The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM)
ISSN: 2602-3199

The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 2024

Volume 27, Pages 50-62

IConTech 2024: International Conference on Technology

Chatbot Development: Framework, Platform, and Assessment Metrics

Wassem Saad Nsaif
University of Diyala

Huda M. Salih
University of Diyala

Hassan Hadi Saleh
University of Diyala

Bashar Talib Al-Nuaimi
University of Diyala

Abstract: It can be difficult for developers to select the best solution for their projects due to the abundance of chatbot development platforms and frameworks. This paper explores the selection of frameworks and platforms for designing chatbots, based on criteria from numerous scientific articles. The introduction covers the axes and sections of the paper, including frameworks, platforms, metrics, and paper details. The second section reviews previous studies on the topic, examining frameworks and platforms used, metrics, and other details. An expansion of software-related services devoted to chatbot development has resulted from the necessity for these services to be produced in large quantities quickly and effectively. The third section examines the latest frameworks and platforms, various sources of articles and scientific research published in prestigious international databases. Large corporations compete with one another and offer comprehensive chatbot development platforms include Google, Microsoft, Amazon, and IBM. We also talk about chatbot platform and measures of evaluation framework while showcasing successful industrial practices. The fourth section proposes methodologies for choosing frameworks or platforms based on findings from numerous scientific research, master's and doctoral theseis, and important scientific books by prominent authors. The fifth section discusses the criteria for measuring chatbot efficiency and the best frameworks and platforms according to these metrics. Scholars, developers, and businesses are given recommendations that point to potential areas for further research and development in this rapidly evolving section. The final section presents the conclusions, listing details and section mentioned in the paper, and a list of references, including about a hundred references from prestigious scientific articles. This scientific paper provides individuals, groups, and large and small companies with mental and intellectual enlightenment, helping them make decisions on their chatbot designing by choosing the most appropriate frameworks and platforms.

Keywords: Conversational agents, Natural language processing (NLP), Chatbot

Introduction

Recently, the world has witnessed a wide spread of chatbots in various fields of work, due to their ease of use 24/7 without getting tired or bored. chatbots respond to the questions and inquiries of users and provide suggestions and solutions to them. This is done by understanding natural languages (NLU) and processing them using natural language processing (NLP). There are also many varied platforms and frameworks available for building and designing chatbots suit to various specialization fields in the labor market (Abd-Alrazaq, 2020) Companies noticed the effectiveness of the chatbot, its ease of use, and the satisfaction of clients and customers

