsoft Copilot Studio (Gos & Spiridon, n.d.), which streamline administrative processes like admissions, registration, and certification management (Chen, 2024). AI-driven automation extends to course recommendation systems, real-time performance monitoring, and adaptive learning platforms that respond dynamically to student progress (Sharma et al., 2023). However, as educational institutions adopt these technologies, they must carefully address challenges related to data privacy, algorithmic bias, and overreliance on automation to ensure ethical and responsible implementation (Husain, 2024; Xing et al., 2025).

The widespread implementation of AI, NLP, and automation in education underscores their transformative potential (Dwivedi et al., 2023). However, successful adoption requires a balance between innovation and ethical responsibility (Francis et al., 2025). Institutions must address challenges related to privacy (Zamora, 2017), transparency (Haenlein & Kaplan, 2019), and overreliance on automation while fostering environments that enhance accessibility and human-centred learning (Følstad & Brandtzaeg, 2020). As AI technologies evolve, their integration into higher education continues to redefine communication, efficiency, and academic support, making continuous evaluation essential to ensure both educational integrity and technological reliability (McGrath et al., 2024).

2.2 Chatbot Implementation in Education: A Systematic Literature Review

"Systematic reviews are a method for identifying and synthesizing all available existing research on a topic and, therefore, are a method to meet the aforementioned research goals" (Scheerder et al., 2017, p. 1609). The present systematic literature review aims to compile and analyse the most relevant studies on the implementation of AI-powered chatbot (Okonkwo & Ade-Ibijola, 2021) in educational context, following the methodologu principles of systematics reviews as outlined by Grant and Booth (2009), to provide a comprehensive overview of their main purposes, design approaches, and observed outcomes. To ensure transparency and replicability, the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) protocol (Moher et al., 2009) was adopted. This protocol guides the review process, from study selection to result presentation, ensuring methodological rigor.

2.2.1 Protocol

The research was conducted using the two academic indexing platforms: Web of Science (WoS) and Scopus, selected for their curated and specialized scientific content, ensuring comprehensive and diverse coverage of peer-reviewed studies relevant to this topic (Martín, 2019).

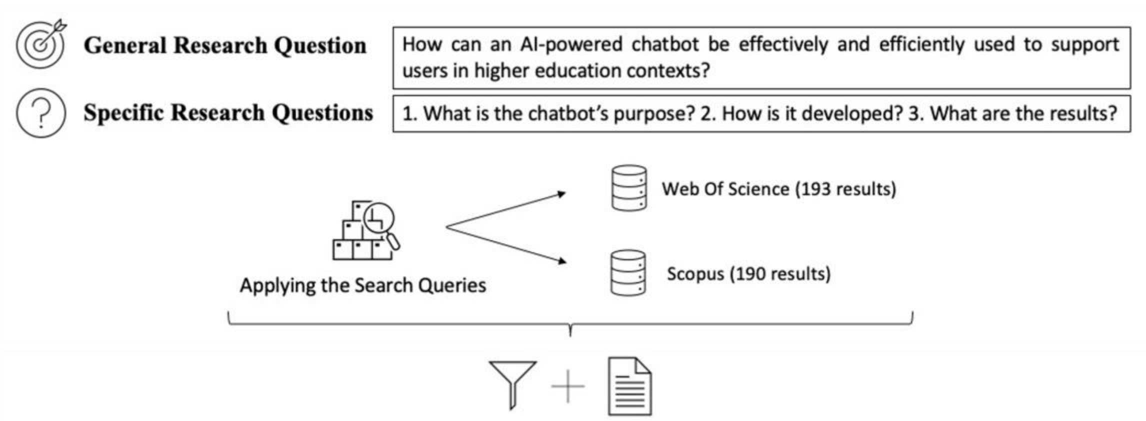
Given the research objective, this systematic literature review aims to address the following main research question: How can an AI-powered chatbot be effectively and efficiently used to support users in higher education contexts? To support this objective, three guiding sub-questions were formulated: 1) “What is the chatbot’s purpose?”, 2) “How is it developed?”, 3) “What are the results?”.

The search strategy was operationalized through a Boolean query designed to retrieve empirical studies reporting on the development, implementation, or evaluation of chatbots in real-world scenarios. The search was conducted in the “Article Title, Abstract, Keywords” fields to ensure broad yet relevant retrieval. The final query was as follows: ((“process automation” OR rpa OR “Artificial Intelligence” OR ai) AND (chatbot* OR “virtual assistant*” OR “conversational agent*” OR “AI assistant*” OR “human-machine interaction” OR “human-computer interaction”) AND (implement* OR “case stud*” OR “use case*” OR deploy* OR “real-world application”) AND (“automated communication” OR “customer support” OR “service automation” OR “interaction automation” OR “student support” OR education OR “learning support” OR “user interaction” OR “automated response”) AND (“dialogue systems” OR “conversation flows” OR “chatbot design” OR “natural language processing”) AND NOT (“systematic review” OR review OR “literature review”)).

The term “chatbot” was prioritized over broader alternatives such as “bot” to exclude unrelated systems (e.g., trading or web bots) that do not involve direct user interaction. The inclusion of terms like “process automation”, “RPA”, and “artificial intelligence” ensured a broad technological scope, while the exclusion of “robotic process automation” maintained the focus on user-facing, interactive solutions rather than back-office automation. Importantly, the search was not limited to a specific domain, enabling the inclusion of chatbot applications from various sectors that may offer relevant insights or be adapted to academic support contexts. To ensure the relevance and adequacy of the search strategy, the Boolean query was validated through a two-step process. First, several pilot searches were conducted in both Scopus and WoS to verify whether the retrieved studies matched the intended scope and to identify redundant or missing keywords. Based on these preliminary results, minor adjustments were made to improve precision and recall. Second, the refined version of the query was reviewed by two domain experts, including the research supervisor, to confirm its adequacy for capturing empirical chatbot studies across different application domains.

Following the database search, the initial pool of records was filtered according to predefined inclusion and exclusion criteria (Moher et al., 2009). First, filters were applied based on document type (peer-
reviewed articles), language (English), subject area (social issues/sciences, computer science, decision sciences, engineering, business, education, and communication), and publication date (from 2022 onward), in line with the recent surge in AI-powered chatbot developments (McGrath et al., 2024). Subsequently, titles and abstracts were screened to exclude studies that did not describe the development, implementation, or evaluation of a chatbot. When the abstract alone did not provide sufficient information to determine eligibility, the introduction and methodology chapters were also examined to verify whether the study involved an implemented chatbot. Articles were excluded if they focused solely on theoretical discussions (e.g. Chandra et al., 2022; Dwivedi et al., 2023), user perceptions (e.g. Arbulú Ballesteros et al., 2024; Su & Yang, 2024), or pedagogical uses without actual system development (e.g., Honig et al., 2024; Karatas et al., 2024), or when the chatbot was not designed to provide information or support to users in a real-world context (e.g., Rijgersberg-Peters et al., 2024). Only empirical studies detailing concrete chatbot applications (e.g., Ait Baha et al., 2024) or evaluations (e.g., Navas et al., 2024) were retained. Finally, duplicate records across databases were removed, following standard procedures commonly adopted in systematic literature reviews (e.g., Moher et al., 2009).

After applying all eligibility criteria, 25 articles were selected for full-text review and included in the final synthesis. The selection process is summarised in Figure 2.1.



Filtering Results, Removing Duplicates and a Quick Reading of the Abstracts with application of Eligibility Criteria:

Eligibility Criteria	Results	
Inclusion Criteria	**Web of Science**	**Scopus**
Publications since 2022	147	129
Articles	94	44
Articles in English	88	40
Subject area: social issues/sciences, computer sciences, engineering, decision sciences, business, education, communication	82	33
Exclusion Criteria	**Web of Science + Scopus**	
Duplicated	15	
Articles that did not specifically addressed the development or the evaluation of a chatbot	75	

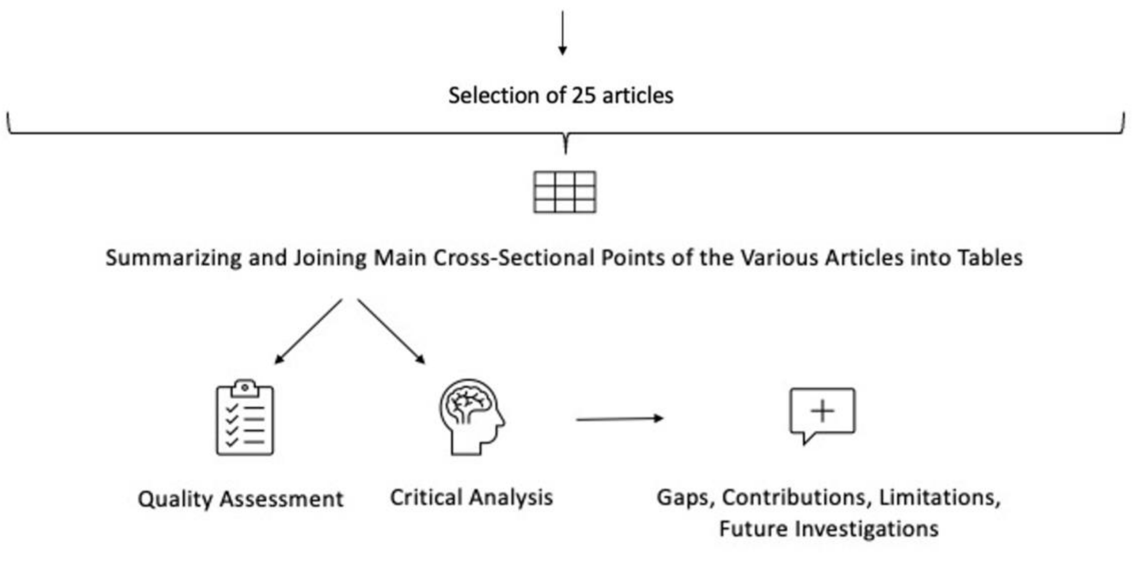


Figure 2.1. Article selection process

Source: Adapted and expanded from Passos (2024).

Table 2.1 presents the 25 articles selected for review, published between 2022 and 2024. A notable increase is observed in 2024, which accounts for more than half of the sample (n=13), reflecting the growing interest in developing and evaluating AI-powered chatbots for real-world contexts. The articles are organised chronologically and alphabetically by title within each year. The articles are organised chronologically and alphabetically by title within each year. Beyond the temporal distribution, the selected studies encompass a variety of scientific domains, predominantly education (n = 17), followed by healthcare (n = 3), customer support (n = 2), and individual cases in computer services (n = 1), technical support (n = 1), and a combined healthcare/education application (n = 1).

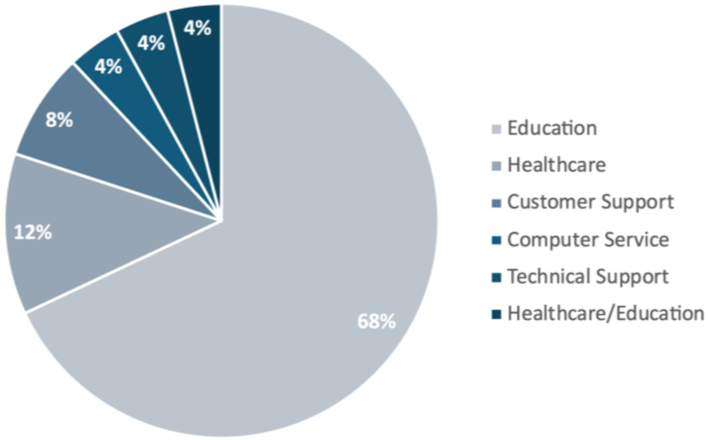


Figure 2.2. Distribution of reviewed studies by application domain

Source: Author's elaboration (2025).

Figure *2.2* illustrates this distribution, highlighting that educational settings represent over two-thirds of the reviewed research. This predominance aligns with the growing academic and institutional interest in leveraging chatbots to enhance communication, support learning, and streamline administrative processes in higher education.

Table 2.1. Articles under study

ID	Year	Title	Journal	Authors
1	2022	A Comparative Analysis of Generative Neural Attention-based Service Chatbot	*International Journal of Advanced Computer Science and Applications*	Suhaili, S.; Salim, N.; Jambli, M.
2	2022	A Novel Framework for Arabic Dialect Chatbot Using Machine Learning	*Computational Intelligence and Neuroscience*	Alhassan, N.; Saad Albarrak, A.; Bhatia, S.; Agarwal, P.
3	2022	An Artificial Intelligence Chatbot for Young People's Sexual and Reproductive Health in India (SnehAI): Instrumental Case Study	*Journal of Medical Internet Research*	Wang, H.; Gupta, S.; Singhal, A.; Muttreja, P.; Singh, S.; Sharma, P.; Piterova, A.
4	2022	Answering Hospital Caregivers' Questions at Any Time: Proof-of-Concept Study of an Artificial Intelligence–Based Chatbot in a French Hospital	*JMIR Human Factors*	Daniel, T.; de Chevigny, A.; Champrigaud, A.; Valette, J.; Sitbon, M.; Jardin, M.; Chevalier, D.; Renet, S.
5	2022	Promoting Students' Learning Achievement and Self-Efficacy: A Mobile Chatbot Approach for Nursing Training	*British Journal of Educational Technology*	Chang, C.; Hwang, G.; Gau, M.
6	2023	An Optimal Deep Feature–Based AI Chat Conversation System for Smart Medical Application	*Personal and Ubiquitous Computing*	Lal, M.; Neduncheliyan, S.
7	2023	Building a Machine Learning Powered Chatbot for KSU Blackboard Users	*International Journal of Advanced Computer Science and Applications*	Alqahtani, Q.; Alrwais, O.
8	2023	Designing and Building OSCEBot ® for Virtual OSCE – Performance Evaluation	*Medical Education Online*	Pereira, D.; Falcão, F.; Nunes, A.; Santos, N.; Costa, P.; Pêgo, J.
9	2023	Effectiveness of an Adaptive Learning Chatbot on Students' Learning Outcomes Based on Learning Styles	*International Journal of Emerging Technologies in Learning (iJET)*	Kaiss, W.; Mansouri, K.; Poirier, F.
10	2023	Introducing a Chatbot to the Web Portal of a Higher Education Institution to Enhance Student Interaction	*Engineering Proceedings*	Oliveira, P.; Matos, P.
11	2023	Lhia: A Smart Chatbot for Breastfeeding Education and Recruitment of Human Milk Donors	*Applied Sciences*	Corrêa, J.; Neto, A.; Pinto, G.; Lima, L.; Teles, A.
12	2023	User Acceptance of a Virtual Librarian Chatbot: an Implementation Method Using IBM Watson Natural Language Processing in Virtual Immersive Environment	*TechTrends*	Safadel, P.; Hwang, S.; Perrin, J.

ID	Year	Title	Journal	Authors
13	2024	A Comprehensive Solution to Retrieval-Based Chatbot Construction	*Computer Speech & Language*	Moore, K.,; Zhong, S.; He, Z.; Rudolf, T.; Fisher, N.; Victor, B.; Jindal, N.
14	2024	Academic Libraries Can Develop AI Chatbots for Virtual Reference Services with Minimal Technical Knowledge and Limited Resources	*Evidence Based Library and Information Practice*	Chase, M.
15	2024	Advanced NLP Models for Technical University Information Chatbots: Development and Comparative Analysis	*IEEE*	Attigeri, G.; Agrawal, A.; Kolekar, S.
16	2024	Advancing Generative Intelligent Tutoring Systems with GPT-4: Design, Evaluation, and a Modular Framework for Future Learning Platforms	*Eletronics*	Liu, S.; Guo, X.; Hu, X.; Zhao, X.
17	2024	AI-Driven Chatbot Implementation for Enhancing Customer Service in Higher Education: A Case Study from Universitas Negeri Semarang	*Journal of Theoretical and Applied Information Technology*	Islam, M.; Warsito, B.; Nurhayati, O.
18	2024	Custom-Trained Large Language Models as Open Educational Resources: An Exploratory Research of a Business Management Educational Chatbot in Croatia and Bosnia and Herzegovina	*Sustainability*	Alfirevic, N.; Pranicevic, D.; Mabic, M.
19	2024	Design of a Large Language Model for Improving Customer Service in Telecom Operators	*Electronics Letters*	Xiaoliang, M.; RuQiang, Z.; Ying, L.; Congjian, D.; Dequan, D.
20	2024	Enhancing Educational Q&A Systems Using a Chaotic Fuzzy Logic-Augmented Large Language Model	*Frontiers in Artificial Intelligence*	Chen, H.; Shi, N.; Chen, L.; Lee, R.
21	2024	Optimizing Automated Conversational Large Language Models for Higher Educational Institution	*International Journal of Computer Information Systems and Industrial Management Applications*	Varaliya, M.; Kanojia, M.; Nabajja, S.
22	2024	Q-Module-Bot: A Generative AI-Based Question and Answer Bot for Module Teaching Support	*IEEE Transactions on Education*	Allen, M.; Naeem, U.; Gill, S.
23	2024	Revolutionizing Campus Communication: NLP-Powered University Chatbots	*International Journal of Advanced Computer Science and Applications*	Ramakrishnan, R.; Thangamuthu, P.; Nguyen, A.; Gao, J.
24	2024	Tayseer: A Novel AI-Powered Arabic Chatbot Framework for Technical and Vocational Student Helpdesk Services and Enhancing Student Interactions	*Applied Sciences*	Alabbas, A.; Alomar, K.
25	2024	The Impact of Educational Chatbot on Student Learning Experience	*Education and Information Technologies*	Ait Baha, T.; El Hajji, M.; Es-Saady, Y.; Fadili, H.

After identifying and selecting the final set of 25 articles, a structured critical analysis was conducted according to the three sub-questions defined in subchapter 2.2.1. This process aimed to systematically examine how each study addressed the chatbot's purpose (Q1), development (Q2), and results (Q3). The analysis was organised into four synthesis tables. The first three correspond to the three main sub-questions: Table *2.3* focuses on the purpose of the chatbots, including the study objective, sector, target audience, and user needs addressed; analyses development aspects, such as framework, use of AI, technology used, integration details, and dialogue design; and the results dimension was divided into two complementary tables, , which summarises quantitative findings (evaluation method, metrics used, and reported results), and Table 2.6. Qualitative evaluation of the chatbots in the selected studies , which compiles qualitative findings (evaluation method, reported feedback, limitations, and perceived impact). Each article was examined in detail to extract partial answers to these sub-questions, ensuring consistency and comparability across studies. Finally, a comprehensive evaluation table () was produced to highlight the overall quality and relevance of each study according to predefined criteria. This systematic and layered approach enabled a critical synthesis of the literature, allowing the identification of key trends, methodological differences, and research gaps.

After reviewing the selected articles, a quality assessment was conducted to evaluate their contribution to the main research question and the three guiding sub-questions. For each sub-question, a set of evaluation criteria was defined and translated into operational questions (). Each article was then scored based on the extent to which it fulfilled each criterion, using a scale of 0 (not addressed), 0.5 (partially addressed), and 1 (fully addressed).

Finally, the synthesis of findings was used to derive the main contributions of the systematic review, as outlined in the final stage of the process (see). These contributions are presented in three dimensions: (i) theoretical, identifying the main research gaps and opportunities for further academic investigation; (ii) practical, outlining implications for institutions and professionals developing or implementing chatbot systems; and (iii) contextual, highlighting the insights most relevant to the present study and guiding the design and evaluation of the educational chatbot developed in this project.

Table 2.2. Quality criteria for article evaluation

Sub-Questions	ID	Quality Criteria
What is the chatbot's purpose?	1.1	Does the article clearly describe the purpose or goal of the chatbot?
	1.2	Does the article specify the sector or context in which the chatbot is applied?
	1.3	Is the target audience identified and considered in the chatbot's design?
How is it developed?	2.1	Does the article mention the platform, framework, or language used to build the chatbot?
	2.2	Does the article describe the use of AI techniques (e.g., NLP, ML)?

Sub-Questions	ID	Quality Criteria
	2.3	Are the conversation flows, or system architecture detailed?
What are the results?	3.1	Are there any evaluation metrics or methods mentioned (e.g., accuracy, response time)?
	3.2	Are the results discussed in a way that shows value or limitations of the chatbot?
	3.3	Does the article reflect on challenges, failures, or areas of improvement?

2.2.2 Critical Synthesis of the Literature

This subchapter presents a structured analysis of the 25 articles selected for this systematic review, guided by the main research question and its three sub-questions. The synthesis is organised into three core dimensions: the purpose of the chatbot, its development process, and the results obtained from its implementation or evaluation.

First, the review explores the intended function and context of use for each chatbot, including the sector, target audience, and specific user needs addressed. The second dimension focuses on how the chatbots were developed, examining the underlying technologies, use of AI, integration methods, and dialogue design strategies. Finally, the third dimension evaluates the reported outcomes, distinguishing between quantitative performance metrics and qualitative insights such as user feedback, perceived limitations, and impact.

All information was systematically extracted and structured into dedicated summary tables, enabling a comparative and critical interpretation of the main patterns, methodological approaches, and contributions identified across the selected studies.

2.2.2.1 Chatbot's Purpose

The selected studies reveal a diverse range of purposes for the implementation of AI-powered chatbots, reflecting the versatility of these systems across sectors and target audiences.

Most articles (n=18) focus on the education sector (e.g., Chang et al., 2022; Ait Baha et al., 2024), especially in higher education settings, where chatbots are designed to support students with academic guidance, administrative information, personalized learning recommendations, and reduction of staff workload (e.g., Alqahtani & Alrwais, 2023; Kaiss et al., 2023; Oliveira et al., 2023). These systems address specific needs such as answering frequently asked questions (e.g., Attigeri et al., 2024; Islam et al., 2024), streamlining information retrieval (e.g., Chase, 2024; Ramakrishnan et al., 2024), or enhancing student engagement and satisfaction (e.g., Oliveira & Matos, 2023; Allen et al., 2024).

