



EVALUATING ARABIC CHATBOTS USING TRANSFORMER MODELS AND TRANSFER LEARNING TECHNIQUES

Prepared By Mohammed Naeem



JULY 24, 2025

CONTENTS

Problem statement.....	2
Research Objectives.....	2
Literature review.....	3
Summary of research papers:	3
Paper 1: Arabic chatbot technologies: A scoping review.....	3
Paper 2: TOWARDS USER-CENTRIC GUIDELINES FOR CHATBOT CONVERSATIONAL DESIGN	3
Paper 3: Chatbots for Robotic Process Automation	3
Paper 4: Exploring consumers' response to text-based chatbots in e-commerce.....	3
Paper 5: Evaluation of an Arabic Chatbot Based on Extractive Question-Answering Transfer Learning and Language Transformers	4
Paper 6: Metrics of Success: Evaluating User Satisfaction in AI Chatbots.....	4
Critical Analysis:	4
Connections.....	4
Conflicts between the findings.....	5
Synthesis Matrix for Literature Review	5
References	6

PROBLEM STATEMENT

Arabic chatbots are still not performing well compared to English ones. This is because Arabic is a complex language, and there are not many tools or datasets made for it. Many current Arabic chatbots are built with simple rules, so they don't always understand users or give correct answers. This makes users lose trust in the chatbot and feel unsatisfied.

In recent years, transformer models like AraBERT and AraElectra became available. These models, especially when combined with transfer learning, can help Arabic chatbots improve. Some research already tested these models and found that they give more accurate and confident answers (Alruqi & Alzahrani, 2023). However, not many studies focus on evaluating these models deeply, especially for chatbot use. Also, few works compare different transformer models to see which one gives better performance in real conversations.

Another issue is that most evaluations only focus on technical results like accuracy or confidence score. But users also care about how useful, natural, and clear the chatbot responses are. These parts are not always studied.

Because of these gaps, this research will evaluate Arabic chatbots that use transformer models and transfer learning. It will compare their performance and also look at how they affect user satisfaction and trust.

RESEARCH OBJECTIVES

The goal of this research is to evaluate how transformer models with transfer learning can improve Arabic chatbots.

To reach this goal, the study will:

1. To identify the transformer models used in Arabic chatbots and how transfer learning is applied.
2. To compare the performance of different transformer models (like AraBERT and AraElectra) in answering Arabic questions.
3. To measure how well these models perform using accuracy, confidence score, and similarity score.
4. To examine how users feel about chatbot answers by looking at satisfaction and trust levels.
5. To suggest which model gives the best balance between technical performance and good user experience.

LITERATURE REVIEW

SUMMARY OF RESEARCH PAPERS:

PAPER 1: ARABIC CHATBOT TECHNOLOGIES: A SCOPING REVIEW

SUMMARY

The paper explores different studies on Arabic-language chatbots and shows the gap in their development compared to English.

PAPER 2: TOWARDS USER-CENTRIC GUIDELINES FOR CHATBOT CONVERSATIONAL DESIGN

SUMMARY

This paper provides user-centered guidelines for designing effective chatbot conversations based on a systematic literature review. It emphasizes key design elements and explores how they affect trust, satisfaction, and engagement, and then confirms this through surveys.

PAPER 3: CHATBOTS FOR ROBOTIC PROCESS AUTOMATION

SUMMARY

This paper studies how chatbot design affects trust and user satisfaction in tasks related to Robotic Process Automation (RPA). The researchers tested this with 277 people using four tasks with different levels of difficulty. They found that when the task was harder, people trusted the chatbot less and felt less satisfied. Users liked short and clear replies more than long answers. Also, users were more willing to use chatbots for simple tasks, but for difficult tasks, they preferred help from a human.

PAPER 4: EXPLORING CONSUMERS' RESPONSE TO TEXT-BASED CHATBOTS IN E-COMMERCE

SUMMARY

This study looks at how people respond to text-based chatbots in online shopping. It focuses on how personal, helpful, and human-like the chatbot feels, and how these affect trust, satisfaction, and the chance of buying something. The results show that when chatbots give personalized and human-like answers, users are more interested and likely to buy. But if the chatbot feels too automatic and not relevant, it can make the experience worse.

PAPER 5: EVALUATION OF AN ARABIC CHATBOT BASED ON EXTRACTIVE QUESTION-ANSWERING TRANSFER LEARNING AND LANGUAGE TRANSFORMERS

SUMMARY

This paper looks at Arabic chatbots that use transformer models like AraBERT and AraElectra with transfer learning for answering questions. The researchers tested the models using two Arabic question-and-answer datasets. They found that the AraElectra-SQuAD model gave the best results, with high accuracy and confidence. The study shows that transformer models can make Arabic chatbots better, especially in giving answers that match the context..

PAPER 6: METRICS OF SUCCESS: EVALUATING USER SATISFACTION IN AI CHATBOTS

SUMMARY

This paper presents a way to measure how satisfied users are with AI chatbots. It looks at things like how useful the chatbot is, how easy it is to use, how it makes the user feel, and how good the answers are. The authors used both surveys and chatbot interaction data to test their method. The paper shows that user satisfaction is not only about finishing tasks, but also about how the user feels during the chat. It also says that chatbot evaluation should match the situation where the chatbot is used.

CRITICAL ANALYSIS:

The papers have different goals and methods. Some papers like “Metrics of Success” and “Consumers’ Response” focus more on user satisfaction and how users interact with chatbots, but they don’t look at Arabic users or different languages. On the other hand, the “Evaluation of an Arabic Chatbot” paper uses new AI models and gives good results in Arabic, but it talks more about technical scores and not much about the user experience.