with its performance, which prompted them to competition to develop platforms and frameworks for developing chatbots. Thus, it became important to carefully choose between these platforms and frameworks for building a chatbot and integrating it with applications (Suhaili, 2021) The rate of increase in online conversations with chatbots over the past few years has attracted the attention of artificial intelligence (AI) and machine learning (ML) developers, as well as the diversity of its areas of use in health, education, entertainment, customer support, etc., in addition to the possibility of integrating it with other applications such as Siri, Messenger, Watson, etc. (Radziwill, 2017). In light of the rapid emergence of chatbots, the search for criteria for evaluating it has become an urgent issue, and there are no accurate metrics for evaluating its performance. Existing studies primarily focus on technical aspects, lacking human or business perspectives and exploring motivators (Io, 2017). Although there is a lot of research on chatbot development and design, there is still a big knowledge vacuum about user motivation (Brandtzaeg, 2017). Goal-oriented dialogue systems have previously been evaluated using TRAINS, PARADISE, SASSI, and MIMIC. PARADISE is the most popular framework for evaluating chatbots; it measures subjective aspects including user happiness, clarity, and ease of use (Venkatesh, 2018). In this regard, developing a methodical process for assessing and contrasting various chatbot frameworks and platforms becomes essential. By using evaluation criteria in an organized manner, chatbot development experts can make accurate decisions that are consistent with their goals (Gupta, 2022). This study attempts to offer recommendations on the ideal framework or platform for developing chatbots by incorporating knowledge from the literature already in existence and industry best practices. We will cover the many assessment metrics and criteria that developers should take into account as well as the important elements that affect platform selection through a thorough analysis of pertinent research articles, case studies, and industry reports. In order to choose the most appropriate platforms and frameworks, we examined these proposed methodologies to supply programmers and developers with wide insights that will help them create a chatbot and facilitate their decision-making process in choosing the best and most appropriate platform or framework for building and designing their chatbot. Using a methodical search of the Scopus, Elsevier ,Web of Science, IEEE, and Science Direct platforms, the study examined current chatbot assessment measures with an emphasis on the computer science, telecommunications, education, and engineering domains. Papers addressing chatbot assessment, evaluation measures, and quality factors were deemed pertinent. A combination of qualitative and quantitative measurements were found when prior research on chatbot evaluation was refined. The majority of evaluation metrics, which were based mostly on user satisfaction surveys, lacked quantitative components. In Hung (2009) assessed LifeLike's efficacy and naturalness using the PARADISE technique, paying particular attention to efficiency and quality costs in dialog performance. The functioning of commercial chatbots, including their appearance, implementation, speech synthesis, knowledge base, conversational skills, context sensitivity, personality, customisation choices, and user rating, was examined in Kuligowska, (2015). The study concentrated on quality factors and was subjective. A paradigm for assessing chatbots based on user involvement, topic coverage, consistency, content variety, and conversation depth was presented by Venkatesh (2018) to objectively evaluate chatbot performance, the metrics were combined. Using user feedback as a baseline, the approach's validity was confirmed by the significant correlation observed between the suggested measures and user evaluations. In Chakrabarti (2013) experts evaluated the conversational agents, factual accuracy, information sufficiency, manner, and relation. In [(Jwalapuram, (2017, September))] Grice's Maxims were employed to assess the conversations and rated chatbot conversations using a Likert scale. In Shawar (2007) using qualitative methods to assess the efficacy, quality, and user satisfaction of conversations by analyze the relationship between user input and chatbot responses. In Cahn, (2017) and Hussein et al. (2020) the study assesses chatbot performance from multiple angles, including information retrieval, user experience, linguistics, Using user feedback.

Chatbot Framework and Platform

There are two kinds of tools for creating chatbots: frameworks and platforms. Frameworks provide developers greater control and freedom when building chatbots from scratch by offering tools, libraries, and standards. Developers that have the resources and programming abilities to oversee the complete development process can use them. Conversely, platforms provide a whole ecosystem for creating, implementing, and overseeing chatbots without requiring a lot of coding knowledge or technical know-how. They provide pre-built templates, chat platform connections, and user-friendly interfaces. Development is made easier by the fact that platforms frequently come with NLP, analytics, and deployment options pre-installed. On the other hand, compared to using frameworks directly, users can have less customization and freedom. The decision between a framework and a platform is based on particular needs and level of experience. Developers can create and implement chatbots more effectively by using a chatbot framework, which is a collection of guidelines and tools.

Table 1. Platforms and frameworks

Platform	Features	Suitablefor	Pricing
Chtfuel	A/B testing, AI/ML, alerts/notifications, auto-responders, automated publishing, automated scheduling, and campaign management are all provided by Chatfuel.	Chatfuel's user interface is simple to use, and can quickly build up a chatbot using it. The fact that each chatbot flow is exclusive to a single channel and cannot be duplicated across channels is the only negative could detect.	Enterprise plans start at $300 and provide customized plans, priority support, and a dedicated bot building specialist. Business plans start at $14.39 and include 500 monthly talks plus $0.03 additional conversations each interaction.
anychat	Popular for building and administering chatbots on Facebook Messenger, ManyChat provides capabilities including broadcasting, audience segmentation, custom design, and connectivity with marketing tools.	Creating bots for sales, support, and marketing is possible with ManyChat, a top Facebook Messenger marketing solution. ManyChat gives all the growth tools needed to turn anyone into a subscriber, so you can easily expand the Messenger audience.	Manychat has a range of business options, which include bespoke features, specialized automation specialists, sophisticated Pro features beginning at $15/mo, and free basic capabilities.
I.B.M	Central repository for enterprise vocabulary that helps users with governance initiatives, asset comprehension, compliance, and Working With IT Teams.	Enhancing decision-making confidence for businesses by providing accurate, consistent, and complete information definition and organization.	With a committed use account or subscription upgrade, the service provides discounted rates on over 350 products, a 30-day credit, and free access to over 40 services. pay-as-you-go options are also available.
Google-Dialog-Flow	With features like a visual flow builder and omnichannel implementation, dialogflow expedites the creation of generative aı agents, cutting down on development time and enhancing conversational capabilities.	Use text virtual agents to quickly and accurately respond to common inquiries and provide specific information while interacting with customers on their preferred platform at any time and from any location.	Monthly pricing for dialogflow is determined by edition and number of requests. a $600 credit is given to new customers for a risk-free trial that lasts for a year.)
Azure AI Bot Service	With power virtual agents, developers can build conversational aı bots without writing any code thanks to a fully hosted low-code platform.	With the least amount of code modifications, the aı system generates conversational interactions for clients by combining natural language, discourse, and vision.	The cost of the azure AI Bot service is determined by how many messages are sent over premium channels (standard channels are free).