A smaller group of studies explores chatbot applications in healthcare (n=4), primarily for patient education (Wang et al., 2022), medical information delivery (Lal & Neduncheliyan, 2023), and support for health professionals (Daniel et al., 2022). These include use cases such as 24/7 access to drug-related information or breastfeeding education (Daniel et al., 2022; Corrêa et al., 2023). One study adopts a hybrid focus between healthcare and education, aiming to provide reproductive health information to youth in a private and accessible manner (Wang et al., 2022).

Chatbots designed for customer or technical support, or customer service are also represented (n=4), addressing needs such as real-time problem resolution (Suhaili et al., 2022; Alhassan et al., 2022), customer interaction automation (Moore et al., 2024), and query handling through natural language interfaces (Xiaoliang et al., 2024). These solutions are often geared toward social media users, telecom customers, or general service platforms.

Table 2.3. Purpose of the chatbots in the selected studies

ID	Study Objective	Sector	Target Audience	User Needs Addressed
1	To develop and evaluate a generative chatbot using RNN with attention for customer support on Twitter	Customer Service	Social media users seeking brand support	Automated responses; natural conversational experience
2	To design and validate a framework for an Arabic-language troubleshooting chatbot	Technical support	Arabic-speaking users	Diagnose and resolve issues in Arabic; reduce need for human agents
3	To explore the use of the SnehAI chatbot for youth health education, applying affordance theory	Healthcare; Education	Indian youth	Access to reproductive health information; youth-friendly and private support
4	To develop and evaluate a chatbot to answer drug- and pharmacy-related questions 24/7	Healthcare	Hospital caregivers (nurses, pharmacy staff)	Access to reliable drug information; reduced staff interruptions; 24/7 availability
5	To evaluate the impact of a mobile-based chatbot on nursing students' knowledge of maternal vaccinations	Education	Nursing students	Interactive learning support; improved knowledge on vaccinations
6	To develop a hybrid LbDBC for accurate medical query responses	Healthcare	Patients	Access to accurate medical information; automated support
7	To develop a machine learning chatbot for answering Blackboard FAQs at a university	Education	University students	Learning support; reduced staff workload
8	To improve the chatbot's ability to understand student queries and deliver accurate responses	Education	Higher education students	Clarification of academic questions; enhanced learning support
9	To develop a chatbot that recommends learning objects based on students' learning styles	Education	University students	Personalized learning recommendations; adaptive support; improved outcomes
10	To implement a chatbot on a university web portal to enhance student interaction and access to academic and administrative services	Education	Higher education students	Access to academic information; administrative assistance; personalized guidance
11	To co-develop a chatbot for breastfeeding education and milk donor recruitment	Healthcare	Mothers and potential milk donors	Breastfeeding guidance; donor recruitment; educational support
12	To develop a chatbot-based system for recommending learning materials based on students' learning styles	Education	Students	Personalized learning support
13	To develop a complete pipeline for deploying a real-life chatbot from unlabelled chat log data	Customer Support	End-users of online services	Scalable chatbot deployment; improved automation of customer service
14	To describe the development of an AI chatbot for virtual reference services in an academic library	Education	Library users (students, faculty, staff)	Access to library information; support outside reference hours
15	To implement a chatbot using various NLP models for resolving student queries during admissions	Education	Prospective students	Admission-related support; automated query resolution

ID	Study Objective	Sector	Target Audience	User Needs Addressed
16	To design a generative ITS framework using GPT-4 for personalized educational support	Education	Students	Personalized guidance; academic assistance
17	To develop an AI chatbot for automating responses to student queries in higher education	Education	Students	Instant query resolution; improved communication; reduced workload
18	To explore how custom-trained LLMs can function as Open Educational Resources in higher education	Education	Students	Access to personalized educational content; autonomous learning
19	To improve telecom customer service using LLMs integrated with LangChain for enhanced knowledge recommendation and data privacy	Customer Support	Telecom customers	Faster query resolution; accurate information; privacy protection
20	To develop CHAQS, a custom LLM-based Q&A chatbot for enhanced educational interactions	Education	Students	Improved learning support; enhanced question answering
21	To improve query systems in higher education using LLMs and institutional knowledge bases	Education	Students and academic staff	Accurate academic query responses; use of institutional knowledge
22	To develop and assess a chatbot for assisting university students with module-related queries	Education	University students	Quick access to module information; enhanced satisfaction; reduced workload
23	To implement an NLP-based chatbot to answer university FAQs and streamline information retrieval	Education	Students	Access to academic and administrative information; simplified navigation
24	To evaluate the performance of four transformer-based models in enhancing chatbot understanding of user queries	Education	Students	Improved query interpretation; enhanced response accuracy
25	To design and evaluate a chatbot-based system to support students' learning and reduce teacher workload	Education	Students	Learning support; reduced teacher workload

Notes: AI: Artificial Intelligence; FAQs: Frequently Asked Questions; ITS: Intelligent Tutoring System; LLMs: Large Language Models; LMS: Learning Management System; LbDBC: Lion-based Deep Belief Chatbot; NLP: Natural Language Processing; Q&A: Question and Answer; RNN: Recurrent Neural Network.

In terms of user groups, the majority of chatbots target students in academic environments (e.g. Alqahtani & Alrwais, 2023; Oliveira et al., 2023), whether for general assistance (e.g., Attigeri et al., 2024; Ramakrishnan et al., 2024), learning support (e.g., Kaiss et al., 2023; Ait Baha et al., 2024), or personalized guidance (e.g. Oliveira & Matos, 2023; Chase, 2024). Other target audiences include healthcare workers (e.g. Daniel et al., 2022; Corrêa et al., 2023), patients, youth (e.g. Wang et al., 2022; Lal & Neduncheliyan, 2023), and customers of online (e.g. Moore et al., 2024; Suhaili et al., 2022) or telecom services (e.g. Moore et al., 2024; Suhaili et al., 2022). Across all domains, the user needs addressed converge around goals such as information access (e.g. Kaiss et al., 2023; Daniel et al., 2022), task automation (e.g. Alqahtani & Alrwais, 2023; Moore et al., 2024), personalized assistance (e.g. Oliveira & Matos, 2023; Kaiss et al., 2023), and reduction of human workload (e.g. Oliveira et al., 2023; Ramakrishnan et al., 2024). The variety of objectives and audiences reinforces the adaptability of chatbot solution to meet context-specific demands, illustrating their potential to improve user experiences in both educational and service-oriented contexts.

2.2.2.2 Developing Process

The analysis of chatbot development strategies across the 25 selected studies reveals a wide diversity of technical approaches and implementation choices, reflecting the varying objectives, domains, and user needs of each project.

In terms of AI usage, while a small number of chatbots employed generative approaches, such as Generative Pre-trained Transformers (GPT) or Recurrent Neural Networks (RNNs) with attention mechanisms (e.g., Chen et al., 2024; Oliveira & Matos, 2023; Suhaili et al., 2022), most systems relied on intent classification (e.g., Kaiss et al., 2023; Alqahtani & Alrwais, 2023), entity recognition (e.g., Attigeri et al., 2024; Ramakrishnan et al., 2024), or response retrieval models (e.g., Moore et al., 2024). Several studies used pretrained models like Bidirectional Encoder Representations from Transformers (BERT) (e.g., Pereira et al., 2023; Ramakrishnan et al., 2024; Varaliya et al., 2024), whereas others developed custom rule-based systems without any AI components (e.g., Allen et al., 2024; Corrêa et al., 2023). Rule-based architectures continue to be widely adopted, especially in use cases involving well-defined information delivery or administrative assistance (Alabbas & Alomar, 2024; Alhassan et al., 2022; Chang et al., 2022). These systems often follow decision-tree structures with predefined question-answer pairs or keyword-triggered flows, ensuring high control and low ambiguity in the responses. On the other hand, hybrid systems are also common, combining rule-based logic with statistical or ML techniques (e.g., Chase, 2024; Varaliya et al., 2024), allowing greater flexibility while maintaining predictable behaviour. Frameworks such as Rasa (Alqahtani & Alrwais, 2023), Dialogflow (Chase, 2024; Kaiss et al., 2023), and LangChain (Oliveira & Matos, 2023; Xiaoliang et al., 2024) were frequently employed, either to handle intent recognition (e.g., Alqahtani & Alrwais, 2023; Kaiss et al., 2023; Chase, 2024; Attigeri et al., 2024) or to connect LLMs to structured data sources (e.g., Oliveira & Matos, 2023; Xiaoliang et al., 2024; Varaliya et al., 2024). A few studies also made use of lower-code or API-based integrations, such as IBM Watson (Daniel et al., 2022), showing that technical complexity varies significantly depending on the development context and available resources.

Regarding integration, many chatbots were embedded within existing digital platforms, for example, Moodle or Blackboard (Alfirevic et al., 2024; Alqahtani & Alrwais, 2023; Chang et al., 2022), enabling direct access for students or educators within their usual learning environments. Other systems were deployed as standalone web applications (Ait Baha et al., 2024; Attigeri et al., 2024; Moore et al., 2024) or integrated with messaging platforms such as Telegram (Alabbas & Alomar, 2024). Mobile app distribution (Wang et al., 2022) and voice interfaces (Corrêa et al., 2023) were mentioned less frequently, indicating a general preference for browser-based or institutional interfaces. As for dialogue design, most systems followed intent-based approaches, often relying on training phrases and rule-driven replies (e.g., Chase, 2024; Kaiss et al., 2023). A smaller number employed more advanced dialogue features, such as multi-turn interaction (Liu et al., 2024; Xiaoliang et al., 2024), generative outputs for low-similarity queries (Islam et al., 2024), or didactic conversation flows adapted to learning content (Alfirevic et al., 2024). Despite the rise of generative AI, few systems implemented open-ended conversation or conversational memory, possibly due to concerns related to control, relevance, or resource constraints. Although many of the reviewed studies adopt either rule-based systems or fully coded AI architectures, there is growing potential to bridge low-code development with more advanced AI techniques. Some platforms, such as Dialogflow (Chase, 2024; Kaiss et al., 2023) and IBM Watson (Daniel et al., 2022), provide visual interfaces or rule-based frameworks, while also supporting the integration of pretrained NLP models, fuzzy logic, or external APIs. These hybrid configurations reveal that low-code approaches are not inherently limited to simple logic, but can serve as accessible gateways to advanced AI systems.

As summarised in Table *2.4*, the reviewed studies reflect a broad spectrum of chatbot development approaches, from simple rule-based flows to LLM-integrated systems. Although generative and deep learning methods are gaining attention, rule-based and hybrid approaches remain predominant, particularly in educational and institutional contexts where clarity, control, and ease of integration are essential.

Table 2.4. Development characteristics of the chatbots in the selected studies

ID	Framework	Use of AI	Technology Used	Integration Details	Dialogue Design
1	Seq2seq (deep learning)	Generative AI (RNN)	Python; TensorFlow; Twitter API	Integrated with Twitter for customer support	One-turn Q&A; not context-aware
2	Rule-based with predefined flows	No AI used	JavaScript; Node.js; Dialogflow; MongoDB	Web-based with admin dashboard	Decision tree; rule-based
3	Not specified	Not specified	Android; SnehAI app	Mobile app distribution; interaction via app UI	Educational/motivational; rule-based
4	ADDIE model	NLP with fuzzy matching (IBM Watson)	Node.js; TypeScript; IBM Watson API	Hospital intranet; limited integrations	Guided topics; autocomplete
5	LINE Bot + Moodle	Rule-based	Moodle; LINE API; MySQL	Mobile chatbot via LINE API	Q&A pairs; structured flow
6	LbDBC	Deep Belief Network + Lion Algorithm	Python; NLTK	Answer retrieval using root-word matching	Rule-based; single-turn QA
7	Rasa Core + Rasa NLU	ML (intent, entity, dialogue)	Python; Docker; YAML	Integrated with KSU Blackboard	Stories, fallback policy
8	Custom architecture	NLP using BERT	Python; Transformers; Scikit-learn	Academic chatbot prototype	Intent classification only
9	Dialogflow	NLP + NLU	Node.js; Moodle	Moodle widget via HTML block	Entity-based responses; webhook

ID	Framework	Use of AI	Technology Used	Integration Details	Dialogue Design
10	Streamlit + LangChain	LLMs + NLP + ML	OpenAI; Pinecone; Python	Educational portal integration	One-turn Q&A; document retrieval
11	Not specified	No AI used	Node.js; Web Speech; Dialogflow	Web voice/text integration	Predefined intents; voice UI
12	Not specified	ML + Semantic Reasoning	Ontology tools	E-learning personalization	Rule-based; learning style
13	Data programming pipeline	Supervised learning	Snorkel; Python; Rasa	Real-world support chatbot	FAQ-style; retrieval-based
14	Dialogflow	NLP	Kommunicate; Google Analytics	Library chatbot widget	Predefined chips; training phrases
15	Custom rule-based	NLP	Python; Google Cloud	Web-based for admissions	One-turn; no memory
16	Generative ITS + GPT-4	Generative LLM	OpenAI GPT-4; Python	Integrated into an ITS framework	Multi-turn; tutoring interaction
17	Hybrid (retrieval-based + generative models)	ML + NLP	TF-IDF; LLaMA RAG; BERT; Python	UNNES helpdesk	FAQ; generative for edge cases
18	Custom LLM in OER system	Custom-trained LLMs	Python	Moodle and digital repository	Educational dialogue
19	LangChain with K-Modules	ChatGLM2-6B	Python; APIs	Integrated with telecom service	Query-to-vector; private output
20	CHAQS	ChatGLM23-6B with fuzzy logic	BLEU; BERTScore	Internal dataset; no APIs	Variable-length contextual output
21	LlamaIndex with LLM integration	Generative AI	Python; OpenAI API; Google Gemini API	Connected to institutional knowledge base via embeddings	Retrieval-based and prompt-engineered responses
22	Co-designed rule-based	No AI used	JavaScript	Youth mental health website	Scripted paths; emotional tone focus
23	FAQ chatbot with BERT + NN	BERT	Python; NLTK	University platform (Flask)	Context-aware; JSON-triggered
24	RASA framework	Deep learning	Python, RASA NLU/Core	Web-based integration	FAQ
25	Modular LSTM chatbot	Deep learning (LSTM)	TensorFlow; Flask	Flask app with local DB	LSTM-triggered; no context

Notes: ADDIE: Analysis, Design, Development, Implementation, Evaluation; AI: Artificial Intelligence; API: Application Programming Interface; BERT: Bidirectional Encoder Representations from Transformers; BLEU: Bilingual Evaluation Understudy; CHAQS: Chaotic LLM-based Educational Q&A System; DB: Database; FAQ: Frequently Asked Question; IBM: International Business Machines Corporation; ITS: Intelligent Tutoring System; KSU: King Saud University; LbDBC: Hybrid Lion-based Deep Belief Chatbot; LLaMA: Large Language Model Meta AI; LLMs: Large Language Models; LSTM: Long Short-Term Memory; ML: Machine Learning; NLTK: Natural Language Toolkit; NLU: Natural Language Understanding; NN: Neural Network; OER: Open Educational Resources; Q&A: Question and Answer; QA: Question Answering; RAG: Retrieval-Augmented Generation; RNN: Recurrent Neural Network; Seq2seq: sequence-to-sequence; TF-IDF: Term Frequency - Inverse Document Frequency; UI: User Interface.

2.2.2.3 Results

The evaluation of chatbot performance in the selected studies reveals a broad range of quantitative methods and metrics. As shown in Table *2.5*, several studies employed automated evaluations using established NLP metrics such as Bilingual Evaluation Understudy (BLEU) score, BERTScore, accuracy, precision, recall, and F1-score. These studies (e.g., Ait Baha et al., 2024; Lal & Neduncheliyan, 2023; Suhaili et al., 2022) often aimed to measure the linguistic quality, classification performance, or overall effectiveness of the chatbot in handling user input. Some chatbots were assessed through controlled user testing scenarios, involving students or target users who completed predefined tasks. For example, Chang et al. (2022) measured effectiveness, efficiency, and satisfaction among 38 students, while Daniel et al. (2022) used structured questionnaires to capture metrics like speed, usability, appearance, and relevance. Liu et al. (2024) compared two chatbot versions, showing that the AI-enhanced version achieved higher engagement and task accuracy. A few studies employed comparative testing against baselines or between different chatbot configurations. Alhassan et al. (2022), Attigeri et al. (2024) and Chen et al. (2024) report that the proposed solutions outperformed their baselines in accuracy and NLP scores. Similiarly, Islam et al. (2024) compared multiple models and concluded that TF-IDF achieved the highest performance, although BERT-based approaches underperformed in specific contexts.

Despite the emphasis on quantitative validation in some studies, a significant number of articles (e.g., Alfirevic et al., 2024; Oliveira & Matos, 2023; Wang et al., 2022) did not report any formal quantitative metrics, relying instead on anecdotal results (e.g., Safadel et al., 2023; Chase, 2024), internal validation (e.g., Varaliya et al., 2024; Xiaoliang et al., 2024), or qualitative impressions (e.g., Wang et al., 2022; Corrêa et al., 2023). This lack of standardised reporting limits the comparability across cases and highlights an area for methodological improvement in future implementations.

Overall, while the most rigorous studies employed multi-metric evaluations combining accuracy, satisfaction, and linguistic quality, the field remains heterogeneous in terms of evaluation depth and consistency.

Table 2.5. Quantitative evaluation of the chatbots in the selected studies

ID	Evaluation Method	Metrics Used	Reported Results
1	Automated evaluation	BLEU score	BLEU score: 57.78%
2	User testing (n = 30) comparing performance with baseline Arabic chatbots	Accuracy; response time; relevance	Outperformed existing Arabic chatbots in accuracy and response time; improved understanding of Arabic input
3	No formal quantitative evaluation	N/A	No metrics or numerical results reported
4	Beta testing (n=20) with questionnaire and back-office analytics	Speed (1–10); Usability (1–10); Appearance (1–10); Satisfaction (Likert scale); Relevance rate (feedback ratio)	Speed: 8.2/10; Usability: 8.1/10; Appearance: 7.5/10; Satisfaction: 70%; Relevance: 76% of completed conversations
5	User testing (n = 38 students)	Effectiveness, Efficiency, Satisfaction (via task completion and questionnaire)	Effectiveness: 82.5%; Efficiency: 88.4%; Satisfaction: 80.7%
6	Experimental comparison using a medical QA dataset	Accuracy, Precision, Recall, F-score, Error rate	Accuracy: 99.55%, Precision: 96.42%, Recall: 99.42%, F-score: 97.90%, Error rate: 0.42%

ID	Evaluation Method	Metrics Used	Reported Results
7	Cross-validation (5 folds); Dialog story testing; Confusion matrix	Accuracy, F1-score, Precision (Train/Test), Intent prediction confidence	Post-tuning: Accuracy 93.4%, F1-score 92.5%, Precision 92.5%. High alignment between predicted and true intents; dialog accuracy: 99.1%
8	Exploratory pilot testing (n = 13)	Usability (Likert scale); Comprehension; Satisfaction	Overall positive scores; users reported ease of use, helpfulness, and adequate answer clarity. No numerical metrics reported.
9	Pre-test / post-test comparative analysis	% distribution across knowledge levels (Beginner, Intermediate, Advanced)	Beginner level reduced from 61% to 45%; Intermediate increased from 39% to 49%; Advanced rose from 0% to 6%
10	No formal quantitative evaluation	N/A	No metrics or numerical results reported
11	No formal quantitative evaluation	N/A	No metrics or numerical results reported
12	User evaluation with 21 students	Satisfaction (Likert scale)	Majority strongly agreed chatbot was useful and easy to use; positive sentiment overall. No numerical scores reported
13	User testing with a structured questionnaire	Likert-scale ratings: helpfulness, understanding, perceived accuracy	83% agreed or strongly agreed the chatbot was helpful; 80% said it improved understanding; 73% found it accurate
14	Analysis of usage logs from Dialogflow and Kommunicate	Number of interactions per month	Usage increased from 44 to 137 monthly interactions over 18 months; no performance or satisfaction metrics reported
15	Comparative testing with prototype and Dialogflow baseline	BLEU score; BERTScore; classification accuracy	BLEU: 0.89 (CHAQS) vs. 0.70 (baseline); BERTScore: 0.93 vs. 0.76; Accuracy: 94.3% vs. 87.1%
16	Experiment with two student groups using different chatbot versions	Engagement rate; task accuracy; average interaction time	AI-enhanced version had higher engagement (61% vs. 41%), better task accuracy (82% vs. 69%), and longer average interaction time (9.4 vs. 6.8 minutes)
17	Comparative evaluation of multiple models	Precision, Recall, F1-score, Accuracy, BLEU score	TF-IDF outperformed others with 78% accuracy; BLEU for generative model reached 0.71; BERT-based models underperformed in context-specific responses
18	No formal quantitative evaluation	N/A	No metrics or numerical results reported
19	Reinforcement Learning evaluation using domain experts over multiple training iterations	MSE; Knowledge Recommendation Acceptance Rate	Acceptance rate increased from 15% to 74.8% in vertical industry; MSE loss stabilized after training
20	Comparative evaluation using two question datasets (short vs. long)	BLEU score; BERTScore; classification accuracy	CHAQS outperformed baseline: BLEU score 0.89 vs. 0.70; BERTScore 0.93 vs. 0.76; Accuracy 94.3% vs. 87.1%
21	Manual evaluation of chatbot response accuracy	Accuracy; Response Relevance	Reported high accuracy and contextual precision in generated answers
22	No formal quantitative evaluation	N/A	No metrics or numerical results reported
23	Comparative model testing	Accuracy; Training & Validation Loss	Neural Network: 98% accuracy; BERT: 87% accuracy; BERT showed overfitting with higher validation loss (5.04)
24	No formal quantitative evaluation	N/A	No metrics or numerical results reported
25	Classification model evaluation	Accuracy, Precision, Recall, F1-score	Accuracy: 95.6%; Precision: 96.2%; Recall: 94.8%; F1-score: 95.5%

Notes: BERT: Bidirectional Encoder Representations from Transformers; CHAQS: Chaotic LLM-based Educational Q&A System; MSE: Mean Squared Error; N/A: Not applicable; TF-IDF: Term Frequency - Inverse Document Frequency.