The paper about robotic process automation shows how task difficulty can change user trust and satisfaction. It is helpful, but it doesn’t study different cultures or languages. Also, “User-Centric Guidelines” gives useful design ideas, but it does not test them with real users in practice.

In general, some papers are strong in technology, and others are better in user feedback, but most do not combine both. More research is needed that works for Arabic users and checks both chatbot quality and how people feel using them.

CONNECTIONS

All the papers are related to chatbots and how to improve them. Most of them talk about user experience, design, or performance. For example, many papers try to make chatbots more helpful and easy to use. Some papers focus on Arabic chatbots, while others look at English ones, but they all want to make chatbots better for users.

Also, some papers use AI models like transformers or machine learning to improve chatbot answers. Others try to study what users think and how they react when using a chatbot. Even if they use different methods, they all aim to improve chatbot communication and usefulness.

CONFLICTS BETWEEN THE FINDINGS

Some papers focus more on technical performance, while others care more about user satisfaction. For example, the paper about Arabic chatbot with transformers shows high accuracy, but it does not study how users feel when using the chatbot. On the other hand, papers like “*Metrics of Success*” and “*Consumers’ Response*” talk about user experience but do not test the chatbot's technical side.

Also, some papers believe that chatbot success comes from strong design and clear answers, while others think it depends more on how the user feels during the conversation. These different opinions show that there is no agreement yet on what makes a chatbot successful.

SYNTHESIS MATRIX FOR LITERATURE REVIEW

Paper Title	Authors	Year	Methodology	Key Findings	Limitations
1. Arabic Chatbot Technologies: A Scoping Review (1)	• Arfan Ahmed , • Nashva Ali • Mahmood Alzubaidi , • Wajdi Zaghouani • Alaa Abd-alrazaq , • Mowafa Househ	5 April 2022	Scoping review of Arabic chatbot research	Shows big gap in Arabic chatbot tools, datasets, and research	No model implementation or evaluation provided, no discussion about Arabic accents.
2. Towards User-Centric Guidelines for Chatbot Conversational Design (2)	• Geovana Ramos Sousa Silva • Edna Dias Canedo	16 Jan 2023	Literature review + expert feedback	Suggests design principles to make chatbot conversations better	No real testing with users
3. Chatbots for Robotic Process Automation (3)	• Alessandro Casadei • Stephan Schlögl • Markus Bergmann	1 Feb 2023	Conceptual framework + literature insights	Chatbots help reduce task complexity and improve efficiency	Focused more on business tasks, not general chatbot use

4. Exploring Consumers' Response to Text-Based Chatbots in E-commerce⁽⁴⁾	• Xusen Cheng • Ying Bao • Alex Zarifis • Wankun Gong • Jian Mou	19 May 2021	User experiment	Users trust chatbots more when they use human-like language	Study limited to English and e-commerce only
5. Evaluation of an Arabic Chatbot Based on Transfer Learning and Transformers⁽⁵⁾	• Tahani N. Alruqi • Salha M. Alzahrani	16 August 2023	Model testing (AraBERT, AraElectra, etc.)	AraElectra-SQuAD showed strong performance in Arabic QA tasks	Only technical evaluation, no user testing
6. Metrics of Success: Evaluating User Satisfaction in AI Chatbots⁽⁶⁾	• Møller, Cecilie Grace • Ang, Ke En; • Bongiovanni, Maria de Lourdes • Khalid, Md Saifuddin • Wu, Jiayan	2025	Survey + scale development	Developed a validated tool to measure chatbot satisfaction	Not applied to Arabic or technical chatbot models

REFERENCES

1. Ahmed A, Ali N, Alzubaidi M, Zaghouani W, Abd-alrazaq A, Househ M. Arabic chatbot technologies: A scoping review. *Computer Methods and Programs in Biomedicine Update* [Internet]. 2022 Jan 1 [cited 2025 Jul 24];2:100057. Available from: <https://www.sciencedirect.com/science/article/pii/S2666990022000088?via%3Dihub>
2. Silva GRS, Canedo ED. Towards User-Centric Guidelines for Chatbot Conversational Design. *Int J Hum Comput Interact* [Internet]. 2024 [cited 2025 Jul 24];40(2):98–120. Available from: <https://www.tandfonline.com/doi/abs/10.1080/10447318.2022.2118244>
3. Casadei A, Schlogl S, Bergmann M. Chatbots for Robotic Process Automation: Investigating Perceived Trust and User Satisfaction. *Proceedings of the 2022 IEEE International Conference on Human-Machine Systems, ICHMS 2022*. 2022;
4. Cheng X, Bao Y, Zarifis A, Gong W, Mou J. Exploring consumers response to text-based chatbots in e-commerce: The moderating role of task complexity and chatbot disclosure. *Internet Research* [Internet]. 2024 Jan 20 [cited 2025 Jul 24];32(2):496–517. Available from: <http://arxiv.org/abs/2401.12247>
5. Alruqi TN, Alzahrani SM. Evaluation of an Arabic Chatbot Based on Extractive Question-Answering Transfer Learning and Language Transformers. *AI 2023*, Vol 4, Pages 667-691 [Internet]. 2023 Aug 16 [cited 2025 Jul 24];4(3):667–91. Available from: <https://www.mdpi.com/2673-2688/4/3/35/htm>

6. Møller CG, Ang KE, de Lourdes Bongiovanni M, Khalid MS, Wu J. Metrics of Success: Evaluating User Satisfaction in AI Chatbots. ICAAI 2024 - Conference Proceedings of the 2024 8th International Conference on Advances in Artificial Intelligence [Internet]. 2025 Mar 3 [cited 2025 Jul 24];168–73. Available from: <https://dl.acm.org/doi/pdf/10.1145/3704137.3704182>