Amazon-Lex	Rich insights and dashboards are provided by amazon lex, an aı service that creates, develops, tests, and launches conversational interfaces. ıt also integrates With AWS Lambda.	Scalable, safe, and easy to use, amazon lex works with aws lambda to solve deep learning issues such as language comprehension and speech recognition.	Amazon lex is a service that offers voice and text conversational interfaces for applications, processing up to 10,000 text requests per month without any upfront commitment or minimum fee.
R.A.S.A	AI/machine learning software is appropriate for developers and pre-configured bots because it can modify behavior based on data, assist in intent recognition, and process natural language for contextual guidance.	Millions Of Developers Use Rasa open source, rasa x, and rasa enterprise to create conversational aı applications for small teams and large enterprises alike.	Rasa Io plus and pro plans offer customizable features for small to mid-sized businesses, simplifying newsletter management and saving time with varying pricing.
WIT. AI	Multimodal interaction across multiple platforms is made possible by the platform, which facilitates the development, testing, and deployment of free, open, and extensible natural language experiences.	Signing Up, creating a wit app, improving detection, querying it, adding new intents, and adding entities to capture more data are all possible with Facebook.	Wit is available for free, even for business use. thus, our terms apply to both private and public wit apps, which are free.
PANDORA-BOT	A comprehensive artificial intelligence software for businesses and startups, pandorabots offers complete Windows Solutions, İncluding Chatbots And Multilingual Features.	With its easy-to-use platform, pandorabots makes it possible to quickly deploy chatbots and virtual assistants without requiring complex infrastructure management.	The $199/month pro plan offers live training, email, chat, phone support, an unlimited message limit, chat widget Access, API Access, Development Sandbox, And Third-Party Channels.
BOT-PRESS	Use multiple conversations, css, and react to customize the gui messages that your bot sends. you can host on aws s3 or bpfs, share files via shareable links, and set and delete conversations privately.	Botpress offers an easy-to-use interface for creating chatbots, with pre-built themes and plugins that make customization and functionality enhancement simple.	No cost with restrictions, obtain $5 credit every month for aı spend, five bots, two thousand incoming messages or events per month, three partners, five thousand table rows, one hundred
SAP-CONVERSATIONAL-AI	Make use of our cutting-edge natural language processing technology to interpret text inputs, enhance important information, and construct flexible bots that can speak different languages. you can configure triggers, parameters, and responses right within the bot-builder.	Using Cutting-Edge NLP technology and low-code features for quicker development, the platform offers an intuitive user interface for training, developing, testing, connecting, and monitoring chatbots integrated into sap and third-party solutions.	Model Of Pricing: Free, Trial Offer: Obtainable

Google Gemini AI	Google created Google Gemini AI, formerly known as Bard, as an AI chatbot tool to mimic human discussions using machine learning and natural language processing.	Because Google Gemini takes strong privacy and confidentiality precautions, such as severing communications from users' accounts before reviewers can access them, it is safe.	For individuals 18 years of age and older who have a personal Google Account or Google Workspace account with admin access allowed, Google Bard AI, now known as Gemini, is available for free.