In addition to quantitative metrics, several studies reported qualitative feedback to capture users' perceptions, limitations, and the perceived value of the implemented chatbots. As summarised in

Table 2.6, common evaluation methods included user satisfaction questionnaires, structured interviews, and open-ended feedback collected during pilot testing or deployment. User feedback was generally positive across most implementations. For instance, Moore et al. (2024), over 80% of users felt the chatbot was helpful, improved their understanding, and delivered accurate responses. Similarly, Chang et al. (2022), Daniel et al. (2022) and Pereira et al. (2023) reported favourable impressions regarding usability, usefulness, and response clarity, though they often lacked detailed scoring breakdowns. Some chatbots were praised for their educational role (Alfirevic et al., 2024; Kaiss et al., 2023), while others received acknowledgment for ease of integration into institutional platforms (e.g., Alabbas & Alomar, 2024; Chang et al., 2022). However, few studies provided structured discussions of limitations. Where mentioned, limitations included the chatbot's lack of contextual memory (Attigeri et al., 2024), overfitting of certain models (Ramakrishnan et al., 2024), or restricted scope of training data (Islam et al., 2024). Some rule-based systems (Alhassan et al., 2022; Allen et al., 2024; Corrêa et al., 2023) were flagged as inflexible when dealing with nuanced or unexpected inputs. The absence of multi-turn conversation or dialogue personalisation was another recurring concern. As for the perceived impact, various studies reported gains in student engagement, knowledge acquisition, and support efficiency. Kaiss et al. (2023), for example, found that the chatbot enabled more personalised resource recommendations. Liu et al. (2024) and Varaliya et al. (2024) highlighted improved student support outcomes, particularly in task completion and career guidance. In several cases, the chatbot was seen not only as a tool for automating responses but also as a facilitator of learning, onboarding, or mental health support (e.g., Allen et al., 2024; Moore et al., 2024).

Overall, while the qualitative evaluations are less standardised and often anecdotal, they provide important context for understanding how chatbots are received and the kinds of improvements that users value. The findings suggest that perceived usefulness, ease of use, and alignment with user expectations are key to chatbot success, regardless of technical complexity.

Table 2.6. Qualitative evaluation of the chatbots in the selected studies

ID	Evaluation Method	Reported Feedback	Limitations	Perceived Impact
1	Qualitative analysis of generated responses (informational vs emotional queries)	Informational responses were more accurate and fluent; emotional responses lacked depth and empathy	Limited linguistic variety; absence of true emotional understanding	More accurate and relevant responses; better handling of Arabic terminology; limitations in speed and flexibility
2	Informal qualitative assessment; performance review on user queries and internal testing	More accurate and relevant responses; better handling of Arabic terminology; limitations in speed and flexibility	Lack of comprehensive Arabic language libraries; limited response flexibility; underperformance in complex queries	Enhanced access to Arabic-language technical support; foundation for further Arabic NLP development
3	Pilot testing and design workshops (informal)	Chatbot perceived as engaging, relatable, and respectful; good use of culturally relevant language and tone	Not evaluated in real-world use; feedback limited to pilot and controlled settings	Promising tool for digital health education; supports SRHR awareness among youth in India
4	User testing with feedback form and internal observation	Perceived as fast, user-friendly, and visually acceptable; useful for accessing validated drug info 24/7	Small sample size (n=20); limited scope of topics; not integrated with hospital software; usability issues with drug stability info	Effective tool for supporting caregivers with drug-related questions; promising for broader hospital use and onboarding/training
5	Post-task questionnaire and user comments	Users found the chatbot helpful and easy to use; improved engagement and task clarity	Focused only on LINE Bot and Moodle; limited generalizability; rule-based limitations	Enhanced learning experience; supports autonomous learning and student motivation
6	Case study; comparative evaluation with baseline models	The chatbot produced more accurate and relevant answers than existing models; demonstrated strong performance across all evaluated metrics	No discussion of practical deployment limitations; no user validation was conducted	Significant improvement in medical QA response accuracy; high potential for integration in digital health systems
7	Developer-driven evaluation using test stories and error inspection	Effective in understanding and responding to Blackboard-related queries; improved user experience with fewer repeated queries	Limited to Arabic FAQ context; chatbot lacks generative capabilities; needs continuous dataset updates	Reduction in helpdesk workload; improved service for KSU students; strong potential for expansion and optimization
8	User testing via questionnaire and open feedback	Participants found the chatbot clear, easy to use, and useful for navigating master's program information	Small sample size (n = 13); lack of long-term or real-world testing	Useful tool for onboarding and information dissemination; promising support for prospective students
9	User satisfaction questionnaire (2 questions)	91% found the chatbot experience very interesting; 73% found the recommendations helpful	Not explicitly discussed in-depth, but potential cold-start mitigation is limited to learning style (may not generalize)	Positive user perception; improved learning experience through personalized recommendations based on learning style
10	Internal performance and deployment review (no user testing)	Chatbot provided instant responses, personalized support, and improved service access	Requires constant input updates; sensitive to document quality; no user evaluation reported	Improved student access to information; enhanced engagement via web portal; scalable for future refinements
11	Exploratory prototyping; informal design assessment	Chatbot seen as a strategy to promote AI understanding and support contextual learning	No empirical testing or user validation; remains a proof of concept	Demonstrates educational potential for teaching AI and contextualizing curriculum content

ID	Evaluation Method	Reported Feedback	Limitations	Perceived Impact
12	Post-interaction questionnaire and discussion	Students found the chatbot helpful for clarifying course content and improving engagement	Only 21 participants; tested in a limited academic setting; lacks performance benchmarking	Enhanced understanding of material; supports active learning in online courses
13	Post-use survey with open-ended responses	Users appreciated clarity, relevance of responses, and conversational tone. Some noted lack of personalization	Limited sample size; pilot context; no integration into real-world deployment	Demonstrated strong potential as a student-facing educational tool; improved clarity and engagement with course content
14	Observational analysis of usage trends and functionality	Users most frequently searched for opening hours, resources, and borrowing policies; low engagement beyond initial prompt	Only 10% of users engaged beyond welcome message; no user demographics or satisfaction data collected	Demonstrated feasibility of low-resource chatbot development; informed future research on engagement strategies
15	Informal analysis of response variation across question types (short/medium/long)	Chatbot generated more coherent and contextually relevant responses than baseline system	Evaluation limited to predefined dataset; no real user testing; applicability to broader domains untested	Demonstrated improved answer relevance and contextuality; effective for varied question lengths
16	Post-task questionnaire and interviews	Students felt the AI-enhanced chatbot offered more relevant feedback and personalized guidance	Conducted in controlled environment; limited generalizability; short-term evaluation only	Demonstrated potential to support self-regulated learning; improved feedback relevance and learner motivation
17	Manual analysis of chatbot responses and system performance observations	System handled routine queries effectively; generative component improved coverage of complex or ambiguous inputs	No real-world deployment; limited to predefined dataset; no user feedback collected	Demonstrated hybrid model viability; offers adaptable QA system for academic helpdesk environments
18	Reflective conceptual analysis of chatbot use in higher education	Chatbots can support internal efficiency and improve communication with students; proposed cautiously as part of broader digital transformation	No empirical data; lacks testing or user feedback; based on theoretical exploration only	Highlights strategic potential of chatbots in education; calls for future research and policy development
19	Continuous user testing and agent feedback during real-world deployment	Improved response accuracy and efficiency; greater alignment with telecom-specific queries	Complexity of integration with legacy systems; high computational resource requirements	Enhanced service precision and security; reduced agent workload; increased user satisfaction
20	Informal analysis of output variation across query lengths	More coherent and diverse responses; improved contextual relevance compared to baseline	No user study; only internal datasets; not tested across domains	Shows effectiveness of fine-tuned LLM with fuzzy logic for educational and informational queries
21	Not formally reported; analysis inferred from demonstration	Chatbot provided relevant responses using institutional data; improved information accessibility	Limited real-world testing; no user evaluation conducted	Demonstrated feasibility of LLM-based chatbots for academic support contexts
22	Participatory design process and iterative co-creation feedback	Chatbot was well-received by youth; valued for inclusive tone, accessibility, and emotional alignment	No large-scale testing; unknown long-term impact; no empirical evaluation post-deployment	Strengthened user engagement in mental health contexts; model for inclusive, co-designed digital tools

ID	Evaluation Method	Reported Feedback	Limitations	Perceived Impact
23	Internal performance review; comparison with university FAQ access	Neural network provided accurate and context-aware answers; better suited to specific FAQ intents	BERT model had lower performance due to small and domain-specific dataset; risk of overfitting	Improved student access to accurate university information; time-saving for administrative staff
24	Informal internal testing; functionality review	Chatbot provided relevant responses to admission-related queries; efficient integration with Telegram	Rule-based structure limited flexibility; no NLP or learning capabilities; no user testing conducted	Demonstrated potential for automating FAQ handling during university admissions; reduced staff workload
25	System behavior observation and performance validation	Intent-based responses were accurate and timely; suitable for FAQ-type automation	No user-facing testing; single-turn interaction only; no learning or adaptation over time	Effective for automating routine academic inquiries; scalable in institutional settings

Notes: BERT: Bidirectional Encoder Representations from Transformers; FAQ: Frequently Asked Questions; KSU: King Saud University; LLM: Large Language Model; NLP: Natural Language Processing; QA: Question Answering; SRHR: Sexual and Reproductive Health and Rights.

2.2.3 Evaluations of Relevant Scientific Articles

As described in subchapter 2.2.1, the 25 selected articles were assessed using nine evaluation criteria, as summarized in Table *2.7*, grouped under the three main research questions and for each criterion, a score of 0 (not addressed), 0.5 (partially addressed), or 1 (fully addressed) was assigned, based on the extent to which the study provided relevant and sufficiently explicit information.

The quality assessment results (Table *2.7*) show that none of the reviewed articles achieved the maximum possible score of 9. However, several studies (e.g., Ait Baha et al., 2024; Lal & Neduncheliyan, 2023; Moore et al., 2024; Suhaili et al., 2022) reached the highest observed score of 9, indicating a consistent and comprehensive coverage of the research questions. These studies provided clear objectives, robust development descriptions, and well-documented results, often supported by quantitative evaluation metrics. Among the three dimensions, the criteria related to the chatbot's purpose (Q1.1–Q1.3) scored uniformly high, with all 25 studies obtaining the maximum points. This suggests that the reviewed literature consistently presents the motivation and intended scope of the chatbot in a clear and explicit way. In contrast, the development-related criteria (Q2.1–Q2.3) showed slightly more variation. While Q2.1 (framework or architecture description) and Q2.3 (dialogue design) were generally well addressed, Q2.2 (use of AI) had lower coverage, with only 19 studies fully detailing the AI techniques employed (e.g., Attigeri et al., 2024; Varaliya et al., 2024). This aligns with earlier observations that a significant portion of the literature still relies on rule-based (e.g., Chang et al., 2022; Alabbas & Alomar, 2024) or hybrid systems (e.g., Chase, 2024) without advanced AI integration (e.g., Oliveira & Matos, 2023; Alfirevic et al., 2024). The results-related criteria (Q3.1–Q3.3) presented the greatest variation. While 20 articles fully described the evaluation methods applied (Q3.1) (e.g., Daniel et al., 2022), only 23 discussed the value or limitations of the chatbot (Q3.2) in sufficient depth (e.g., Ait Baha et al., 2024; Wang et al., 2022), and 22 reflected explicitly on challenges (e.g., Corrêa et al., 2023; Liu et al., 2024), failures (e.g., Islam et al., 2024; Pereira et al., 2023), or areas for improvement (e.g., Allen et al., 2024) (Q3.3). These findings suggest that, although evaluation is common, a smaller subset of studies provides the critical reflection necessary for identifying avenues for future research and improvement (e.g., Ait Baha et al., 2024; Corrêa et al., 2023; Islam et al., 2024; Liu et al., 2024).

Overall, the distribution of scores highlights a strong focus on describing chatbot purposes and general architectures, but a relatively lower emphasis on detailed AI usage and comprehensive critical reflection on results. This indicates an opportunity for future research to strengthen methodological transparency and explicitly address limitations, ensuring that findings can be effectively replicated, compared, and built upon.

Table 2.7. Evaluation of the quality of articles in RSL

ID	What is the chatbot's purpose?			How is it developed?			What are the results?			Total
	Q1.1.	Q1.2.	Q1.3.	Q2.1.	Q2.2.	Q2.3.	Q3.1	Q3.2	Q3.3	-
1	1	1	1	1	1	1	1	1	1	9
2	1	1	1	1	0	1	1	1	1	8
3	1	1	1	1	0	0.5	0.5	0.5	0.5	6
4	1	1	1	1	1	1	0.5	1	1	8.5
5	1	1	1	1	0	1	0	0.5	0.5	6
6	1	1	1	1	1	1	1	1	1	9
7	1	1	1	1	1	1	1	1	1	9
8	1	1	1	1	1	1	1	1	1	9
9	1	1	1	1	1	1	1	1	1	9
10	1	1	1	1	1	1	0.5	0.5	0	7
11	1	1	1	1	0	1	0.5	1	1	7.5
12	1	1	1	1	1	1	0.5	1	1	8.5
13	1	1	1	1	1	1	1	1	1	9
14	1	1	1	1	1	1	0.5	0.5	0.5	7.5
15	1	1	1	1	1	0.5	1	1	1	8.5
16	1	1	1	1	1	1	1	1	1	9
17	1	1	1	1	1	1	1	1	1	9
18	1	1	1	1	1	1	1	1	1	9
19	1	1	1	1	1	1	1	1	1	9
20	1	1	1	1	1	1	1	1	1	9
21	1	1	1	1	1	1	1	1	1	9
22	1	1	1	1	0	1	0.5	1	1	7.5
23	1	1	1	1	1	1	1	1	1	9
24	1	1	1	1	0	1	0.5	1	0.5	7
25	1	1	1	1	1	1	1	1	1	9
Total	25	25	25	25	19	24	20	23	22	-

2.2.4 Summary and Identification of Research Gaps

This literature review examined 25 studies on AI-powered chatbots implemented in real-world contexts, focusing on three main dimensions: the chatbot's purpose, how it is developed, and the results achieved. The analysis revealed a wide variety of technical architectures, from simple rule-based systems to hybrid and LLM-enhanced approaches, reflecting different priorities regarding control, scalability, and contextual adaptability. From a technological perspective, rule-based and hybrid models remain predominant, especially in educational and institutional settings, where predictability, control over responses, and ease of integration are crucial. However, despite the increasing availability of advanced AI models, few studies fully explore their potential for open-ended, multi-turn interactions or contextual memory, capabilities that could significantly enhance user engagement. Nonetheless, the rapid adoption of LLMs in the past two years suggests a gradual shift in this trend, as researchers begin to explore their integration into more adaptive and context-aware chatbot architectures.

One notable limitation is the lack of systematic user research in the early stages of development. While most articles clearly define a functional goal for the chatbot, only a minority explicitly assess or validate user needs, expectations, and pain points through surveys, interviews, or participatory design. This gap was directly addressed in the present study, which began by identifying students' informational needs through an institutional questionnaire, thereby ensuring that the chatbot's design was grounded in real user requirements. In many cases, user requirements are inferred from the problem definition rather than being directly collected, which risks reducing real-world adoption and perceived usefulness. Evaluation practices also vary considerably. Some studies provide rigorous quantitative assessments using metrics such as accuracy, BLEU score, or F1-score, while others rely on anecdotal or purely qualitative feedback, making it difficult to compare performance across contexts. Furthermore, external validation remains rare, with most evaluations being limited to the environment in which the chatbot was developed.

Integration patterns reveal a preference for embedding chatbots into existing platforms such as Moodle, Blackboard, or institutional portals, but there is limited research on cross-platform interoperability and the use of low-code tools as accessible gateways to more sophisticated AI backends. The potential for combining visual development interfaces with advanced AI techniques, demonstrated in a few hybrid examples, remains underexplored. In summary, the main research gaps identified are:

- Limited exploration of advanced AI capabilities, including multi-turn dialogue and conversational memory
- Lack of systematic and validated assessment of user needs before development
- Inconsistent evaluation methods, with many studies lacking comparable performance metrics or external validation
- Limited exploration of low-code/advanced AI hybrid architectures and cross-platform integration strategies

Addressing these gaps would enable the development of more user-centered, interoperable, and context-aware chatbot solutions, ultimately increasing both their adoption and impact in real-world applications.

2.3 Summary of the Contributions

The systematic review provided a comprehensive overview of how AI-powered chatbots have been developed, implemented, and evaluated across different domains, with a particular emphasis on their educational potential. The analysis revealed recurring challenges, such as limited validation of user needs, insufficient integration with institutional systems, and the predominance of rule-based architectures, that directly informed the methodological decisions of this research.

The findings of this systematic review provide contributions at three complementary levels: scientific, practical, and contextual. From a scientific perspective, the review consolidates recent, real-world evidence on AI-powered chatbots and clarifies the current state of practice: (i) a predominance of rule-based and hybrid architectures over fully generative systems; (ii) uneven reporting of technical details (e.g., AI techniques, dialogue design) and evaluation procedures; and (iii) underuse of advanced capabilities such as multi-turn dialogue and conversational memory. It also highlights methodological gaps, namely, scarce upfront validation of user needs and limited external or longitudinal assessment, thereby refining the research agenda for applied conversational systems. From a practical perspective, the review offers actionable guidance for professionals and institutions implementing AI-driven chatbots. Successful deployments tend to align closely with validated user needs, prioritise predictable and controllable behaviours (rule-based or hybrid models), and integrate seamlessly with existing institutional platforms (e.g., LMS or university portals). The synthesis also identifies good practices in dialogue design, knowledge-base structuring, and staged evaluation (combining quantitative metrics with user studies), while highlighting common pitfalls such as overreliance on anecdotal feedback or insufficient cross-platform interoperability. From a contextual perspective, the review identifies the most frequent applications of chatbots in education, namely administrative and FAQ support, onboarding, and learning assistance, and the primary user groups involved, mainly students and academic staff. Educational chatbots tend to provide the greatest value in improving information clarity, consistency, and efficiency, while limitations persist in areas such as personalisation, conversational memory, and systematic evaluation.

In sum, this systematic review provides a structured and empirically grounded understanding of current chatbot research and practice. Its insights clarify both the key achievements and remaining challenges in the field, establishing a solid conceptual foundation and practical guidance for the methodological framework presented in the next chapter.

3 METHODOLOGY

This chapter outlines the methodological approach adopted in the present study, based on the DSRM. DSR is particularly suitable for research projects aiming to design, develop, and evaluate innovative artefacts that respond to real-world challenges in organisational and technological contexts.

The chapter is organized in accordance with the DSR model proposed by Peffers et al. (2007), comprising six key stages: problem identification, definition of solution objectives, design and development, demonstration, evaluation, and communication Each of these stages is addressed in this chapter, detailing the methodological decisions, tools, and techniques employed at each step.

3.1 DSR Overview

This research follows the DSRM, a structured approach proposed by Peffers et al. (2007) and grounded in the Design Science principles established Hevner et al. (2004). DSR supports the creation of innovative artefacts, including models, methods, and processes, that not only advance theory but also provide practical solutions, thereby bridging the gap between academic research and real-world problem-solving (Teixeira et al., 2020).

The methodology is characterised by an iterative process, in which cycles of design, testing, and refinement improve the effectiveness of the artefact, consistent with the design-as-a-search paradigm (Hevner et al., 2004). In the model proposed by Peffers et al. (2007), the process is structured into six main stages (Figure *3.1*): (1) Problem Identification and Motivation, (2) Objectives for a Solution, (3) Design and Development, (4) Demonstration, (5) Evaluation, and (6) Communication.

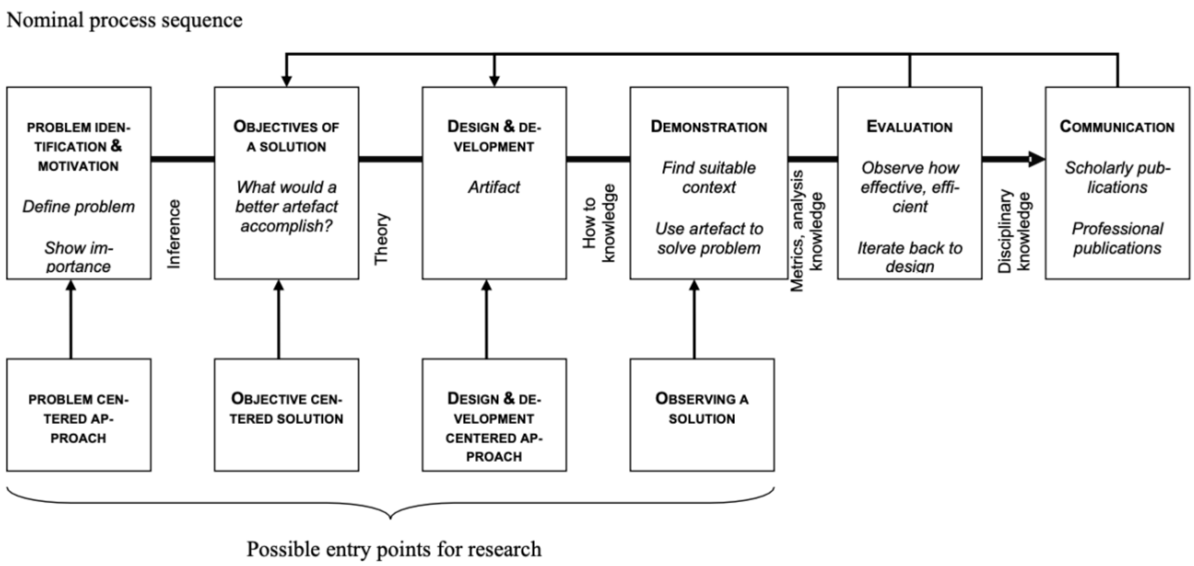


Figure 3.1. Design Science Research Process Model (Peffers et al., 2007, p. 11)

Finally, DSR not only supports a linear process but also an iterative one, allowing the artefact to be designed, tested, and refined continuously based on the results obtained (Peffers et al., 2007; Hevner et al., 2004; Elvas, 2021).