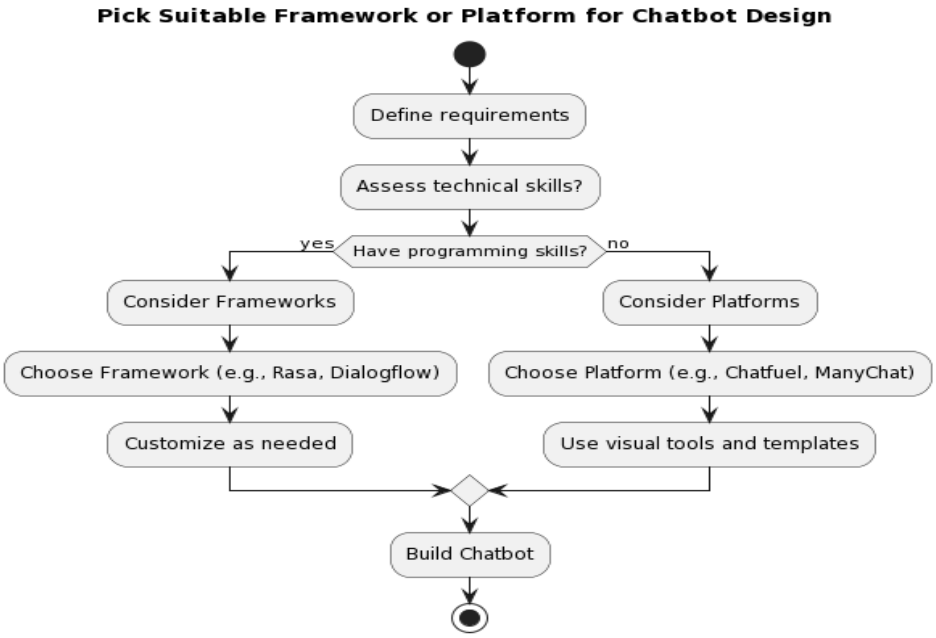

Figure 1. Path for picking up the framework or platform.

These frameworks provide developers with defined techniques and reusable code for basic tasks, freeing them up to concentrate on personalizing the chatbot's behavior and connecting it with certain applications. Microsoft Bot Framework, Google's Dialogflow, and Rasa are popular frameworks (Wei, 2018). Chatbot frameworks and platforms are distinct tools for developing and deploying chatbots. Frameworks offer a foundation for customizing chatbots, requiring coding expertise, while platforms offer an end-to-end solution with a user-friendly interface and pre-built components, making them suitable for individuals with varying technical expertise (Framework., 2019). Microsoft Bot Framework is an AI chatbot framework made for communicating and interacting with customers. It can be taught with current conversations and Azure cognitive services, and it connects with well-known Microsoft programs like Cortana and Office 365. With help of the free and open-source (NLP) API Wit.ai, companies may develop voice- and text-based chatbots. It extracts useful information using machine learning methods and supports multiple languages. Google owns Dialog Flow, a system that translates speech into text to digitize corporate operations and save time and money. It has special voice navigation functions. It enables automatic human-computer interaction through speech-to-text and natural language dialogues (Thorat, 2020). Neural networks (NN) are used in the DeepQA project I.B.M. Watson to provide organically processed responses. It is widely used in healthcare facilities to diagnose possible illnesses and prescribe appropriate care and medications. These frameworks are made to function as dynamic dialogue-flowing question-answering platforms that let organizations get information and crucial data. With the help of the AIaaS platform Pandorabots, companies can easily create, implement, and refine chatbots. To process human language organically, it makes use of the Artificial Intelligence Modelling Language (AIML) and the (A.L.I.C.E) Artificial Linguistic Internet Computer Entity. Pandorabots includes SDKs for Java, Node.js, Python, Ruby, PHP, and Go and supports GIFs and static images. Voice interfaces, eCommerce, customer support, and marketing have all made use of it (Følstad, (2021)). Botpress is a chatbot available under two licenses that provides enterprises with a modular blueprint to add new features to pre-existing code frames. Businesses can use their choice cloud hosting service and create chatbots locally with its three-step installation method. More than 7,000 developers use the Botkit platform, which includes integrated LUIS.ai (NLU).

Numerous plugins, open source libraries, a visual builder for conversations, integrated analytics and statistics, and a free edition are all available. A sophisticated collection of open-source machine learning tools called RASA Stack can be used to build assistants and chatbots. Among its characteristics are conversational functions, interactive and supervised (ML), and (NLU) (Pérez-Soler, (2020, October)). From the table1 below: there are several chatbot systems and frameworks on the market, each with unique features, cost structures, and intended user bases. Several well-liked choices are Amazon Lex, Microsoft Bot Framework, I.B.M Watson-Assistant, Google-Cloud's Dialog-flow, and Facebook-Wit.ai. Building conversational interfaces is made easy with Dialogflow, a free platform that provides natural language comprehension, messaging platform integration, and simple deployment. Pay-as-you-go Amazon Lex is a service that can build conversational interfaces based on voice and text that can be integrated with messaging apps and AWS services. Although the Microsoft Bot Framework is free, using Azure services and resources for the creation and implementation of bots may cost money. An AI-powered platform called IBM Watson Assistant is used to create, train, and use conversational bots on a variety of platforms.

Proposed Methodologies

The combination of AI-driven skills with chatbot frameworks is what will define chatbots in the future. (AI) chatbot frameworks, such as those for (ML), (NLU), and contextual understanding, are making it possible for chatbots to react more human-like and comprehend user inputs more precisely (Makatchev, 2010). Through enhanced personalization, adaptability, and dynamic dialogue management, this connection offers a more smooth and intuitive user experience. In order to prove concepts using appropriate frameworks, it is preferable to create prototypes in order to evaluate their applicability. It is also preferable to make modifications to the suggestions and knowledge collected in the stages of testing the prototypes. The success of a chatbot depends on evaluating frameworks and platforms (ElGibreen, 2020). For companies, frameworks and platforms have become extremely important because they help in developing advanced chatbots in all fields. Choosing the most appropriate among the frameworks and platforms depends on the needs of the company or institution and provides many specifications to achieve the required goals (Qaffas, 2019; Al-Khazraji, 2023). There are a number of criteria through which the best framework and platform for developing a chatbot can be determined, including cost, accuracy, ease of use, and compatibility with other applications (Denecke, 2020). We designed a comparison table above to highlight the most prominent comparisons between the most important frameworks and distinct platforms in developing chatbots. The optimal option is determined by the technical capabilities, financial restrictions, and project needs.

Chatbot Metrics

One of the important issues in evaluating the performance of a chatbot is user comments, ratings, and opinion polls (Balaji, 2019). When a chatbot's responses are accurate to a certain level, this is considered an important metric, and this is done using one of these metrics (F1 score), (recall), and (precision) (Goodman, (2023)). As is the case with humans, clear understanding and accurate interpretation are also important in a chatbot through understanding connotations, intentions, and entities. This is done by (NLU) and is considered a measure of the extent of understanding (Cañizares, 2022). When the user interacts, continues the conversation, and prolongs the dialogue with the chatbot, this is considered an important measure and is called the (Engagement metrics) (Schuetzler, 2020). When a chatbot is asked to perform a specific task and it completes it, the extent of completion, accuracy of completion, speed of completion, and other considerations related to this task, such as the error rate and success rate, are all considered a evaluate of the success chatbot (Schumaker, 2007). There are responses from a chatbot that require personalization ,so the extent of the chatbot’s flexibility in personalizating answers and adapting and being flexible with them is an important metric for a chatbot (Abd-Alrazaq,2020). Classification of some chatbot features based on quality, efficiency, and satisfaction in an analytical hierarchy process that has been proposed in order to evaluate chatbot performance. (Radziwill, 2017). User input and chatbot responses were evaluated using the correction rate and response satisfaction criteria to evaluate mixed dialogue systems and pure dialogue systems (Schumaker, 2007). To evaluate features such as human support, language diversity, command integrity through the use of helper commands such as cancellation, typos, key keywords and synonyms, as well as usage and response times, all of these are presented as criteria to evaluate the quality of the chatbot. (Pereira, 2018). Using a question or command that is already known to be answered to test a chatbot's responsiveness was presented as a criterion for measuring the efficiency of a chatbot. Assessors tallied mistakes and examined students' proficiency in grammar, spelling, and vocabulary. The acceptability of the responses was graded from both the grammatical and semantic perspectives (Coniam, 2014). Five categories were found in Brandtzaeg and Følstad's research on people's reasons for utilizing chatbots:

productivity, amusement, social/relational, novelty/curiosity, and other reasons. According to the study, users expect social and enjoyable interactions, while productivity was identified as the primary motivator. They also underlined the necessity of more research and the need for chatbots to offer useful and practical information (Brandtzaeg, 2017). In his 2017 study, Zamora looked at how people expected and perceived chatbots. With the chatbot, participants discussed their habits, observations, and experiences. The study discovered that a small vocabulary impedes communication and that delicate subjects should not be covered. Emotional needs were stated as the top priority, while privacy concerns around improper data handling were voiced (Zamora, 2017). Assessing chatbot effectiveness and important metrics indicators including job completion rates, customer satisfaction scores, and user engagement are crucial for gauging the efficacy of a chatbot when analyzing its performance. For more information see Table 2: Chatbot Metrics. It is crucial to comprehend how to calculate these metrics and analyze the information in order to consistently enhance and optimize chatbot functionality (Peras,2018). In the ever-evolving chatbot development ecosystem is essential to building chatbots that provide users with meaningful and value interactions. Developers can make well-informed decisions that result in the production of more efficient and user-friendly chatbots by studying comparative assessments of platforms and tools and comprehension of the evolution of chatbot capabilities. For chatbot deployment ,it's critical to take into account the chatbot deploying aspect like Facebook Messenger, WhatsApp, Slack, and Telegram that developers may decide which one is best for their particular chatbot project (Nuruzzaman, 2018).

Table 2. Chatbot metrics

Sorting by Category	Metrics	Articles
Metric with an automatic basis	F1-Score	(Zhang, 2018; Bashir, 2018; Alshammari, 2022; Nuruzzaman, 2020),
	ROUGE	(Omoregbe, 2020; Zhang, 2018; Kapočiūtė-Dzikienė, 2020; Hori, 2019).
	Accuracy	(Boussakssou, 2022;Peng, 2020; Wael, 2021;Wijaya, 2020; Niculescu, 2020; Grosuleac, 2020; Alshammari,2022)
	Recall	(Mai, 2021; Omoregbe, 2020; Zhang et al.,2018)
	Precision	(Boussakssou, 2022; Peng, 2020; Wael, 2021;Wijaya, 2020; Niculescu, 2020; Grosuleac, 2020; Alshammari, 2022; Mai, 2021; Omoregbe , 2020; Zhang et al., 2018)
	BLEU	(Yang, 2018; Aleedy, 2019; Palasundram, 2019; Alshareef, 2020; Kim, 2020;Tran, 2019;Zhou, 2020;Kim et al.,2019),
	PPL	(Song, 2021; Zhang et al., 2020; Wu, 2018)
	MRR, MAP, and P@1	(Prassanna, 2020; Liu, 2020; Candra, 2019)
	Skip thoughts cosine similarity, BOW and greedy matching scores, vector extreme cosine similarity, embedding average cosine similarity	(Mai, 2021; Omoregbe, 2020; Zhang et al.,2018)
	Other	(Hu, 2018; Sajjapanroj, 2020; Mohialden, 2021; Mavridis, 2011)
NA		(Roca, 2020; Zahour, 2020;Ranavare, 2020; Alotaibi, 2020; Kasinathan, 2020;Vanjani, 2019)
Metric based on human	H: User Satisfaction	(Hijjawi, 2014; Noori, 2014; Sweidan, 2021; Octavany, 2020;El Hefny, 2021;Al-Ajmi, 2021; Chete, 2020; Oguntosin, 2021; Mageira, 2022)

To create meaningful and interesting interactions, it is essential to understand user preferences and satisfaction levels. As is the case in all other applications, the development of a chatbot will necessarily make it important to choose platforms or frameworks for its design based on their advantages and disadvantages, taking into account the rapid development in (NLU) (Dagkoulis, 2022). There are many metrics that measure the coherence and fluency of the responses generated, and among these measures is the (F1 Score), which combines recall, accuracy, and the confusion measure. (Yuwono, 2019). In order to evaluate the quality of the text created by the chatbot in terms of comparing response units and text references, a metric called (ROUGE) Recall-Oriented Understudy for Gisting is used (Duong,2022). To measure one of the forms of accuracy known (precision at K (P@K)) as the percentage of responses. To evaluate the effectiveness of retrieval systems, a special metric

called Mean Average Precision (MAP) is used (Gu, 2019). To rank relevant responses, a measure called the Mean Reciprocal Rank (MRR) is used to evaluate how well the retrieval system classifies relevant responses. These automated metrics offer insightful information on a range of topics related to chatbot behavior, such as response relevancy, correctness, completeness, and fluency (Singh, 2021). The methods used by different chatbot tactics differ, and metrics are essential to assessing their efficacy. While human-based metrics offer qualitative evaluations of user happiness and interaction quality, automatic-based analytics offer quantitative insights (Naous,2020).