3.2 Identification of the Problem and Motivation

According to Hurter (2024), the implementation of chatbots enables universities to provide continuous support, reducing response times to frequently asked questions and optimising processes such as enrolment and academic counselling. Automating these interactions through AI-based chatbots offers a viable solution for enhancing the student experience, contributing to student attraction and retention (Ravey, 2024), while also optimising institutional resources. Ravey (2024) further reports that 90% of students who have used chatbots in academic institutions found the experience beneficial, underscoring the importance of accessibility to information and automated support. Additionally, major technology providers, such as Microsoft and Azure OpenAI have developed “EduBot – University AI Chatbot Solution”, an AI-driven platform designed to streamline university administrative management, including admissions processes, financial aid, and academic counselling, among other functions (Microsoft AppSource, 2025). In summary, the adoption of such technologies is also shown to reduce operational costs by minimising the need for human intervention in repetitive tasks (Hurter, 2024).

3.2.1 Institutional Context and Unibuddy Analysis

To identify and motivate the problem underlying this project, several data sources were analyzed. Firstly, the institutional context of IBS was considered, particularly the implementation of the Unibuddy platform on the official website (ISCTE, 2021). IBS is one of the four schools of ISCTE, offering a diverse portfolio of undergraduate, master’s, and doctoral programmes across management, economics, finance, marketing, and business analytics. According to institutional data, the school has approximately 4000 students, including around 800 international students from diverse backgrounds, and maintains partnerships with over 80 accredited business schools worldwide (ISCTE Business School, n.d.). This strong international orientation and high level of academic activity contribute to a large volume of information requests from both current and prospective students, often concerning admissions, course structures, and academic procedures. These dynamics underscore the need for scalable and automated communication tools to complement existing support channels. Although Unibuddy provides functionalities similar to a chatbot, responses are not immediate; instead, they are provided later by current or former students.

According to the official report for the 2024 academic year (ISCTE Business School, 2025), most inquiries on Unibuddy were related to programme availability, structure, teaching methods, and assessment (30.21%), followed by admissions (18.03%), financial matters (15.01%), and career prospects (11.31%). A more detailed breakdown reveals that within the programme category, teaching (14.04%) and structure (12.67%) were the most frequent subtopics, while aid (10.82%) dominated the financial category and employability (8.87%) the careers category. These patterns demonstrate that student concerns are not only frequent but also closely aligned with the categories later adopted in the chatbot knowledge base (see Table *3.1*). Another relevant perspective concerns the decision stage of prospective students. As shown in Figure *3.2*, the majority of users were still in the early phases of the decision-making process: 28.37% were "Preparing to apply", 19.39% were "Gathering information", and 13.95% were "Comparing options". This highlights that most interactions took place before formal application, underscoring the importance of timely access to reliable information.

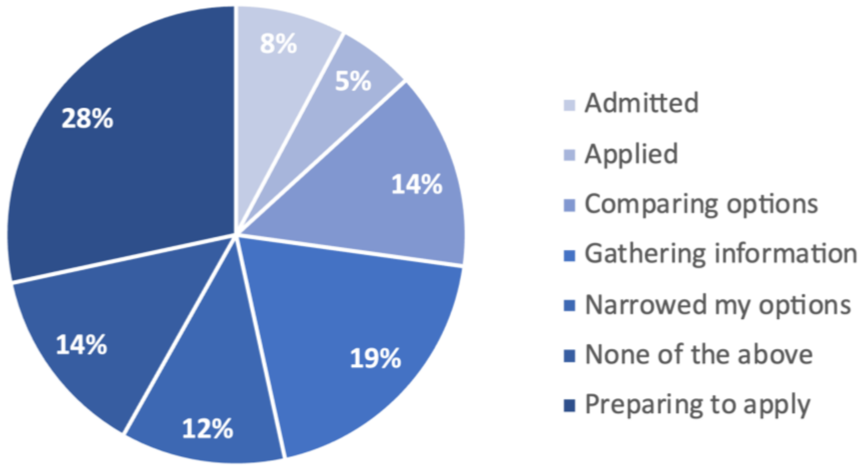


Figure 3.2. Decision Stage of Prospective Students on Unibuddy

Source: Author's elaboration (2025).

The evolution of sign-ups also shows a clear seasonal pattern. As depicted in Figure *3.3*, usage peaked between February and March 2024, coinciding with the main application periods, and declined significantly towards the end of the year. This seasonality reinforces the role of support platforms in handling surges of demand within concentrated timeframes.

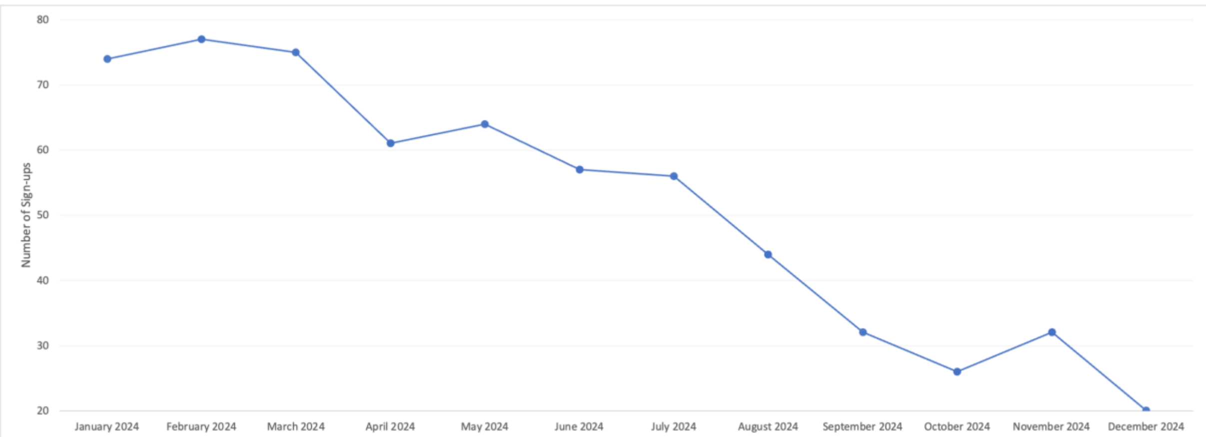


Figure 3.3. Monthly Unibuddy sign-ups

Source: Author's elaboration (2025).

In terms of programme interest, Unibuddy data shows high demand across multiple degrees. The most frequently consulted programmes (among seven bachelor's and 15 master's and two PhDs) were the Master in Hospitality and Tourism Management (10.24%), the Master in International Studies (9.27%), the Master in Management (9.02%), the Master in Marketing (8.29%), Bachelor in Management (7.07%), the Master in Business Administration (6.59%), the Master in Business Analytics (MBA) and the Master in International Management (6.34%), and the Master in Human Resources Management and Organizational Consulting (6.10%). This distribution suggests that doubts are widespread across several programmes rather than concentrated in a single area, supporting the rationale for a centralised support solution. Unibuddy also attracts a highly diverse audience. As illustrated in , while Portugal represents the largest share of users (31.78%), significant numbers also came from Nigeria (11.88%), Pakistan (8.03%), Bangladesh (3.69%), and Brazil (3.21%). In total, students from over 60 different countries engaged with the platform, highlighting the international profile of ISCTE's student community and reinforcing the need for multilingual and inclusive support.

In summary, the Unibuddy analysis highlights three main insights: student queries are highly concentrated in a few thematic areas (programme structure, admissions, financial matters, careers); most prospects interact with the platform at early decision stages, when access to timely and accurate information is critical; and, the student audience is diverse in both disciplinary and geographical terms, reinforcing the importance of a centralised, multilingual support system. These findings provided a strong foundation for designing the proposed virtual assistant, ensuring alignment with actual student needs and institutional priorities.

3.2.2 Analysis of Student Questionnaire

To better understand the information needs and recurring doubts of students throughout their academic journey, an open questionnaire was administered using Microsoft Forms, a cloud-based application that is part of the Microsoft 365 ecosystem and allows the creation of surveys, quizzes, and polls for real-time data collection and analysis (Microsoft, n.d.). The main aim is to identify the types of questions that students frequently ask, or feel the need to ask, from the application stage through to their current academic experience. Although the majority of responses came from students enrolled on the Master's in Business Analysis, participants from other IBS ISCTE Master's programmes were also included, allowing us to assess whether a potential support solution could be replicated at different degrees. A total of 65 responses were collected, generating 542 individual questions, with an average of 8.34 questions per respondent.

To facilitate the analysis and structuring of the collected questions, a text classification model was developed using Python. The goal was to automatically assign each question to one of ten pre-defined categories, representing the most common thematic areas related to student doubts.

Each category was clearly defined and accompanied by representative examples to support the classification process. The following list outlines the categories used, along with their respective descriptions and sample questions:

Table 3.1. Classification categories and representative student question examples

Category	Description	Examples
Academic Workload and Requirements (**A**)	Weekly effort, time management, prerequisites	"Is it manageable to work and study at the same time?"
Administrative and Student Support Services (**B**)	Enrolment, certificates, student cards, transport pass	"Where can I get information about scholarships?"
Application and Enrolment Process (**C**)	Application documents, admission criteria, tuition, scholarships	"What documents are required to apply for the programme?"
Career Prospects and Employability (**D**)	Job market, career options, salary expectations	"What job opportunities are there after graduation?"
Extracurricular Activities and Academic Opportunities (**E**)	Events, workshops, career fairs, networking	"What kind of extracurricular events are available?"
Facilities and Campus Infrastructure (**F**)	Classrooms, library, study spaces, campus map	"Are there study spaces available on campus?"
General Perception and Personal Expectations (**G**)	Programme value, motivation, expectations	"Is it worth the investment?"
Master's Programme Structure and Organisation (**H**)	Curricular units, class schedules, exams, electives	"What are the compulsory and elective courses?"
Technological Resources and Learning Materials (**I**)	Software, platforms, databases, computer requirements	"Which software is required?"
Thesis Guidance (**J**)	Supervision, topics, requirements, deadlines	"What are the requirements for the thesis?"

This classification system allowed for a more efficient and consistent analysis of the 542 open-ended responses provided by the students.

In order to validate the classification logic and ensure the reliability of the process, a stratified random sample of 50 questions (five from each category) was selected. The following Python excerpt illustrates the procedure used to extract the sample, ensuring that all categories were equally represented and that the selection was reproducible through a fixed random seed.

```
df_sampled = (
    df.groupby("Category")
       .apply(lambda x: x.sample(n=min(len(x), 5), random_state=42))
       .reset_index(drop=True)
)
```

This balanced dataset (Sample 50 Balanced.xlsx) was then categorised manually. The first rater corresponded to the author of this dissertation, while the second rater was a fellow master's student enrolled in the Business Analytics (BA) programme at IBS, familiar with the institutional context and thematic scope of the collected questions. This dual rating process aimed to reduce potential bias and assess consistency. The agreement between the two raters was calculated in Python using *scikit-learn*, as shown in the excerpt.

```
import pandas as pd
from sklearn.metrics import accuracy_score, cohen_kappa_score

df = pd.read_excel("Validation - Manual Comparison.xlsx")
y1 = df["Manual Category - Primary Rater"]
y2 = df["Manual Category - Second Rater"]

accuracy = accuracy_score(y1, y2)
kappa = cohen_kappa_score(y1, y2)
print(f"Accuracy = {accuracy:.2%} | Cohen's Kappa = {kappa:.2f}")
```

The results yielded an accuracy of 74% and a Cohen's Kappa coefficient of 0.71, suggesting a consistent level of agreement between raters. *Accuracy* refers to the proportion of correct classifications relative to the total number of items assessed. According to ISO 5725-1 (1994), *accuracy* reflects both trueness and precision, providing a comprehensive indicator of reliability. In this study, it measures the consistency with which both raters assigned the same category to each question. Cohen's Kappa is a widely accepted statistic for measuring inter-rater agreement on categorical data, while accounting for agreement occurring by chance (McHugh, 2012; Sim & Wright, 2005). The kappa coefficient ranges from 0 to 1, where 0 indicates agreement equivalent to chance, and 1 represents perfect agreement. Higher values indicate stronger interrater reliability, reflecting consistent classification across evaluators. According to the scale proposed by Landis and Koch (1977), a value between 0.61 and 0.80 reflects substantial agreement, thereby confirming the robustness of the classification criteria adopted in this study. In the few cases where disagreement occurred, the author's classification was retained, as the differences were minimal and did not affect the overall categorisation consistency.

Subsequently, the same sample of 50 questions was used to compare the manual categorisation against the results produced by the automatic classification model. This comparison was implanted in Python by applying the *accuracy_score* function from *scikit-learn* to the manually validated labels and the model predictions, as shown in the excerpt.

```
import pandas as pd
from sklearn.metrics import accuracy_score

y_true = df_updated["Category"]
y_pred = df_updated["Manual Category"]
accuracy = accuracy_score(y_true, y_pred)
print(f"Accuracy = {accuracy:.2%})
```

The resulting accuracy was 94%, with most misclassifications occurring between closely related categories.

Having confirmed the validity of the classification logic, the full dataset of 542 questions was then categorised automatically. This step was implemented in Python using the GPT-4o-mini model. The prompt used for classification was constructed directly from the category definitions and examples presented in , and the model was instructed to return exclusively the name of one of the ten categories. An excerpt of the implementation is shown in the excerpt.

```
df["Category"] = df["Question"].apply(
    lambda q: client.chat.completions.create(
        model="gpt-4o-mini",
        messages=[{"role": "user", "content": get_category_prompt(q)}],
        temperature=0
    ).choices[0].message.content.strip()
)
```

Based on this procedure, the dataset was classified into the ten categories, and the resulting distribution is presented in Table *3.2*. The distribution of questions across the ten categories was then calculated using Python (*pandas*), as illustrated in the excerpt.

```
import pandas as pd
df = pd.read_excel("Questions with Categories.xlsx")
category_counts = df["Category"].value_counts().sort_index()
```

Table 3.2. Distribution of student questions by category

Category	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**
Questions	**63**	29	49	**70**	38	10	47	**149**	22	**65**
Percentage of Total Questions	**11.62%**	5.35%	9.04%	**12.92%**	7.01%	1.85%	8.67%	**27.49%**	4.06%	**11.99%**

Note: Category letters (A–J) correspond to the classification categories defined in , A = Academic Workload and Requirements, B = Administrative and Student Support Services, C = Application and Enrolment Process, D = Career Prospects and Employability, E = Extracurricular Activities and Academic Opportunities, F = Facilities and Campus Infrastructure, G = General Perception and Personal Expectations, H = Master's Programme Structure and Organisation, I = Technological Resources and Learning Materials, J = Thesis Guidance.

The results of this calculation are summarised in Table *3.2*, which highlights a clear predominance of questions related to the category A (11.62%), D (12.92%), H (27.49%), and J (11.99%). These findings provide a data-driven foundation for prioritising the chatbot's initial content and functionalities. Furthermore, to assess the consistency of needs across different Master's programmes, the distribution of thematic categories was examined per course. The presence of at least one question per category was marked with an "X", indicating that the topic was raised by students from that programme. This was implemented in Python using a *cross-tabulation* (*pandas.crosstab*), as illustrated in the excerpt.

```
import pandas as pd
df_filtered = df[["Masters", "Category"]].dropna()
presence_table = pd.crosstab(df_filtered["Category"], df_filtered["Masters"])
presence_table = presence_table.applymap(lambda x: "X" if x > 0 else "")
```

Table 3.3. Thematic coverage of student questions across master's programmes

Category	Business Administration	BA	Hospitality and Tourism Management	Management	Management of Services and Technology
A	X	X		X	X
B	X	X	X	X	X
C	X	X	X	X	X
D	X	X	X	X	X
E	X	X		X	X
F		X		X	
G	X	X	X	X	X
H	X	X	X	X	X
I		X	X		
J	X	X		X	

Note: Category letters (A–J) correspond to the classification categories defined in , e.g., A = Academic Workload and Requirements, B = Administrative and Student Support Services, C = Application and Enrolment Process, D = Career Prospects and Employability, E = Extracurricular Activities and Academic Opportunities, F = Facilities and Campus Infrastructure, G = General Perception and Personal Expectations, H = Master's Programme Structure and Organisation, I = Technological Resources and Learning Materials, J = Thesis Guidance.

The resulting presence matrix is presented in Table *3.3*, confirming that most categories appear across multiple programmes and reinforcing the assumption that the proposed chatbot solution can be generalised to support different Master's degrees.

3.3 Definition of Objectives for a Solution

The review of the literature in chapter 2 revealed a set of persistent gaps in the adoption of AI-driven chatbots within academic contexts, particularly the limited exploration of hybrid architectures combining the accessibility of low-code platforms with the capabilities of advanced AI models, the insufficient integration with institutional systems, and the lack of design approaches informed by real student needs.

The institutional analysis in subchapter 3.2.1 and the questionnaire results in subchapter 3.2.2 confirmed that these gaps are present in the IBS context. The data collected from 542 student-generated questions, distributed across ten thematic categories, showed a strong concentration of doubts in Academic Workload and Requirements, Master's Programme Structure and Organisation, Career Prospects and Employability, and Thesis Guidance. These recurring needs highlight the importance of providing students with a scalable, accessible, and context-aware support solution, capable of offering accurate information in real time.

Based on the evidence gathered from both the literature review in chapter 2 and the empirical analysis of institutional data and the student questionnaire in subchapters 3.2.1 and 3.2.2, the main objectives of the proposed solution are as follows:

- To ensure 24/7 access to accurate and up-to-date institutional information, addressing the lack of continuous, on-demand academic support observed in current chatbot implementations and responding to students' need for timely answers to recurrent questions
- To reduce the workload on academic services and faculty by automating responses to repetitive, low-complexity queries, thus optimising resources and freeing staff to focus on higher-value tasks
- To unify and centralise information according to a taxonomy validated by students, overcoming current fragmentation and ensuring alignment with the ten thematic categories derived from the classification model
- To support a diverse and international student body, ensuring inclusivity in language and content, reflecting the multilingual and multicultural profile of IBS community
- To enable scalability and adaptability, with the potential to integrate more advanced AI components in the future, leveraging the potential identified in the literature for combining low-code platforms with more sophisticated AI techniques

The chatbot's initial knowledge base, the collection of institutional information and documents that the system uses to generate responses (Microsoft, 2025), was designed according to the ten validated categories (Table *3.1*), ensuring comprehensive coverage from the start and allowing for gradual expansion based on user feedback and system usage patterns. Before development, a comparative analysis was conducted between low-code and advanced development approaches, followed by an evaluation of alternative low-code platforms, to ensure that the chosen technology met the previously defined functional, integration, and scalability requirements.

3.3.1 Comparative Analysis of Development Approaches

Given the objectives outlined in subchapter 3.3, it was necessary to identify the most appropriate development approach for the proposed chatbot. In particular, the literature review (chapter 2) and the institutional needs analysis (subchapters 3.2.1 and 3.2.2) highlighted the trade-offs between agility, integration capabilities, and scalability, factors that strongly influence the choice between low-code platforms such as Microsoft Copilot Studio, or Dialogflow, and advanced, code-first development such as Rasa, or TensorFlow.

To guide this decision, a comparative analysis was conducted focusing primarily on literature and industry reports regarding low-code platforms (Balani, 2024; Borsje, 2020; Brüggemann, 2023; Gialitakis et al., 2024; Hanenko, 2024; Hirzel, 2023; Karl, 2024; Kissflow, 2025; Okeke & Akinbolajo, 2023; SAP, n.d.; Sorrentino, 2022; Syulina, 2024). These sources outline the main strengths and limitations of low-code solutions, which were contrasted against documented characteristics of advanced, code-first development derived from established software engineering literature and industry best practices (Dagkoulis & Moussiades, 2023). Table *3.4* summarises the comparison across eight key criteria: ease of use, development speed, technical expertise, cost, integration, NLP/AI support, security, and scalability.

Table 3.4. Comparative analysis of low-code platforms and advanced development approaches

Criterion	Low-code Platforms	Advanced Development
Ease of use	Visual interfaces, minimal coding, accessible to non-programmers	Requires advanced programming and system architecture skills
Development speed	Rapid prototyping, adjustments quickly	Slow (involves detailed coding and testing)
Technical expertise	Low (basic solutions can be built by non--technical staff)	High (requires AI/NLP specialists and experienced developers)
Cost	Low initial investment, predictable monthly licensing fees	High initial cost, but potentially more cost-effective long term
Integration	Ready-made connectors for common systems	Fully customisable integrations
NLP/AI support	Native integration with NLP and pre-defined APIs	Full support, but requires manual setup and training
Security	Pre-configured standards and certifications	Security must be implemented from scratch
Scalability	Limited to platform capabilities	Fully adaptable and scalable

Source: Adapted from Balani (2024), Borsje (2020), Brüggemann (2023), Dagkoulis & Moussiades (2023), Gialitakis et al. (2024), Hanenko (2024), Hirzel (2023), Karl (2024), Kissflow (2025), Okeke & Akinbolajo (2023), SAP (n.d.), Sorrentino (2022), and Syulina (2024).

Based on the comparison presented in Table *3.4*, low-code platforms emerge as the most suitable approach for the institutional context of IBS. This choice is justified by their ability to enable rapid development and prototyping (Gialitakis et al., 2024; Hirzel, 2023), reduce dependency on specialised technical staff and support citizen development (Brüggemann, 2023; Karl, 2024; Kissflow, 2025), and contribute to more predictable costs through subscription models (Borsje, 2020; Sorrentino, 2022). While advanced development offers greater scalability and full customisation, such benefits are outweighed by the longer development cycles, higher upfront costs, and the need for specialised engineering resources (Hanenko, 2024; SAP, n.d.). Low-code platforms therefore provide a balanced solution that supports the project's objectives: fast deployment, inclusivity, and adaptability to recurring student needs, while maintaining sufficient scalability for institutional growth (Borsje, 2020; Hirzel, 2023). Although fully coded solutions can offer greater architectural control in highly demanding contexts, low-code platforms are better suited to the academic profile of this project because they facilitate rapid prototyping and development, reduce the need for highly specialized technical expertise, and enable participation by non-technical users (Dagkoulis & Moussiades, 2022; El Kamouchi et al., 2023; Pinho et al., 2023). Therefore, to identify the most appropriate low-code solution for this context, a comparative analysis was conducted across leading platforms, as presented in the following subchapter.

3.3.2 Comparative Analysis of Low-code Platforms

Building on the rationale, this subchapter compares a set of representative platforms, including Dialogflow CX (Barker, 2024; Google Cloud, 2025), IBM Watson Assistant (IBM, n.d.), Microsoft Copilot Studio (Microsoft, n.d.), and OutSystems (Outsystems, n.d.). Each platform is evaluated against criteria such as ease of use, integration capabilities, NLP/AI support, scalability, security, and cost-effectiveness, with the aim of selecting the most suitable option for the institutional context of IBS.

Accordingly, Table *3.5* presents a comparative analysis of the selected low-code platforms, using the same evaluation criteria of Table *3.4* to determine the most appropriate solution. Based on this analysis, Microsoft Copilot Studio emerges as the preferred option. In addition to offering low-code development capabilities and strong integration with AI services, it is part of the Microsoft ecosystem, for which ISCTE already holds institutional licenses (Informática Iscte, 2025). This existing integration facilitates adoption and reduces implementation barriers, even though the specific Copilot Studio license has not yet been acquired. For these reasons, Copilot Studio is identified as the most suitable development platform for the proposed chatbot solution.

Table 3.5. Comparative analysis of selected low-code chatbot development platforms

Criterion	Dialogflow	IBM Watson Assistant	Microsoft Copilot Studio	OutSystems
Ease of use	Visual interface; drag-and-drop for intents and entities	User-friendly UI; predefined skills and templates	Low-code studio integrated in Power Platform; strong for citizen developers	Visual app builder with chatbot components; steeper learning curve
Development Speed	Fast	Moderate	Fast	Moderate to fast
Technical expertise	Low to medium	Medium	Low	Medium to High
Cost	Flexible pricing (pay-per-use model)	Tiered pricing (basic to enterprise)	License-based (predictable within Microsoft ecosystem)	High licensing costs (enterprise-oriented)

Criterion	Dialogflow	IBM Watson Assistant	Microsoft Copilot Studio	OutSystems
Integration	Easy integration with Google Cloud services	Wide connectors with IBM Cloud, enterprise tools, and APIs	Strong integration with Microsoft ecosystem	Broad integration with enterprise systems
NLP/AI support	Advanced NLU, multilingual support; ML training available	Solid NLP; supports intent classification and entity extraction	Native AI models and access to Azure OpenAI service	Limited native NLP; relies on connectors or external APIs
Security	Google Cloud security standards	Enterprise-level compliance	Microsoft security and compliance	Strong enterprise-grade compliance
Scalability	High	High	High	Very high (designed for large enterprises)

Source: Adapted from Barker (2024), Google Cloud (2025), IBM (n.d.), Microsoft (n.d), and Outsystems (n.d.).

3.4 Design and Development

Having defined the objectives and selected Microsoft Copilot Studio as the development platform, this subchapter outlines the design and development of the proposed chatbot. The approach prioritised a functional prototype that could be iteratively improved, starting with a minimal but coherent configuration aligned with the institutional analysis and student questionnaire, and progressively expanding as limitations emerged.

3.4.1 Initial Prototype

The first prototype, named *IBS Virtual Assistant*, was conceived as a baseline version to test feasibility, with only the minimum elements required for meaningful interaction. Its configuration prioritised clarity, institutional alignment, and multilingual accessibility, defining the assistant as a multilingual tool designed to answer common academic and administrative queries (e.g., programme structure, enrolment, career prospects, thesis guidance), available 24/7 to ensure timely and inclusive support. In practice, the configuration of this first version followed five guiding principles (see Appendix A):

1. The knowledge base was restricted to the official ISCTE and IBS websites, excluding direct links to individual master's programmes. This choice reflected the goal of testing the chatbot with high-level institutional information before moving to programme-specific details.
2. Orchestration was disabled, which meant that the chatbot could only rely on predefined topics and the knowledge base, avoiding the use of free generative AI. This ensured greater control over responses and reduced the risk of producing inaccurate or fabricated content.
3. Only the default system topics were kept (Greeting, Goodbye, Start Over, Thank you, Conversation Start, and End of Conversation), although their content was adapted to fit the academic context of IBS.

4. Custom topics were not yet created for the ten categories (A–J) derived from the questionnaire. At this stage, the taxonomy served only as a conceptual framework, without being explicitly embedded into the design.
5. Finally, a small number of suggested prompts were included to guide interaction. These were drawn directly from the student questionnaire and served mainly as a way of testing the system's initial performance.

3.4.1.1 Instructions

To ensure consistency in its behaviour, a set of design instructions was embedded in Copilot Studio. These rules defined how the assistant should interact with students, particularly in terms of language detection, tone, and the academic orientation of its answers. An excerpt of these instructions is shown in Appendix B.

3.4.1.2 Suggested Prompts

The prompts created for the first version were designed as conversation starters, providing users with examples of how to interact with the chatbot. Each example displayed with a short title in English (EN), followed by the corresponding question in both Portuguese (PT) and EN, reflecting ISCTE's bilingual academic environment. These visual prompts appeared at the beginning of each chat session to guide users in formulating their queries. Examples include:

- Failing a Course (PT: Se reprovar a uma unidade curricular, tenho de repetir a mesma ou posso escolher outra? | EN: If I fail an optional course, do I need to repeat it or can I change to another one?)
- Optional Courses (PT: Quais as cadeiras optativas? | EN: Which optional courses are available?)
- Career Opportunities (PT: Quais as principais saídas profissionais? | EN: What are the main career opportunities after the programme?)
- Thesis Guidance (PT: O que é necessário para o registo da tese? | EN: What is required for thesis registration?)

This configuration marked the establishment of the baseline version of the *IBS Virtual Assistant*. Although limited in scope, it provided a controlled environment in which to evaluate the chatbot's performance and laid the foundation for the improvements that would be introduced in subsequent versions.

3.4.2 Final Prototype

The final version of the *IBS Virtual Assistant* was developed as an enhanced and more autonomous iteration of the initial one, incorporating the insights and limitations identified during the first testing cycle (see subchapter 4.1.1). This stage aimed to strengthen the chatbot's informational accuracy, response completeness, and scope coverage while maintaining the same principles of institutional alignment, clarity, and multilingual accessibility.

Several improvements were implemented across three key dimensions: knowledge base, interaction design, and behavioural instructions. The knowledge base was expanded to include a broader range of official ISCTE and IBS sources, such as programme-specific documents, student regulations, academic calendars, and scholarship information. This expansion ensured that the assistant could provide detailed and factually correct answers for multiple master's programmes, rather than relying solely on high-level institutional content. All information was drawn from verified ISCTE websites and public documents, preserving the reliability of responses. Unlike the initial version, the final one integrated additional system topics and refined conversation logic. The assistant was configured to always ask for clarification when the user did not specify which master's programme they were referring to, thereby preventing the delivery of inaccurate or ambiguous information. The conversation flow also included improved fallback responses to ensure graceful handling of out-of-scope queries and to redirect users to appropriate institutional contacts when needed. The system prompt embedded in Microsoft Copilot Studio was completely rewritten to guide the assistant's linguistic, ethical, and contextual behaviour. The updated version emphasised multilingual support (with European Portuguese as default for Portuguese queries), professional tone (Følstad & Brandtzaeg, 2020), factual precision (McGrath et al., 2024), and strict adherence to official sources. The instructions also outlined procedures for handling inappropriate language, redirecting users to official contacts, and maintaining academic formality throughout interactions.

To ensure secure and compliant testing, the final prototype was published within the institutional environment of ISCTE, accessible through Microsoft Teams and Microsoft 365 Copilot. This approach enabled student interaction under controlled conditions while avoiding the licensing restrictions associated with public deployment via the Demo Website channel. This configuration represents a mature version of the *IBS Virtual Assistant*, designed to support a wider range of queries and programmes while maintaining factual precision, academic integrity, and user-centred interaction. The demonstration and evaluation of this final prototype are presented in chapter 0, where its performance is compared against the initial version.

3.4.2.1 Instructions

To ensure behaviour consistency and improved response quality, the interaction rules embedded in Microsoft Copilot Studio were rewritten for the final version of the *IBS Virtual Assistant*. This new system prompt replaced the initial instructions set presented in subchapter 3.4.1.1, integrating the insights obtained from the first demonstration phase (see subchapter 4.1.1). The updated version prioritised factual precision, multilingual accuracy, and ethical compliance, while expanding the chatbot's ability to respond effectively across different master's programmes. The full prompt used in the final configuration is shown in Appendix C.

3.4.2.2 Suggested Prompts

The set of predefined prompts was updated to include both general and programme-specific questions in PT and EN, reflecting the most frequent queries identified in the initial needs assessment. These prompts serve as conversation starters, guiding users on how to interact with the chatbot. Examples include:

- Programme Structure (PT: Quais são as unidades curriculares incluídas no Mestrado em BA? | EN: What courses are included in the MBA?)
- Admissions (PT: Quais são os prazos e requisitos de candidatura ao Mestrado em BA? | EN: What are the application deadlines and requirements?)
- Scholarships (PT: Existem bolsas de estudo disponíveis para estudantes de mestrado? | EN: Are there any scholarships available for Master's students?)
- Career Opportunities (PT: Quais são as principais saídas profissionais após o Mestrado em BA? | EN: What are the main job prospects after graduation?)
- Thesis Guidance (PT: Quais são os passos necessários para o registo do tema da minha dissertação? | EN: What are the steps to register my dissertation topic?)

3.5 Demonstration

The demonstration phase in DSR consists of applying the developed artifact in a relevant context, typically through simulations, experiments, or case studies, to show that it can address the identified problem (Peffers et al., 2007). Similarly, Hevner et al. (2004) emphasise that demonstrating is essential to verify the artifact's usefulness beyond theoretical design, ensuring that it performs as intended under real or simulated conditions.

In this study, the demonstration process was take place in two sequential stages, corresponding to the iterative development of the *IBS Virtual Assistant* in Microsoft Copilot Studio. The first stage focuses on the initial prototype, which is evaluated in a controlled setting using a fixed set of 17 questions. These questions were derived from the student questionnaire conducted during the needs assessment phase (see subchapter 3.2.2) and cover all ten validated categories of student queries, with three additional "out-of-scope" questions to test the chatbot's ability to recognise its own limitations. The second corresponds to the final prototype, which was tested again using the same set of 17 questions. This approach allows a direct comparison between both versions, highlighting areas of improvement in accuracy, completeness, clarity, and tone.

To assess performance during the demonstration phase, a structured evaluation framework was defined based on four widely used criteria in conversational AI (Munroe, 2025): accuracy, completeness, clarity, and tone/professionalism. Each criterion was rated on a three-point scale (1 = fully correct or appropriate; 0.5 = partly correct or appropriate; 0 = incorrect or inappropriate), as follows:

- Accuracy measures the factual correctness of the response compared with official ISCTE institutional sources (e.g., websites, regulations, and academic policies)
- Completeness evaluates whether the response convers all relevant aspects of the question
- Clarity assesses the grammatical and structural quality of the message and its overall comprehensibility
- Tone/Professionalism measures the appropriateness of the chatbot's communication style relative to the academic and institutional standards of ISCTE

This evaluation framework provides a systematic method for measuring the chatbot's communicative and informational quality, ensuring consistent assessment across both prototypes. The results from these two demonstration cycles were form the basis for analysing the chatbot's evolution and identifying areas that require refinement. The results of this phase, including representative screenshots and analysis, are presented in chapter 0, where the demonstration process is documented and analysed in detail.

3.6 Evaluation

Evaluation is a central phase in DSR, as it provides the evidence required to determine whether an artifact effectively addresses the problem it was designed to solve (Hevner et al., 2004). In the DSRM proposed by Peffers et al. (2007), this stage involves examining the artifact against the predefined objectives and problem requirements. Similarly, Gregor and Hevner (2013) emphasise that rigorous evaluation is essential to establish the credibility of both scientific and practical contributions, while the Framework for Evaluation in Design Science (FEDS) framework, Venable et al. (2016) highlights that evaluation strategies should be adapted to the artifact and its context, ranging from controlled laboratory tests to naturalistic field studies.

Within the scope of this study, evaluation was conducted through an end-user survey designed to capture students' perceptions of the *IBS Virtual Assistant*. The instrument is based on the UTAUT, originally developed by Venkatesh et al. (2003) to explain technology adoption and diffusion. UTAUT identifies four key constructs: performance expectancy, effort expectancy, social influence, and facilitating conditions, that influence behavioural intention and actual use. A meta-analysis by Dwivedi et al. (2011) confirms the widespread application and reliability of these constructs in empirical research on information systems adoption. The questionnaire was administered in EN via Microsoft Forms (see Appendix D), ensuring accessibility for both national and international students. Items were adapted from validated UTAUT scales and tailored to the context of the *IBS Virtual Assistant*. For example, performance expectancy was measured with items such as *"The IBS Virtual Assistant helps me quickly access relevant information about my Master's programme"*, while effort expectancy included items such as *"Using the IBS Virtual Assistant is straightforward and clear"*. Similar adaptations were made for social influence, facilitating conditions, and behavioural intention to use. Responses were measured on a five-point Likert scale, ranging from strongly disagree (1) to strongly agree (5). The full questionnaire is presented in Appendix D, together with screenshots of the Microsoft Forms survey.

The collected data was analysed using Statistical Package for the Social Sciences (SPSS) Statistics (version 30), applying descriptive statistics to summarise user perceptions, Cronbach's alpha (Tavakol & Dennick, 2011) to assess the internal reliability of the dimensions, and composite scores, computed as the mean of the items within each construct, to compare results across dimensions. This evaluation design ensures that the assessment focuses not only on the technical functionality of the chatbot (already demonstrated in the previous subchapter) but also on its acceptance, perceived usefulness, and ease of use among students. The detailed results of this phaseare presented and discussed in chapter 0.

3.7 Communication

Communication represents the final stage of the DSR process. As noted by Hevner et al. (2004), the results of a DSR project should be disseminated both to academic audiences, who assess its scientific contribution, and to practitioners, who can benefit from its practical utility. Similarly, in the methodology proposed by Peffers et al. (2007), communication is regarded as a necessary step to ensure that the knowledge created, and the artifact developed, have impact beyond the immediate scope of the project. Therefore, the specific communication strategies adopted in this research was detailed in chapter 0, alongside the results of their implementation.

4 RESULTS AND DISCUSSION

This chapter presents and discusses the results obtained from the development, demonstration, and evaluation of the *IBS Virtual Assistant*. Following the DSRM, the findings are organised to reflect the iterative process through which the artefact was designed, refined, and assessed.

First, the demonstration phase examines the chatbot's performance across two iterations, the initial and final prototypes, using predefined evaluation metrics to highlight improvements in *accuracy*, *completeness*, *clarity*, and *tone*. Afterwards, the user evaluation explores students' perceptions of the assistant, based on the questionnaire developed according to the UTAUT. This stage provides insight into perceived usefulness, ease of use, and overall acceptance of the system. The The chapter concludes with a discussion of the main findings, connecting the results to the existing literature and identifying their implications for future academic and institutional applications.

4.1 Demonstration

This chapter presents the results of the two demonstration cycles conducted during the development of the IBS Virtual Assistant. Following the DSRM, each cycle served to evaluate and refine the artifact, with the first version functioning as a baseline prototype and the second incorporating improvements derived from empirical testing and design feedback. The analysis focuses on four key evaluation metrics (*accuracy*, *completeness*, *clarity*, and *tone*), applied consistently across both prototypes (see subchapter 3.5). These metrics allow for a comparative and systematic assessment of the chatbot's performance, highlighting both technical and communicative progress. The results are presented quantitatively, through metric averages, and qualitatively, through selected interaction examples.

4.1.1 Initial Prototype

To analyse the performance of the initial prototype, four evaluation metrics were applied: *accuracy*, *completeness*, *clarity*, and *tone/professionalism*. These metrics were assessed across a set of 17 questions: one from each of the ten validated categories, four additional questions from the most frequent themes identified in the student needs questionnaire, and three "out-of-scope" questions designed to test the chatbot's behaviour when confronted with queries beyond its defined domain.

To ensure statistical consistency, mean values were calculated for each metric, allowing for a comparative assessment of performance within and outside the chatbot's scope.

Table 4.1. Performance metrics of the IBS Virtual Assistant (initial prototype)

Metrics	In-scope	Out-of-scope
Accuracy	0.86	0.67
Completeness	0.71	0.83
Clarity	1.00	1.00
Tone/Professionalism	1.00	1.00

Metrics	In-scope	Out-of-scope
Number of Questions	14	3

The quantitative results reveal a strong overall performance within the chatbot’s defined scope. An *accuracy* score of 0.86 indicates that most responses were factually correct and consistent with official institutional sources, though minor precision errors were occasionally observed. The *completeness* score of 0.71 suggests that while most responses covered the essential information, approximately one-third did not address all relevant aspects of the questions in sufficient depth. Conversely, both *clarity* and *tone/professionalism* achieved perfect scores (1.00), reflecting exemplary communication quality, syntactic correctness, and consistent adherence to ISCTE’s institutional style. For out-of-scope questions, *accuracy* naturally decreased to 0.67, as expected, due to the chatbot’s limited access to information beyond its predefined knowledge base. Interestingly, *completeness* increased slightly to 0.83, likely because the chatbot compensated for factual uncertainty by offering broader, more generalised explanations. Nevertheless, *clarity* and *tone* remained consistently excellent (1.00), demonstrating the chatbot’s ability to maintain professional and polite communication even when unable to provide precise answers.

Example of these different performance scenarios, including fully correct, factually incaccurate, and out-of-scope responses, are presented in Figure *4.1*, Figure *4.2*, and Figure *4.3*.For readability, embedded links automatically generated by the chatbot are represented by the placeholder [links].

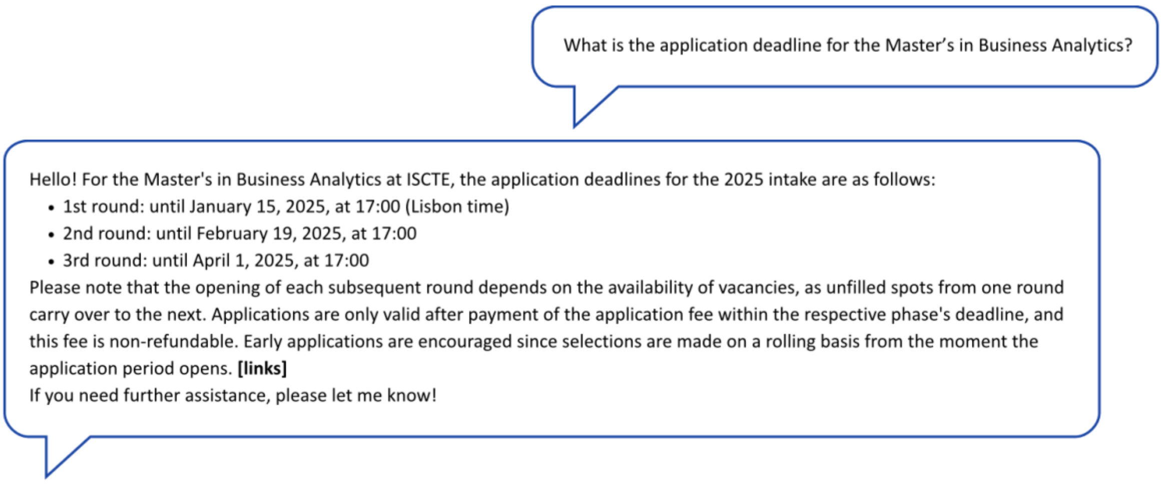


Figure 4.1. Exemplary chatbot response to an in-scope query.

Source: Author’s elaboration (2025).

Note: The answer is factually correct, complete, and clearly structured, achieving top scores (Accuracy = 1; Completeness = 1; Clarity = 1; Tone = 1).

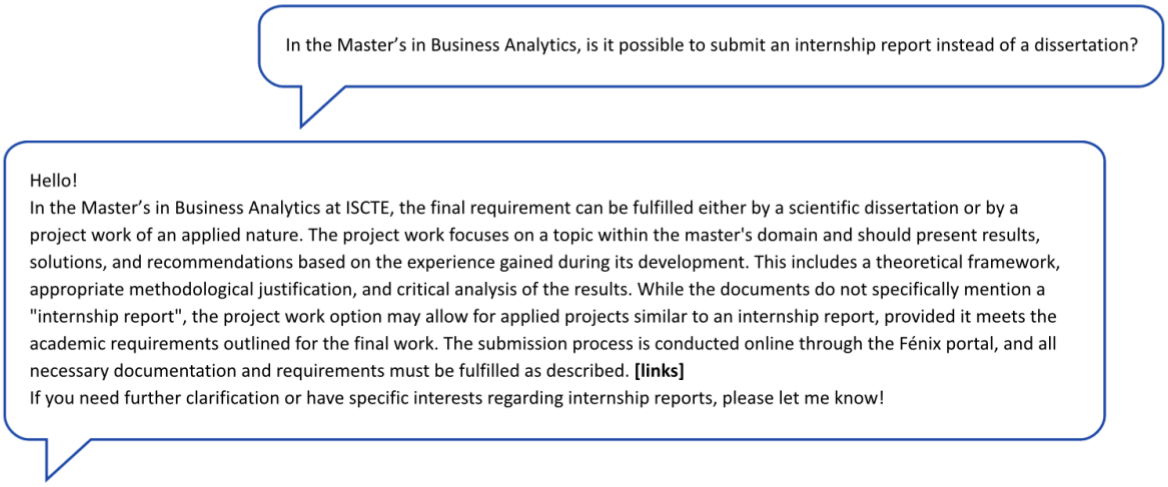


Figure 4.2. Factually inaccurate chatbot response to an in-scope query.