Regarding to table 2 above, metrics that are based on automatic processes showen with related articles infront of each metric are essential for assessing chatbot performance in an automated manner. These metrics include the fluency and coherence of generated responses measured by the Perplexity measure, and the precision and recall measured by the (F1-Score). Additional metrics are (ROUGE), which evaluates the quality of generated text or summaries by analyzing the overlap between (n-grams), and (BLEU), which compares n-grams in machine-generated text to those in a reference answer. Information retrieval metrics like (MAP), Precision at K, and (MRR) are used to assess how well a retrieval system is working. These measures, which include response relevance, accuracy, completeness, and fluency, offer insightful information on chatbot behavior. They are essential for comparing alternative systems and statistically evaluating chatbot performance, offering insightful information on a range of chatbot behavior and Metric based on human, User Satisfaction.

Conclusion

This scientific paper discusses the selection of frameworks and platforms for designing chatbots, focusing on criteria that have been thoroughly examined through numerous scientific articles. The introduction covers the axes and section of the paper, including frameworks, platforms, metrics, and paper details. The second section reviews and delves into previous studies on the topic, examining frameworks and platforms used, metrics, and other details related to the research. The third section examining the latest frameworks and platforms, various sources of articles and scientific research published in prestigious international databases such as Scopus, Web Science, Clarvit, natural Science, and IEEE. The fourth section which are this article's base, proposes methodologies for choosing frameworks or platforms for designing a chatbot, based on our findings from numerous scientific research, master's and doctoral theseis, and important scientific books by prominent authors in the section. The fifth section discusses the criteria for measuring the efficiency of a chatbot, which are this article's core, and the best frameworks and platforms according to these metrics. The final section presents the conclusions, listing the details and section mentioned in this paper, explaining each section briefly and with great clarity. The paper concludes with a list of references, including More than a hundred references of the important articles that are related. This scientific paper provides individuals, groups, and large and small companies in the governmental, private, and mixed sectors with mental and intellectual enlightenment, helping them make decisions on their chatbot designing by choosing the most appropriate frameworks and platforms.

Scientific Ethics Declaration

The authors declare that the scientific ethical and legal responsibility of this article published in EPSTEM Journal belongs to the authors.

Acknowledgements or Notes

* This article was presented as an oral presentation at the International Conference on Technology (www.icontechno.net) held in Alanya/Turkey on May 02-05, 2024.

References

Abd-Alrazaq, A., Safi, Z., Alajlani, M., Warren, J., Househ, M., & Denecke, K. (2020). Technical metrics used to evaluate health care chatbots: Scoping review. *Journal of Medical Internet Research, 22*(6), e18301.

Adiwardana, D., Luong, M. T., So, D. R., Hall, J., Fiedel, N., Thoppilan, R. & Le, Q. V. (2020). Towards a human-like open-domain chatbot. *arXiv preprint arXiv:2001.09977.*

Al-Ajmi, A. H., & Al-Twairesh, N. (2021). Building an Arabic flight booking dialogue system using a hybrid rule-based and data driven approach. *IEEE Access, 9*, 7043-7053.

Aleedy, M., Shaiba, H., & Bezbradica, M. (2019). Generating and analyzing chatbot responses using natural language processing. *International Journal of Advanced Computer Science and Applications, 10*(9), 60-6.

Al-Ghadhban, D., & Al-Twairesh, N. (2020). Nabiha: An Arabic dialect chatbot. *International Journal of Advanced Computer Science and Applications*, *11*(3).

Aljameel, S., O'Shea, J., Crockett, K., Latham, A., & Kaleem, M. (2019). LANA-I: An Arabic conversational intelligent tutoring system for children with ASD. In *Intelligent Computing: Proceedings of the 2019 Computing Conference, 1*, 498-516.

Al-Khazraji, S. H., Saleh, H. H., Khalid, A. I., & Mishkhal, I. A. (2023). Impact of deepfake technology on social media: detection, misinformation and societal implications. *The Eurasia Proceedings of Science Technology Engineering and Mathematics, 23*, 429-441.