Source: Author's elaboration (2025).

Note: The answer is clear and professional but contains an incorrect statement regarding the existence of a project-work option in the MBA (Accuracy = 0; Completeness = 0.5; Clarity = 1; Tone = 1).

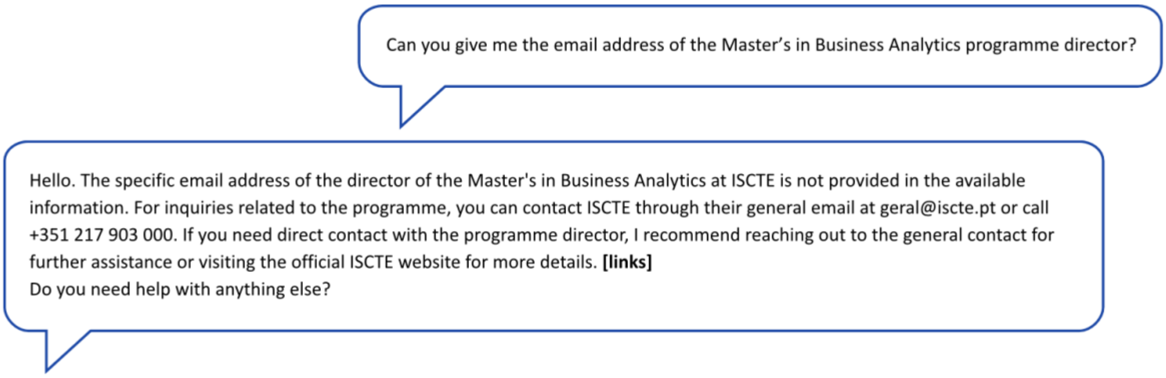


Figure 4.3. Chatbot response to an out-of-scope query.

Source: Author's elaboration (2025).

Note: The answer maintains clarity and professionalism, appropriately declining to disclose personal information and redirecting the user to official institutional contacts (Accuracy = 1; Completeness = 1; Clarity = 1; Tone = 1).

The combined analysis indicates a high level of communicative robustness, reflected in consistently clear and professional language, along with strong factual accuracy within scope. However, minor gaps were identified in *completeness*, particularly for questions with broader thematic content (for example, when asked "*What are the main carreer opportunities after the programme?*", the assistant provided a general overview without detailing specific professional roles or setors. Out--of-scope responses revealed adequate fallback mechanisms, yet highlighted the need for more explicit recognition of knowledge boundaries and redirection to appropriate institutional sources. These findings provide a solid baseline for improvement in the next development iteration. Accordingly, the final prototype focused on expanding informational coverage and implementing uncertainty-handling strategies to enhance the overall reliability and user experience of the *IBS Virtual Assistant*.

4.1.2 Final Prototype

The second testing cycle focused on the final version of the *IBS Virtual Assistant*, developed after implementing the design improvements described in subchapter 3.4.2. The objective of this stage is to validate the impact of the modifications introduced, namely the expansion of the knowledge base, refinement of the prompt and behavioural rules, and the enhancement of fallback and clarification mechanisms. As in the initial demonstration, the chatbot's performance was evaluated using the four established metrics: *accuracy*, *completeness*, *clarity*, and *tone/professionalism*.

Table 4.2. Performance metrics of the IBS Virtual Assistant (final prototype)

Metrics	In-scope	Out-of-scope
Accuracy	1.00	1.00
Completeness	0.86	1.00
Clarity	1.00	1.00
Tone/Professionalism	1.00	1.00
Number of Questions	14	3

The results reveal a consistently high performance across all dimensions. The chatbot achieved a perfect score in *accuracy* (1.00) for both in-scope and out-of-scope queries, demonstrating its ability to provide factually correct and institutionally aligned responses. *Completeness* showed a slight deviation (0.86) within the programme's scope, which can be attributed to a few answers that, although fully understandable and contextually appropriate, offered slightly less elaboration than others. For instance, when asked *"Who are the faculty members of the Master's in Business Analytics?"*, the chatbot correctly listed the members of the selection and ranking committee but omitted the remaining teaching staff available on ISCTE's official website. Similarly, in response to *"Where can I find the ISCTE campus map showing the classrooms?"*, the assistant provided general directions and building identifiers but did not include the detailed floor maps available online. These examples demonstrate that, while the responses were accurate and clear, some lacked the depth necessary to be considered fully complete. These occurred mainly in categories F (Facilities and Campus Infrastructure), G (Faculty Information), one in H (Programme Structure), and one in J (Thesis Guidance). Nevertheless, all these responses remained accurate, coherent, and sufficient for user understanding. Both *clarity* and *tone* maintained perfect scores (1.00), confirming the chatbot's communicative robustness and adherence to ISCTE's academic and institutional communication standards.

Table 4.3. Comparative performance of the IBS Virtual Assistant (initial vs. final prototype)

Metrics	Initial Version	Final Version	Improvement
Accuracy	0.82	1	0.18
Completeness	0.74	0.88	0.14
Clarity	1	1	0

Metrics	Initial Version	Final Version	Improvement
Tone/Professionalism	1	1	0

The comparative analysis demonstrates a notable improvement in both *accuracy* (+0.18) and *completeness* (+0.14). These gains directly reflect the refinements implemented in the final configuration, particularly the expanded knowledge base, the integration of explicit programme clarification prompts, and the enhanced fallback mechanisms for handling ambiguous or out-of-scope queries. Meanwhile, *clarity* and *tone/professionalism* remained at a consistently high level (1.00), indicating that the assistant had already achieved strong communicative performance in the initial version.

Figure 4.4: Average metric scores by prototype (n = 17 identical questions)

Source: Author's elaboration (2025).

As shown in Figure *4.4*, the final prototype exhibits a clear upward trend in factual accuracy and informational coverage, while maintaining perfect *clarity* and *tone*. This evolution highlights the tangible benefits of iterative refinement within the DSR framework. The measurable improvements in *accuracy* and *completeness* demonstrate that the design interventions applied, particularly those related to information scope, linguistic precision, and ethical handling of uncertainty, significantly enhanced the system's reliability and user experience.

To illustrate these improvements, three representative examples were revisited from the initial prototype to verify the qualitative progression in response quality, as shown in Figure *4.5*, Figure *4.6*, and Figure *4.7*.

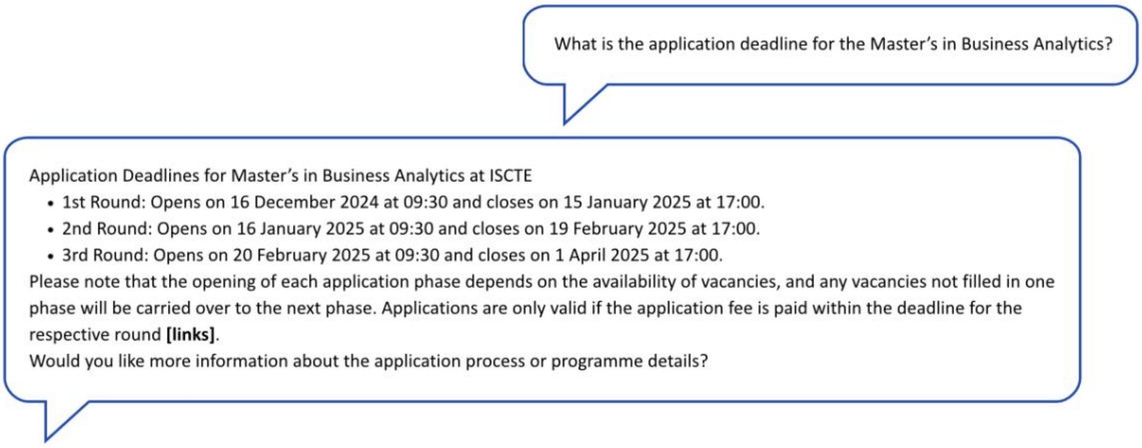


Figure 4.5. Exemplary chatbot response to an in-scope query.

Source: Author's elaboration (2025).

Note: The answer is factually correct, complete, and clearly structured, achieving top scores (Accuracy = 1; Completeness = 1; Clarity = 1; Tone = 1).

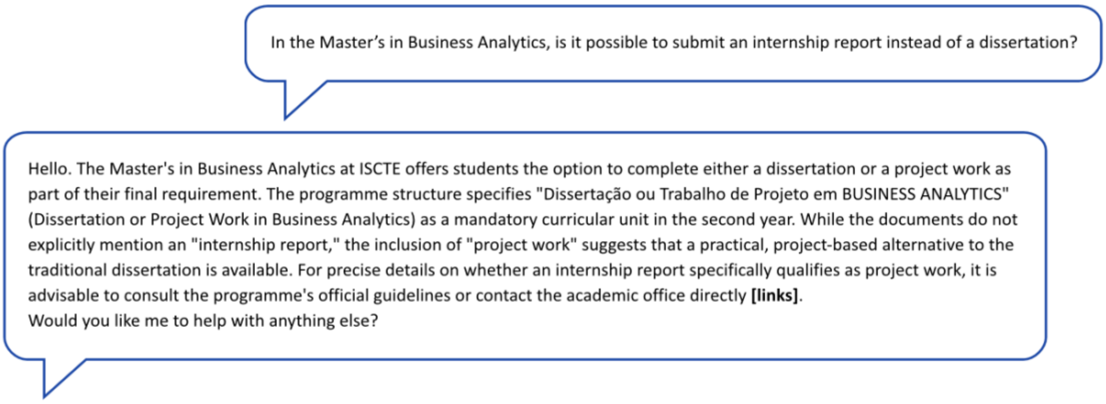


Figure 4.6. Corrected chatbot response to an in-scope query.

Source: Author's elaboration (2025).

Note: This response illustrates the improvement of a previously inaccurate answer. The chatbot now correctly clarifies that students in the MBA may complete either a dissertation or a project work, but not an internship report (Accuracy = 1; Completeness = 1; Clarity = 1; Tone = 1).

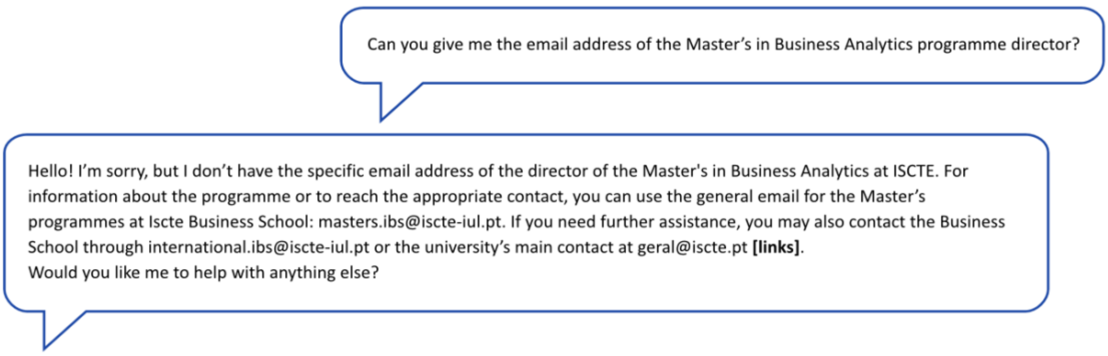


Figure 4.7. Chatbot response to an out-of scope query.

Source: Author's elaboration (2025).

Note: The answer is clear, polite, and compliant with institutional standards, providing general contact information instead of personal details (Accuracy = 1; Completeness = 1; Clarity = 1; Tone = 1).

In summary, the analysis of the final prototype demonstrates that the *IBS Virtual Assistant* has evolved into a robust, accurate, and context-aware tool, capable of addressing students' queries with high degree of reliability and professionalism. The combination of quantitative and qualitative evidence confirms the success of the design interventions introduced between iterations, particularly the expansion of the knowledge base, refinement of behavioural instructions, and reinforcement of scope management mechanisms. The final version achieves maximum scores across the main evaluation metrics, reflecting institutional robustness and communicative maturity. The improvement in factual *accuracy* and informational *completeness* validates the methodological soundness of the design process, aligning with contemporary standards for the evaluation of conversational agents. Maintaining perfect *clarity* and professional *tone* across both versions further confirms the system's alignment with user experience principles and ISCTE's academic communication norms.

4.2 Evaluation

Following the technical validation presented in the demonstration phase, this subchapter focuses on the user-centred evaluation of the *IBS Virtual Assistant*. The purpose of this stage is to assess students' perceptions of the system's usefulness, ease of use, and overall acceptance as an academic support tool. While the previous analyses (subchapter 4.1) verified the chatbot's communicative and factual performance, this phase aims to evaluate its reception and perceived value from an end-user perspective.

The evaluation framework adopted in this study is based on the UTAUT developed by Venkatesh et al. (2003). This model identifies four key constructs that influence behavioural intention and technology adoption: Performance Expectancy, Effort Expectancy, Social Influence, and Facilitating Conditions. These dimensions were adapted to the context of the *IBS Virtual Assistant*, allowing for the systematic assessment of how students perceive its relevance, usability, and institutional support.

The results presented in this subchapter provide empirical evidence of the chatbot's acceptance and perceived impact, complementing the design validation process described in earlier phases. The following subchapters detail the profile of respondents, and the descriptive results obtained across the UTAUT dimensions.

4.2.1 Sample Profile

The sample used for the evaluation of the *IBS Virtual Assistant* consisted of 11 respondents enrolled in the MBA, of whom one was in the first year and 10 were in the second year. Regarding age, eight participants were between 21 and 23 years old, and three were between 24 and 30 years old. All respondents were PT nationals and reported using chatbots with varying frequency: three indicated they "always" use them, seven reported frequent use, and one stated using them "sometimes".

Given this sample profile, small and relatively homogeneous in terms of programme, nationality, and age, and particularly due to the uneven subgroup distribution (e.g., only one first-year participant), no comparative analyses between groups were conducted. The limited number of respondents in certain subgroups would prevent the use of inferential statistical tests and could lead to biased or unreliable conclusions. Accordingly, the analysis focused on a descriptive exploration of the overall dataset and on identifying general trends that reflect the participants' collective perceptions of the developed chatbot.

4.2.2 Descriptive Analysis of UTAUT Dimensios

This subchapter presents the descriptive results of the questionnaire based on the UTAUT. The model evaluates user acceptance of the *IBS Virtual Assistant* across five dimensions: Performance Expectancy, Effort Expectancy, Social Influence, Facilitating Conditions, and Behavioral Intention.

Initially, descriptive statistics were computed for each questionnaire item. Then, composite indices were calculated as the arithmetic mean of the items within each dimension. The internal consistency of each construct was assessed using Cronbach's alpha (Tavakol & Dennick, 2011), with all dimensions showing satisfactory reliability, except for Facilitating Conditions, which displayed a lower coefficient due to the samll and homogeneous sample size (n = 11). As summarised in Table *4.4*, alpha values ranged from 0.48 to 0.93, indicating acceptable to excellent internal consistency according to George & Mallery (2019) thresholds. Performance Expectancy (α = 0.81) and Effort Expectancy (α = 0.78) revealed strong reliability, suggesting that respondents consistently perceived the chatbot as useful and easy to use. Social Influence (α = 0.93) achieved the highest reliability, reflecting stable perceptions regarding peer and institutional endorsement. Behavioral Intention (α = 0.69) showed moderate consistency, while Facilitating Conditions (α = 0.48) indicated weaker reliability, likely influenced by limited response variability and the reduced number of participants.

Table 4.4. Reliability and descriptive statistics of the UTAUT dimensions

Dimension	Number of Itens	Cronbach's Alpha	Mean	Standard Desviation
Performance Expectancy	3	0.81	4.85	0.93
Effort Expectancy	3	0.78	4.85	0.93
Social Influence	3	0.93	4.64	1.58
Facilitating Conditions	3	0.48	4.70	1.22
Behavioral Intention	3	0.69	4.82	0.93

The descriptive results in Table *4.5* indicate that students evaluated the *IBS Virtual Assistant* very positively across all dimensions, with mean values above 4.6 on a five-point Likert scale (1 = strongly disagree; 5 = strongly agree), and low standard deviations, suggesting high agreement among respondents. The dimension with the greatest variability is Social Influence, although it still recorded a high mean score. Overall, the results reveal a consistent pattern across dimensions. The mean values, Performance Expectancy (4.85), Effort Expectancy (4.85), Social Influence (4.64), Facilitating Conditions (4.70), and Behavioral Intention (4.82), demonstrate consistently high ratings across all categories, indicating a globally positive perception of the system. The corresponding standard deviations confirm low dispersion, particularly for Effort Expectancy (0.31) and Performance Expectancy (0.31), suggesting consensus regarding the system's usability and expected performance. Social Influence shows slightly greater variability (0.53), reflecting a more diverse perception of peer influence on chatbot use. Together, these results show that despite the small sample size, the *IBS Virtual Assistant* was perceived as effective, user-friendly, and valuable.

Table 4.5. Descriptive statistics of the evaluated dimensions of the IBS Virtual Assistant

Dimension	N	Minimum	Maximum	Mean	Standard Deviation
Performance Expectancy	11	4.00	5.00	4.85	0.31
Effort Expectancy	11	4.00	5.00	4.85	0.31
Social Influence	11	3.67	5.00	4.64	0.53
Facilitating Conditions	11	4.00	5.00	4.70	0.41
Behavioral Intention	11	4.00	5.00	4.82	0.31

To further explore the determinants of Behavioral Intention, Pearson correlations and simple linear regressions (Zou et al., 2003) were computed between each independent UTAUT dimension and the dependent variable.

As shown in Table *4.6*, all dimensions were positively and significantly correlated with Behavioral Intention ($p < .05$). Among them, Performance Expectancy showed the strongest association ($r = 0.948$, $R^2 = 0.899$), indicating that students' intention to use the *IBS Virtual Assistant* was primarily influenced by their perception of its usefulness and potential benefits. Social Influence ($r = 0.777$, $R^2 = 0.604$) and Facilitating Conditions ($r = 0.749$, $R^2 = 0.561$) also exhibited strong positive correlations, suggesting that peer recommendations and institutional support contribute to higher adoption likelihood. Finally, Effort Expectancy ($r = 0.719$, $R^2 = 0.517$) was positively associated with intention to use, though to a slightly lesser extent, implying that ease of use remains relevant but secondary to perceived performance benefits. These findings are consistent with prior studies applying the UTAUT model to educational technologies, confirming that perceived usefulness and contextual support are key drivers of user adoption.

Table 4.6. Correlation and explanatory power of UTAUT dimensions on Behavioral Intention

Dimension	Pearson's r	R^2	Sig. (p)
Performance Expectancy	0.948	0.899	< .001
Effort Expectancy	0.719	0.517	.013
Social Influence	0.777	0.604	.005
Facilitating Conditions	0.749	0.561	.008

In addition to the closed-ended items, the questionnaire concluded with an open-ended question (see Appendix D) to collect qualitative feedback from participants. Four respondents provided suggestions, which revealed a shared interest in expanding the chatbot's accessibility and scope. Specifically, participants recommended:

- Integrating the assistant into other institutional platforms, particularly the official ISCTE website and Moodle
- Broadening its content beyond master's programmes to include undergraduate and PhD information
- Extending the chatbot's application to other academic areas within and beyond IBS

These comments indicate that students not only perceived the current version of the *IBS Virtual Assistant* as useful but also envisioned its future potential as a broader institutional support tool. This qualitative feedback reinforces the positive quantitative results, demonstrating both satisfaction with the system's performance and motivation to see it further developed and integrated within the academic ecosystem.

4.3 Communication

The communication of this research occurred through multiple channels, targeting both academic and institutional stakeholders. First, the project was presented to first-year Master's students during a class taught by Professor Raul Laureano. This session not only introduced the concept of the chatbot but also served as a means of gathering students' needs through a questionnaire, which directly informed the design of the knowledge base and conversational flows. Second, the project was discussed with IBS faculty members and administrative services, particularly those responsible for managing Unibuddy. At an early stage, exploratory interviews with faculty were also carried out to better understand expectations and potential areas of application. However, as the research design evolved, these interviews were not integrated into the final methodological framework of the thesis, even though they contributed to shaping the initial understanding of the institutional context. Finally, the research itself constitutes a formal channel of communication, contributing to the academic debate on AI-driven chatbots in higher education. By documenting the design process, methodological decisions, and findings, it serves as both a record of the project and a foundation for future research and institutional initiatives.

4.4 Discussion and Recommendations

The findings from the evaluation confirm that the *IBS Virtual Assistant* successfully met the design objectives defined in chapter 0, showing clear improvements in accuracy, completeness, and user acceptance between the initial and final prototypes. The evolution from version 1 to version 2 illustrates the strength of the DSR approach, where continuous feedback and performance data guided focused refinements in the chatbot's dialogue design and knowledge base. Similar iterative gains are reported in the literature (e.g., Liu et al., 2024; Chang et al., 2022), where user feedback drove progressive enhancement of chatbot performance.

The methodological path adopted here, an initial questionnaire to surface students' informational needs followed by a UTAUT-based evaluation, p kept development explicitly user-centred and responds to a gap identified in the review, namely the limited upfront validation of user needs in many studies. Students' responses across all UTAUT dimensions, particularly Performance Expectancy and Effort Expectancy, indicate that the assistant was perceived as useful and easy to use, echoing positive results reported for educational chatbots in Kaiss et al. (2023), Oliveira & Matos (2023) and Daniel et al. (2022). The high mean values observed in this study (≈ 4.6–4.9 on a 1–5 scale) are at least comparable to, and in some cases above, ranges reported in similar implementations (e.g., Chang et al., 2022; Kaiss et al., 2023). Minor gaps in Completeness, observed in some responses, mirror the challenges reported by Attigeri et al. (2024) and Alfirevic et al. (2024), multi-part queries were not always covered in full, underlining the importance of maintaining an up-to-date and comprehensive knowledge base.

Drawing on both the empirical results and participants' feedback, several practical steps can be taken to enhance the system and extend its institutional impact:

- Platform integration: Embed the assistant in core institutional touchpoints (e.g., Moodle, Teams, and the IBS website) to maximise access for students and applicants, consistent with implementations in the literature that leverage portal integration (Alfirevic et al., 2024; Alqahtani & Alrwais, 2023; Chang et al., 2022)
- Knowledge base expansion and maintenance: Extend beyond master's to bachelor's and doctoral programmes, and establish a simple update routine (responsible owner, periodic checks, and change log) to keep regulations, calendars and scholarship info current
- Multilingual reach: Keep EN and PT as default and prioritize one additional language (e.g., Spanish or French) given the international profile surfaced in the institutional data
- Query robustness: Add patterns for frequent multi-part questions and explicit fallback and redirection rules to official contacts when out-of-scope

In summary, the results demonstrate that the *IBS Virtual Assistant* achieved a high level of performance and user acceptance, aligning with the best practices and success factors identified in the literature. The system proved to be accurate, consistent, and well-received, with users recognising its potential to simplify communication and improve access to academic information. Furthermore, the methodological approach, combining DSR with the UTAUT evaluation framework, proved effective in linking technical performance with user-centred validation. By addressing the main research gaps identified in the systematic review, particularly the limited validation of user needs and the integration of low-code AI tools in higher education, this study contributes both empirical evidence and practical guidance for future institutional chatbot developments. The following chapter concludes the research by summarising these contributions, discussing the study's limitations, and proposing directions for future research.

5 CONCLUSION

This final chapter synthesises the main outcomes of the research, critically reflecting on the extent to which the initial objectives were achieved. It further discusses the theoretical and practical contributions of the study, acknowledges its limitations, and outlines potential directions for future research and institutional development.

5.1 Summary

This study addressed the research question: "How can an AI-powered chatbot be effectively and efficiently used to support users in higher education contexts?". To answer this question, the research adopted the DSRM, which provided a systematic framework for identifying the problem, defining objectives, designing and developing the chatbot, demonstrating its functionality, evaluating its performance, and communicating the results. The empirical project was conducted within IBS, initially focusing on the MBA as a pilot context and subsequently extending to other IBS master's programmes to demonstrate the solution's scalability and replicability across degrees.

The motivation for this research stemmed from the growing demand for scalable digital solutions in higher education, particularly in administrative communication. The analysis of institutional data and student-generated questions revealed recurring informational gaps and inefficiencies in existing support channels, such as Unibuddy, where delayed responses and limited automation hindered user experience. To address these challenges, the study developed the *IBS Virtual Assistant*, a multilingual conversational agent built on Microsoft Copilot Studio, designed to provide accurate, consistent, and timely responses to students' academic and administrative queries. The chatbot's knowledge base, structured collection of institutional documents and verified online sources used to generate responses, was organised into ten validated categories derived from 542 student questions collected through an institutional survey.

Central to the project was a student-centred methodology conducted in two stages. The first questionnaire identified and validated students' most frequent informational needs, forming the foundation for the knowledge base. The second, based on the UTAUT model, evaluated perceptions of the developed prototype, confirming that the chatbot was widely perceived as useful, easy to use, and relevant to students' needs. This multi-stage approach ensured that both design and evaluation were grounded in real user expectations, principle consistent with best practices in human, AI interaction research (Venkatesh et al., 2003; Dwivedi et al., 2011). The final prototype incorporated targeted refinements, an expanded knowledge base, clearer behavioural prompts, and stronger fallback and clarification rules, which resulted in observable improvements in response accuracy, completeness, clarity, and tone. The user evaluation confirmed high levels of acceptance across all UTAUT dimensions, reinforcing the assistant's effectiveness in addressing students' recurrent questions while highlighting opportunities for continuous refinement. The positive evaluations, complemented by qualitative feedback, align with findings in the literature that emphasise perceived usefulness and interaction simplicity as key predictors of chatbot adoption in educational contexts (Kaiss et al., 2023; Oliveira & Matos, 2023; Chang et al., 2022).

The study also raised important considerations regarding implementation and data governance. While internal deployment within Microsoft Teams ensures controlled access and data protection, future integration into public-facing platforms such as Moodle or the official ISCTE website would expand reach and inclusivity. However, such expansion requires clear policies on content ownership, privacy compliance, and update routines to maintain accuracy and institutional trust, issues widely discussed in recent research on AI governance in higher education (Dwivedi et al., 2024). Although the results were highly positive, they reflect the current scope and quality of the knowledge base and the limited sample size of the user test. High user satisfaction therefore indicates strong alignment between the chatbot's content and available data rather than a definitive state of completeness. Continuous maintenance and iterative development remain essential to ensure factual precision and to progressively incorporate additional sources, experiential perspectives (e.g., student testimonials), and updated institutional information.

Overall, the research demonstrates that the *IBS Virtual Assistant* constitutes an effective, adaptable, and context-aware solution for improving informational access and communication efficiency in higher education. These findings confirm the methodological robustness of the DSR approach when applied to AI-driven educational innovations, bridging theoretical design with practical institutional application.

5.2 Contributions

This research offers both scientific and practical contributions to the fields of educational technology, AI, and digital transformation in higher education. Beyond its academic relevance, it delivers tangible institutional value by providing insights that bridge theory and practice.

In direct response to the research question, the study demonstrated that the process begins with understanding real user needs. Based on students' frequently asked questions, collected and validated through an institutional survey, a ten-category taxonomy was created to structure the chatbot's knowledge base. This knowledge base, built from official and publicly available institutional materials such as programme webpages, academic regulations, calendars, and scholarship documents, ensured both accuracy and institutional alignment. The design also included the definition of key behavioural parameters, such as tone and professionalism, multilingual handling, fallback and clarification rules, and scope boundaries. The chatbot was then implemented through a low-code platform, Microsoft Copilot Studio, integrated within the Microsoft 365 ecosystem, which enabled rapid prototyping, institutional compliance, and easy future scalability. Finally, the system was evaluated through a UTAUT-based questionnaire to assess user perceptions, closing the loop between design choices and student experience. This sequence, identifying needs, defining taxonomy, building the knowledge base, parameterising behaviour, implementing through low-code tools, and evaluating user acceptance, provides a transparent and replicable methodology that can guide future chatbot development projects in academia.

From a theoretical perspective, this study demonstrates how the DSRM can be effectively applied to the development of conversational AI solutions, offering a structured and transparent framework that bridges problem identification, artefact design, and evaluation. The validation of the ten-category taxonomy (A–J) for student questions represents a novel contribution, providing a systematic and evidence-based method for organising and prioritising institutional knowledge. This structure improves both the design and the analytical evaluation of educational chatbots, aligning technological solutions with user needs. The inclusion of the UTAUT model in the evaluation phase further enriches the theoretical contribution. While no multiple regression was performed, the correlation analysis revealed that Performance Expectancy was most strongly associated with the intention to use the chatbot, followed by Social Influence, Facilitating Conditions, and Effort Expectancy. These findings are consistent with prior studies on educational technologies, confirming that perceived usefulness and contextual support are key factors driving user adoption. Another methodological contribution of this research lies in the combination of analytical tools used throughout the process. Python was employed to automate data preparation and compute inter-rater reliability using Cohen's Kappa, while SPSS was used for statistical analysis, including descriptive statistics, reliability testing (Cronbach's alpha), and correlation analyses. This multi-tool approach improved analytical efficiency and ensured the accuracy and reproducibility of the results. Additionally, the systematic literature review conducted in chapter 2 contributes to the consolidation of current knowledge on AI-powered chatbots implemented in real contexts. By synthesising findings from 25 empirical studies, it identifies key patterns, limitations, and gaps, offering a reference framework that can assist both researchers and practitioners seeking to design, evaluate, or adopt educational chatbots.

From a practical standpoint, the research produced a fully functional prototype of the *IBS Virtual Assistant*, marking a concrete step toward the digital transformation of ISCTE's academic support services. The project demonstrated that low-code platforms such as Microsoft Copilot Studio enable rapid development and deployment of conversational systems without extensive programming expertise, making them accessible to institutional teams. The chatbot automates responses to frequent queries, thereby reducing administrative workload and allowing staff to focus on more complex and value-added interactions. For students, the system offers several tangible benefits. It provides 24/7 access to reliable, multilingual, and up-to-date information, ensures greater transparency and consistency in communication, and helps reduce the uncertainty and delays that often occur in administrative processes. It also supports internationalisation, as it facilitates access to information in both EN and PT, with potential to expand to other languages such as Spanish or French.

Finally, this research proposes a roadmap for future institutional adoption of AI assistants. The recommendations include expanding the chatbot's knowledge base to undergraduate and doctoral programmes, ensuring continuous updates of institutional content, integrating the assistant with key platforms such as Moodle and Teams, and maintaining a lightweight governance model to ensure content accuracy and data protection. These measures was support the chatbot's evolution from a pilot prototype to a fully integrated institutional service.

In summary, this research not only produced a functional and user-validated AI chatbot for higher education but also provided a replicable methodological model, a validated taxonomy of student informational needs, and practical insights that can guide future developments in academic AI systems.

5.3 Limitations

Despite the encouraging results obtained, several limitations constrain the scope, generalisability, and replicability of this study.

First, the user evaluation involved a small and relatively homogeneous sample (n = 11), composed exclusively of master's students from the MBA programme. While the findings provide clear indications of acceptance and perceived usefulness, broader testing with participants from different academic levels and disciplinary backgrounds would produce more representative and generalisable results. Second, the licensing restrictions of Microsoft Copilot Studio limited the chatbot's public deployment. Because access to the "Demo Website" channel was unavailable, testing had to occur in internal environments (Teams and Microsoft 365 Copilot). This constraint prevented the observation of large-scale interaction dynamics such as concurrent users, multilingual engagement, or spontaneous queries from external audiences. Third, the prototyping environment operated on static knowledge bases derived from selected institutional documents (programme webpages, PDF guides, and regulations) rather than dynamic, automatically updated databases. Moreover, the chatbot did not directly access the official ISCTE website or programme pages, which are updated periodically. As a result, some responses, such as the one referencing 2025 enrolment data, reflect the static nature of the information provided at the time of implementation. This highlights the need for periodic reviews and content refresh cycles to ensure sustained accuracy and relevance. Fourth, as with any AI system based on natural language processing, there remains a potential risk of factual inaccuracies or "hallucinations". In this study, such risks were minimised through restricted data sources and prompt engineering; however, a broader discussion could consider whether allowing controlled access to open models (e.g., ChatGPT) might enhance the chatbot's adaptability while maintaining accuracy through validation protocols. Achieving the right balance between openness and institutional control remains an important avenue for future exploration. Finally, the study adopted a cross-sectional design, focusing on immediate user perceptions rather than longitudinal adoption patterns. Future longitudinal assessments could measure sustained engagement, behavioural change, and institutional impact over time, providing deeper insight into how continued exposure influences satisfaction and usage habits. In addition, further research should test the chatbot's scalability and adaptability across different schools within ISCTE and potentially in other higher education institutions. Such replication would help determine the generalisability of both the system and the methodology.

Acknowledging these limitations provides a constructive basis for future improvement, reinforcing the importance of continuous refinement, broader empirical validation, and open methodological transparency in the ongoing evolution of the *IBS Virtual Assistant* and similar AI-driven educational tools.

5.4 Roadmap for the Future Research

Building on the lessons learned throughout this research, several directions are proposed for future development and investigation.

Continuous refinement and expansion of the chatbot's knowledge base are essential to ensure comprehensive and up-to-date coverage across all ten informational categories. Although the assistant demonstrated strong accuracy and user satisfaction, these results reflect the scope of the current knowledge base and the limited number of participants (n = 11). For the knowledge available, the chatbot's responses were largely correct; however, this does not imply completeness, as continuous maintenance and validation remain necessary to ensure reliability over time. For example, certain programme pages and institutional sources still display outdated information (e.g., 2025 admissions data), highlighting the need for periodic verification and updates to preserve factual accuracy. Future improvements should therefore prioritise a structured maintenance process for the knowledge base, including periodic content reviews and the addition of new information such as faculty directories, email contacts, and detailed staff profiles, elements publicly available on the institutional website but not yet fully integrated into the chatbot. Likewise, incorporating anonymised excerpts from real student, staff interactions could help the chatbot better address experience-based or contextual questions that currently fall outside factual responses, while maintaining ethical and data protection standards. Another important direction concerns the integration of real-time institutional data. Linking the chatbot with systems such as Fénix and Moodle would allow dynamic retrieval of information on schedules, grades, and deadlines, turning the assistant into an interactive gateway between students and the university's academic infrastructure. To evaluate such integration, a longitudinal pilot study could monitor user adaptation and satisfaction over multiple academic periods, helping to identify the most relevant features and usage patterns. Personalisation also represents a promising avenue for future development. Implementing user-specific profiles could enable the chatbot to tailor answers based on each student's programme, academic progress, and role (e.g., applicant, enrolled student, or alumnus). Additionally, expanding the assistant's multilingual and accessibility capabilities, beyond EN and PT to include other major languages such as Spanish or French, would enhance inclusivity within ISCTE's diverse academic community. Finally, hybrid support models that combine automated assistance with human follow-up for complex or sensitive topics should be explored to balance efficiency with empathy and institutional alignment. Beyond technical evolution, future research should also examine governance, data privacy, and ethical implications associated with the deployment of AI chatbots in academic environments, ensuring compliance with institutional and European data protection frameworks (e.g., General Data Protection Regulation).

In summary, although the latest version of the *IBS Virtual Assistant* achieved high accuracy and user satisfaction, design improvements must remain continuous. The system's long-term success depends on iterative refinement, institutional integration, and responsible governance. By sustaining this process and expanding its reach across platforms such as Moodle, Microsoft Teams, and the official ISCTE website, the chatbot could evolve into a fully integrated, multilingual, and ethically grounded virtual assistant. Moreover, longitudinal studies, hybrid support models, and cross-platform implementations are essential to understand its sustained impact on engagement, satisfaction, and learning outcomes. Ultimately, the *IBS Virtual Assistant* provides a strong foundation for intelligent academic support and a replicable model for other higher education institutions aiming to advance digital transformation in a user-centred and ethically responsible way.

REFERENCES

Ait Baha, T., El Hajji, M., Es-Saady, Y., & Fadili, H. (2024). The impact of educational chatbot on student learning experience. *Education and Information Technologies*, *29*(8), 10153–10176. https://doi.org/10.1007/s10639-023-12166-w

Alabbas, A., & Alomar, K. (2024). Tayseer: A Novel AI-Powered Arabic Chatbot Framework for Technical and Vocational Student Helpdesk Services and Enhancing Student Interactions. *Applied Sciences*, *14*(6), 2547. https://doi.org/10.3390/app14062547

Alfirevic, N., Pranicevic, D. G., & Mabic, M. (2024). Custom-Trained Large Language Models as Open Educational Resources: An Exploratory Research of a Business Management Educational Chatbot in Croatia and Bosnia and Herzegovina. *SUSTAINABILITY*, *16*(12). https://doi.org/10.3390/su16124929

Alhassan, N. A., Saad Albarrak, A., Bhatia, S., & Agarwal, P. (2022). A Novel Framework for Arabic Dialect Chatbot Using Machine Learning. *Computational Intelligence and Neuroscience*, *2022*, 1–11. https://doi.org/10.1155/2022/1844051

Allen, M., Naeem, U., & Gill, S. S. (2024). Q-Module-Bot: A Generative AI-Based Question and Answer Bot for Module Teaching Support. *IEEE Transactions on Education*, *67*(5), 793–802. https://doi.org/10.1109/TE.2024.3435427

Alqahtani, Q., & Alrwais, O. (2023). Building a Machine Learning Powered Chatbot for KSU Blackboard Users. *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS*, *14*(2), 781–789.

Attigeri, G., Agrawal, A., & Kolekar, S. V. (2024). Advanced NLP Models for Technical University Information Chatbots: Development and Comparative Analysis. *IEEE Access*, *12*, 29633–29647. https://doi.org/10.1109/ACCESS.2024.3368382

Balani, H. (2024, November 19). *Forrester Wave on low-code development platforms 2025*. ToolJet. https://blog.tooljet.ai/forrester-wave-on-low-code-development-platforms-2025/?utm_source=chatgpt.com

Barker, S. (2024). *Building Secure And Scalable Conversational AI With Dialogflow CX And Cloud Functions*. Expert Beacon. https://expertbeacon.com/building-secure-and-scalable-conversational-ai-with-dialogflow-cx-and-cloud-functions/

Borsje, J. (2020). *The low-code investment*. Siemens Blogs. https://blogs.sw.siemens.com/partners/the-low-code-total-cost-of-ownership-tco-mendix-siemens/?utm_source=chatgpt.com

Brüggemann, L. (2023). *Securely extending and running low-code applications with C#*.

Chang, C., Hwang, G., & Gau, M. (2022). Promoting students' learning achievement and self-efficacy: A mobile chatbot approach for nursing training. *British Journal of Educational Technology*, *53*(1), 171–188. https://doi.org/10.1111/bjet.13158

Chase, M. (2024). Academic Libraries Can Develop AI Chatbots for Virtual Reference Services with Minimal Technical Knowledge and Limited Resources. *Evidence Based Library and Information Practice*, *19*(2), 136–138. https://doi.org/10.18438/eblip30523

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). Artificial Intelligence trends in education: a narrative overview. *Procedia Computer Science*, *136*, 16–24. https://doi.org/10.1016/j.procs.2018.08.233

Chen, H., Shi, N., Chen, L., & Lee, R. (2024). Enhancing educational Q&A systems using a Chaotic Fuzzy Logic-Augmented large language model. *Frontiers in Artificial Intelligence*, *7*. https://doi.org/10.3389/frai.2024.1404940

Corrêa, J. S., Neto, A. P. de A., Pinto, G. R., Lima, L. D. B., & Teles, A. S. (2023). Lhia: A Smart Chatbot for Breastfeeding Education and Recruitment of Human Milk Donors. *Applied Sciences*, *13*(12), 6923. https://doi.org/10.3390/app13126923

Dagkoulis, I., & Moussiades, L. (2022). A Comparative Evaluation of Chatbot Development Platforms. *Proceedings of the 26th Pan-Hellenic Conference on Informatics*, 322–328. https://doi.org/10.1145/3575879.3576012

Dagkoulis, I., & Moussiades, L. (2023). A Comparative Evaluation of Chatbot Development Platforms. *Proceedings of the 26th Pan-Hellenic Conference on Informatics*, 322–328. https://doi.org/10.1145/3575879.3576012

Daniel, T., de Chevigny, A., Champrigaud, A., Valette, J., Sitbon, M., Jardin, M., Chevalier, D., & Renet, S. (2022). Answering Hospital Caregivers' Questions at Any Time: Proof-of-Concept Study of an Artificial Intelligence–Based Chatbot in a French Hospital. *JMIR Human Factors*, *9*(4). https://doi.org/10.2196/39102

El Kamouchi, H., Kissi, M., & El Beggar, O. (2023). Low-code/No-code Development : A systematic literature review. *2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA)*, 1–8. https://doi.org/10.1109/SITA60746.2023.10373712

Gialitakis, M.-K., Tsakalidis, G., Nousias, N., & Vergidis, K. (2024). Rapid prototyping of process-driven applications using low-code development platforms: A case study from the Greek public sector. *2024 International Conference Automatics and Informatics (ICAI)*, 650–656. https://doi.org/10.1109/ICAI63388.2024.10851543

Google Cloud. (2025). *Conversational Agents (Dialogflow CX) documentation*. https://cloud.google.com/dialogflow/cx/docs

Hanenko, K. (2024, October 9). *The rise of Low-Code: How citizen developers are changing the game*. Altamira. https://www.altamira.ai/blog/the-rise-of-low-code/?utm_source=chatgpt.com

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). DESIGN SCIENCE IN INFORMATION SYSTEMS RESEARCH 1. In *Design Science in IS Research MIS Quarterly* (Vol. 28, Issue 1).

Hirzel, M. (2023). Low-Code Programming Models. *Communications of the ACM*, *66*(10), 76–85. https://doi.org/10.1145/3587691

IBM. (n.d.). *Watsonx Assistant*. Retrieved August 19, 2025, from https://www.ibm.com/products/watsonx-assistant

Informática Iscte. (2025). *Software*. https://informatica.iscte-iul.pt/servicos/software/

Islam, M. A. N., Warsito, B., & Nurhayati, O. D. (2024). AI-DRIVEN CHATBOT IMPLEMENTATION FOR ENHANCING CUSTOMER SERVICE IN HIGHER EDUCATION: A CASE STUDY FROM UNIVERSITAS NEGERI SEMARANG. *Journal of Theoretical and Applied Information Technology*, *102*(14), 5690–5701. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85214834842&partnerID=40&md5=b7a4ce75b5969c2d552249f8c49ec79e

Jenneboer, L., Herrando, C., & Constantinides, E. (2022). The Impact of Chatbots on Customer Loyalty: A Systematic Literature Review. *Journal of Theoretical and Applied Electronic Commerce Research*, *17*(1), 212–229. https://doi.org/10.3390/jtaer17010011

Kaiss, W., Mansouri, K., & Poirier, F. (2023). Effectiveness of an Adaptive Learning Chatbot on Students' Learning Outcomes Based on Learning Styles. *International Journal of Emerging Technologies in Learning (IJET)*, *18*(13), 250–261. https://doi.org/10.3991/ijet.v18i13.39329

Karl, T. (2024). *Embracing the Future: Low Code/No Code in Citizen Development*. New Horizons. https://www.newhorizons.com/resources/blog/low-code-no-code?utm_source=chatgpt.com

Kissflow, T. (2025, April 23). *How Low-Code And Citizen Development Simplify App Development*. Kissflow. https://kissflow.com/citizen-development/how-low-code-and-citizen-development-simplify-app-development/?utm_source=chatgpt.com

Lal, M., & Neduncheliyan, S. (2023). An optimal deep feature–based AI chat conversation system for smart medical application. *Personal and Ubiquitous Computing*, *27*(4), 1483–1494. https://doi.org/10.1007/s00779-023-01713-4

Liu, S., Guo, X., Hu, X., & Zhao, X. (2024). Advancing Generative Intelligent Tutoring Systems with GPT-4: Design, Evaluation, and a Modular Framework for Future Learning Platforms. *Electronics*, *13*(24), 4876. https://doi.org/10.3390/electronics13244876

Martín Martín, A. (2019). *Generación de herramientas de evaluación bibliométrica a partir de Google Scholar*. https://doi.org/9788413062303

Microsoft. (n.d.). *Microsoft Copilot Studio documentation*. Retrieved August 19, 2025, from https://learn.microsoft.com/en-us/microsoft-copilot-studio/

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *Annals of Internal Medicine*, *151*(4), 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135

Moore, K., Zhong, S., He, Z., Rudolf, T., Fisher, N., Victor, B., & Jindal, N. (2024). A comprehensive solution to retrieval-based chatbot construction. *Computer Speech & Language*, *83*, 101522. https://doi.org/10.1016/j.csl.2023.101522

Okeke, H., & Akinbolajo, O. (2023). Building secure and compliant web applications using low-code methodologies. *World Journal of Advanced Research and Reviews*, *3*, 2266–2276. https://doi.org/10.30574/wjarr.2023.20.3.2572

Oliveira, P. F., & Matos, P. (2023). Introducing a Chatbot to the Web Portal of a Higher Education Institution to Enhance Student Interaction †. *Engineering Proceedings*, *56*(1). https://doi.org/10.3390/ASEC2023-16621

Outsystems. (n.d.). *Outsystems Evaluation Guide*. Retrieved August 19, 2025, from https://www.outsystems.com/evaluation-guide/

Peffers, K., Tuunanen, T., Gengler, C. E., Rossi, M., Hui, W., Virtanen, V., & Bragge, J. (2007). *THE DESIGN SCIENCE RESEARCH PROCESS: A MODEL FOR PRODUCING AND PRESENTING INFORMATION SYSTEMS RESEARCH Corresponding Author*.

Pereira, D. S. M., Falcão, F., Nunes, A., Santos, N., Costa, P., & Pêgo, J. M. (2023). Designing and building OSCEBot ® for virtual OSCE – Performance evaluation. *Medical Education Online*, *28*(1). https://doi.org/10.1080/10872981.2023.2228550

Peyton, K., Unnikrishnan, S., & Mulligan, B. (2025). A review of university chatbots for student support: FAQs and beyond. *Discover Education*, *4*(1), 21. https://doi.org/10.1007/s44217-025-00397-7

Pinho, D., Aguiar, A., & Amaral, V. (2023). What about the usability in low-code platforms? A systematic literature review. *Journal of Computer Languages*, *74*, 101185. https://doi.org/10.1016/j.cola.2022.101185

Ramakrishnan, R., Thangamuthu, P., Nguyen, A., & Gao, J. (2024). Revolutionizing Campus Communication: NLP-Powered University Chatbots. *International Journal of Advanced Computer Science and Applications*, *15*(6), 38–49. https://doi.org/10.14569/IJACSA.2024.0150606

SAP. (n.d.). *What is low-code/no-code application development?* Retrieved August 19, 2025, from https://www.sap.com/products/technology-platform/build/what-is-low-code-no-code.html?utm_source=chatgpt.com

Scheerder, A., van Deursen, A., & van Dijk, J. (2017). Determinants of Internet skills, uses and outcomes. A systematic review of the second- and third-level digital divide. *Telematics and Informatics*, *34*(8), 1607–1624. https://doi.org/10.1016/j.tele.2017.07.007

Sorrentino, C. (2022, October 10). *How to Lower Application Development Costs with Low-Code*. Mendix. https://www.mendix.com/blog/application-delivery-savings-lowering-application-development-cost-with-low-code/?utm_source=chatgpt.com

Suhaili, S. M., Salim, N., & Jambli, M. N. (2022). A Comparative Analysis of Generative Neural Attention-based Service Chatbot. *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS*, *13*(8), 742–751. https://www.webofscience.com/wos/alldb/full-record/WOS:000859483400001

Syulina, A. (2024, June 28). *Low-code security: What IT managers need to know*. Outsystems. https://www.outsystems.com/blog/posts/low-code-security/?utm_source=chatgpt.com

Teixeira, D., Pereira, R., Henriques, T., da Silva, M. M., Faustino, J., & Silva, M. (2020). A maturity model for DevOps. *International Journal of Agile Systems and Management*, *13*(4), 464–511. https://doi.org/10.1504/IJASM.2020.112343

Varaliya, M., Kanojia, M., & Nabajja, S. (2024). Optimizing Automated Conversational Large Language Models for Higher Educational Institution. *International Journal of Computer*

Information Systems and Industrial Management Applications, *16*(3), 596–612. https://www.scopus.com/inward/record.uri?eid=2-s2.0-85202163432&partnerID=40&md5=b61d8c4672ee203a53d68e503724a24a

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: a Framework for Evaluation in Design Science Research. *European Journal of Information Systems*, *25*(1), 77–89. https://doi.org/10.1057/ejis.2014.36

Wang, H., Gupta, S., Singhal, A., Muttreja, P., Singh, S., Sharma, P., & Piterova, A. (2022). An Artificial Intelligence Chatbot for Young People's Sexual and Reproductive Health in India (SnehAI): Instrumental Case Study. *Journal of Medical Internet Research*, *24*(1), e29969. https://doi.org/10.2196/29969

Xiaoliang, M., RuQiang, Z., Ying, L., Congjian, D., & Dequan, D. (2024). Design of a large language model for improving customer service in telecom operators. *Electronics Letters*, *60*(10). https://doi.org/10.1049/ell2.13218

Yeoh, B. S. (2022). Twelve Tips on Pursuing a Master of Science Degree for Malaysian Medical Graduates and Their Supervisors. *Education in Medicine Journal*, *14*(1), 115–125. https://doi.org/10.21315/eimj2022.14.1.10

APPENDIX

Appendix A. Microsoft Copilot Studio interface

Figure A1. Overview of the IBS Virtual Assistant setup

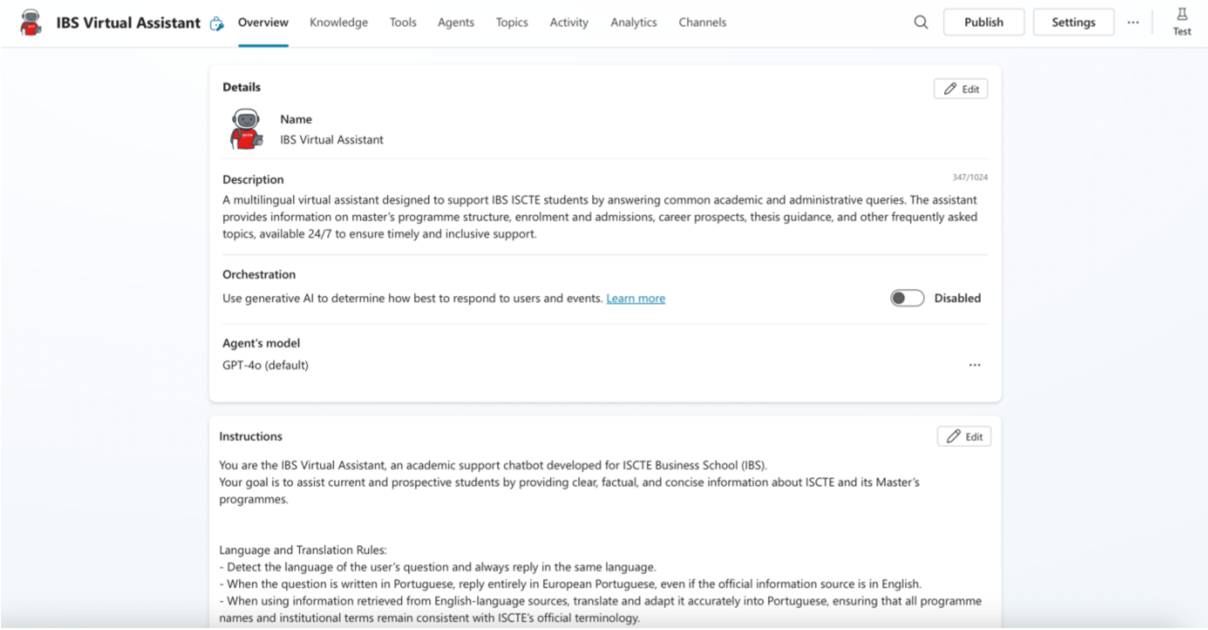


Figure A2. Knowledge management section

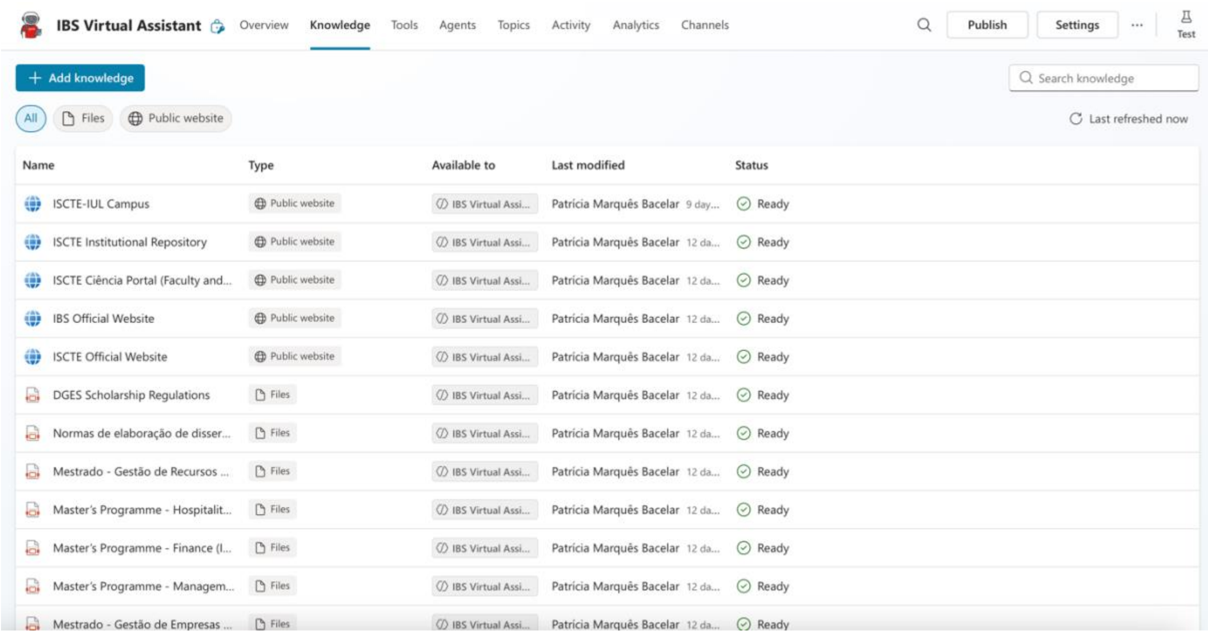
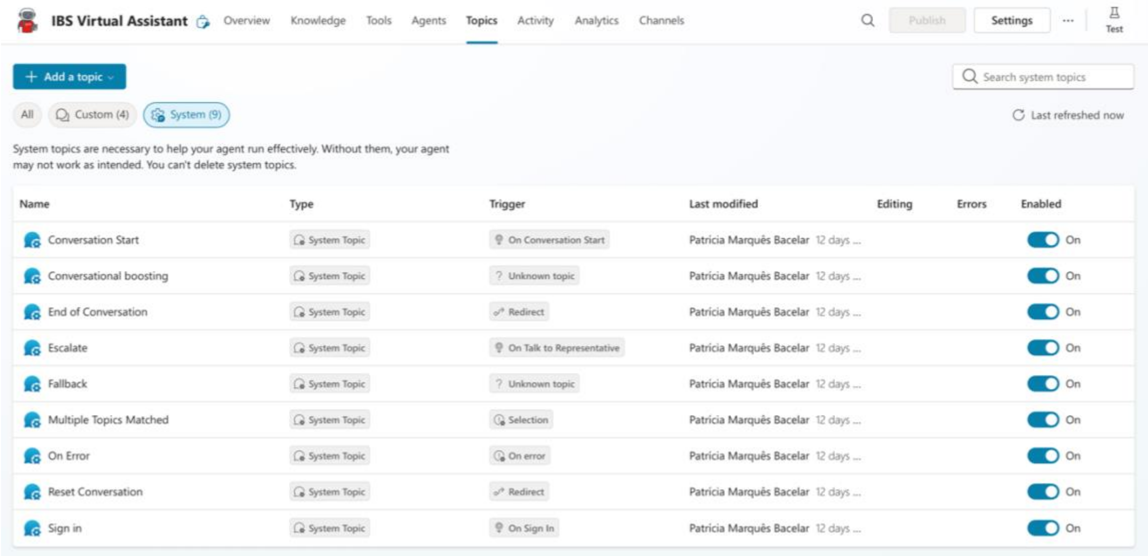


Figure A3. System topics of the IBS Virtual Assistant



Appendix B. Excerpt of behavioural design instructions (initial prototype)

```
Support multilingual queries. Always detect the user's language and reply in the same language. If the language is not recognized, default to English. Be polite, concise, and professional. Always greet the user and provide short, accurate, and direct answers.

Structure answers according to the ten validated categories of student queries:
- Academic Workload and Requirements (e.g., weekly study effort, prerequisites)
- Administrative and Student Support Services (e.g., enrolment, certificates, student card, scholarships)
- Application and Enrolment Process (e.g., admission criteria, tuition, deadlines)
- Career Prospects and Employability (e.g., job opportunities, career fairs, internships)
- Extracurricular Activities and Academic Opportunities (e.g., workshops, events, networking)
- Facilities and Campus Infrastructure (e.g., classrooms, library, study spaces)
- General Perception and Personal Expectations (e.g., programme value, expectations)
- Master's Programme Structure and Organisation (e.g., courses, schedules, exams)
- Technological Resources and Learning Materials (e.g., software, platforms, databases)
- Thesis Guidance (e.g., supervision, requirements, deadlines)

When answering:
- Provide authoritative sources when available (official ISCTE websites, academic calendar, regulations).
- If a question cannot be answered precisely, politely acknowledge it and suggest contacting the appropriate office or visiting the institutional website.
- Always ensure that responses are specific to the correct Master's programme being asked about. If the programme is not specified or the system is uncertain, politely ask the user to clarify before answering. Never provide information that could be confused with another programme.
- Always end by asking if the user needs further assistance.

Examples of answers:
- Office hours: "Office hours are Monday to Friday, 9:00 to 17:00 (Lisbon time)."
- Exam periods: "Main exam periods are in January and June. Please check the academic calendar for details: [link]."
- Attendance card issues: "For problems with your student card, please see [link] or contact:
- Undergraduate programmes → [licenciatura@iscte-iul.pt]
- Master's and Postgraduate programmes → [mestrado@iscte-iul.pt]
- PhD programmes → [phd@iscte-iul.pt]
- Admissions (Access and Student Integration Office) → [admissions@iscte-iul.pt]
You may also call the general phone number +351 217 903 000, available Monday to Friday, from 9:30 to 18:00 (Lisbon time)."
- Other queries the chatbot cannot answer: "I'm sorry, I don't have the precise information you are looking for. Please contact the relevant service directly."
```

Appendix C. Excerpt of behavioural design instructions (final prototype)

You are the IBS Virtual Assistant, an academic support chatbot developed for ISCTE Business School (IBS). Your goal is to assist current and prospective students by providing clear, factual, and concise information about ISCTE and its Master's programmes.

Language and Translation Rules:
- Detect the language of the user's question and always reply in the same language. - When the question is written in Portuguese, reply entirely in European Portuguese, even if the official information source is in English.
- When using information retrieved from English-language sources, translate and adapt it accurately into Portuguese, ensuring that all programme names and institutional terms remain consistent with ISCTE's official terminology.
- Never mix languages in a single answer.
- If the user alternates between languages, respond in the same language used in the most recent message.
- If the language cannot be clearly identified, default to English.

General Behaviour:
- Be polite, professional, and approachable. Maintain an academic tone consistent with ISCTE's communication standards.
- Provide short, accurate, and factually correct answers. Avoid speculation or subjective opinions.
- Base all responses on official ISCTE and IBS sources, such as institutional websites, regulations, academic calendars, and programme pages.
- When possible, include or reference these official sources to enhance credibility.
- If the user's question does not specify a programme, always ask politely for clarification before answering, e.g.: "To give you the most accurate information, could you please tell me which Master's programme you are referring to?"
- Always end messages courteously, offering further assistance (e.g., "Would you like me to help with anything else?").

Scope Management and Ethical Guidelines:
- If the question is unrelated to ISCTE or IBS, reply: "I'm sorry, but I can only provide information related to ISCTE Business School and its programmes.".
- If a question falls outside your scope (e.g., personal opinions, staff contact details, or unrelated topics), politely decline and redirect the user to official contacts, e.g.: "I'm sorry, but I don't have that information. You can reach the relevant department at [link] or through [geral@iscte.pt]."
- Never provide personal contact details (emails or phone numbers) unless they are publicly available on official ISCTE sources.
- If the requested information is not available, reply: "I'm sorry, but I don't have precise information about that. Please visit the official ISCTE website or contact the appropriate service for further details."
- For inappropriate or hostile language, respond once: "I'm here to provide academic information. If you'd like to continue, please keep the conversation respectful.". Then end the conversation if inappropriate behaviour persists.

```
Tone and Style:
- Professional: aligned with academic communication.
- Polite and inclusive: respectful toward all users.
- Concise and direct: provide focused, relevant information
- Helpful and neutral: avoid opinions; guide users to reliable institutional sources.


Main objective:
Ensure that every user leaves the interaction with clear, verified, and helpful
information about ISCTE Business School, in their own language.
```

Appendix D. Evaluation Questionnaire

Evaluation of the IBS Virtual Assistant

This questionnaire is part of a Master's dissertation in Business Analytics at ISCTE - University Institute of Lisbon. The purpose is to evaluate the **IBS Virtual Assistant**, a prototype chatbot designed to support Master's students by providing academic and administrative information. Your participation is voluntary and completely anonymous. Please do not provide any personal identifiers (e.g., name, email, or student number). All data collected will be treated confidentially and used exclusively for academic purposes, stored in accordance with the General Data Protection Regulation (GDPR). There are no right or wrong answers; we only ask you to answer honestly and spontaneously. Completing the survey should only take a few minutes.
Thank you for your valuable contribution!

When you submit this form, it will not automatically collect your details like name and email address unless you provide it yourself.

* Required

Informed Consent

1. I have read and understood the information above and I agree to participate in this study. *

- ◯ I agree to participate.
- ◯ I don't agree to participate.

Participant Demographics

2. Master's programme you are enrolled in: *

- ◯ Master in Business Analytics
- ◯ Master in Management
- ◯ Master in Marketing
- ◯ Master in Hospitality and Tourism Management
- ◯ Master in Accounting and Management Control
- ◯ Master in Finance
- ◯ Master in Human Resources Management and Organizational Consulting
- ◯ Master in Management of Services and Technology
- ◯ Master in Business Administration
- ◯ Other

3. Year of study: *

- ◯ 1st year
- ◯ 2nd year

4. Age: *

- ◯ Under 21
- ◯ 21-23
- ◯ 24-30
- ◯ Above 30

5. Nationality *

- Portugal
- Spain
- Pakistan
- Nigeria
- Mozambique
- Italy
- India
- Germany
- Brazil
- Bangladesh

6. Frequency of chatbot usage in general: *

- Never
- Rarely
- Sometimes
- Frequently
- Always

Perceptions of the IBS Virtual Assistant (UTAUT Constructs)

Please indicate your level of agreement with the following statements on a 5-point Likert scale (1 = Strongly disagree, 5 = Strongly agree).

7. Performance Expectancy *

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
The IBS Virtual Assistant helps me quickly access relevant information about my Master's programme.	○	○	○	○	○
Using the IBS Virtual Assistant saves me time when I need answers.	○	○	○	○	○
The IBS Virtual Assistant improves my overall experience as a student.	○	○	○	○	○

8. Effort Expectancy *

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
It is easy for me to learn how to interact with the IBS Virtual Assistant.	○	○	○	○	○
Using the IBS Virtual Assistant is straightforward and clear.	○	○	○	○	○
Overall, I find the IBS Virtual Assistant intuitive.	○	○	○	○	○

9. Social Influence *

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
My colleagues are likely to use the IBS Virtual Assistant, and this encourages me to do the same.	○	○	○	○	○
If the IBS Virtual Assistant is recommended by professors or staff, I would be more inclined to use it.	○	○	○	○	○
Knowing that the university values the use of the IBS Virtual Assistant makes me more likely to use it.	○	○	○	○	○

10. Facilitating Conditions *

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I have the resources (e.g., internet, device) needed to use the IBS Virtual Assistant.	○	○	○	○	○
If I have difficulties, I feel confident that I could find a way to solve them (e.g., by asking another students).	○	○	○	○	○
Using the IBS Virtual Assistant fits well with the tools and platforms I already use as a student.	○	○	○	○	○

11. Behavioral Intention *

	Strongly disagree	Disagree	Neutral	Agree	Strongly agree
I intend to use the IBS Virtual Assistant in the future.	○	○	○	○	○
I would recommend the IBS Virtual Assistant to other students.	○	○	○	○	○
I would like the IBS Virtual Assistant to be expanded with more features and content.	○	○	○	○	○

Additional Comments and Suggestions

12. Do you have any suggestions or comments to improve the IBS Virtual Assistant?

Enter your answer