Al-Madi, N. A., Maria, K. A., Al-Madi, M. A., Alia, M. A., & Maria, E. A. (2021). An intelligent Arabic chatbot system proposed framework. In *2021 International Conference on Information Technology (ICIT)* (pp. 592-597). IEEE.

Almurtadha, Y. (2019). LABEEB: Intelligent conversational agent approach to enhance course teaching and allied learning outcomes attainment. *Journal of Applied Computer Science & Mathematics*, *13*(27), 9-12.

Alobaidi, O. G., Crockett, K. A., O'Shea, J. D., & Jarad, T. M. (2013). Abdullah: An intelligent arabic conversational tutoring system for modern islamic education. In *Proceedings of the World Congress on Engineering*, *2*, 762-768.

Alotaibi, R., Ali, A., Alharthi, H., & Almehamdi, R. (2020). AI chatbot for tourist recommendations: A case study in the city of Jeddah, Saudi Arabia. *International Journal of Interactive Mobile Technologies, 14*(19), 18-29.

Alshammari, N. O., & Alharbi, F. D. (2022). Combining a novel scoring approach with arabic stemming techniques for arabic chatbots conversation engine. *Transactions on Asian and low-resource language Information Processing, 21*(4), 1-21.

Alshareef, T., & Siddiqui, M. A. (2020). A seq2seq neural network based conversational agent for gulf arabic dialect. *In 2020 21st International Arab Conference on Information Technology (ACIT)* (pp. 1-7). IEEE.

Alsheddi, A. S., & Alhenaki, L. S. (2022). English and arabic chatbots: A systematic literature review. *International Journal of Advanced Computer Science and Applications, 13*(8),662-675.

Balaji, D. (2019). *Assessing user satisfaction with information chatbots: a preliminary investigation* (Master's thesis, University of Twente).

Bashir, A. M., Hassan, A., Rosman, B., Duma, D., & Ahmed, M. (2018). Implementation of a neural natural language understanding component for Arabic dialogue systems. *Procedia Computer Science*, *142*, 222-229.

Boussakssou, M., Ezzikouri, H., & Erritali, M. (2022). Chatbot in Arabic language using seq to seq model. *Multimedia Tools and Applications, 81*(2), 2859-2871.

Brandtzaeg, P. B., & Følstad, A. (2017). Why people use chatbots. *In Internet Science: 4th International Conference, INSCI 2017* (pp. 377-392). Springer International Publishing.

Cahn, J. (2017). *Chatbot: Architecture, design, & development*University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science

Candra, H., & Riki, R. (2019). Designing a chatbot application for student information centers on telegram messenger using fulltext search boolean. *bit-Tech: Binary Digital-Technology*, *2*(2), 71-80.

Cañizares, P. C., Pérez-Soler, S., Guerra, E., & de Lara, J. (2022, April). Automating the measurement of heterogeneous chatbot designs. In *Proceedings of the 37th ACM/SIGAPP Symposium on Applied Computing* (pp. 1491-1498).

Chakrabarti, C., & Luger, G. F. (2013). A framework for simulating and evaluating artificial chatter bot conversations. In *The Twenty-Sixth International FLAIRS Conference.*

Chete, F. O. (2020). An approach towards the development of a hybrid chatbot for handling studentsâ€™ complaints. *Journal of Electrical Engineering, Electronics, Control and Computer Science*, *6*(4), 29-38.

Coniam, D. (2014). The linguistic accuracy of chatbots: Usability from an ESL perspective. *Text & Talk*, *34*(5), 545-567.

Dagkoulis, I., & Moussiades, L. (2022). A comparative evaluation of Chatbot development platforms. In *Proceedings of the 26th Pan-Hellenic Conference on Informatics* (pp. 322-328).

Daniel, G., Cabot, J., Deruelle, L., & Derras, M. (2019). Multi-platform chatbot modeling and deployment with the Jarvis framework. In *Advanced Information Systems Engineering: 31st International Conference CAiSE 2019* (pp. 177-193). Springer International Publishing.

Denecke, K., Vaaheesan, S., & Arulnathan, A. (2020). A mental health chatbot for regulating emotions (SERMO)-concept and usability test. *IEEE Transactions on Emerging Topics in Computing*, *9*(3), 1170-1182

Duong, K. (2022). *A Re-examination of chatbot evaluation metrics (*Master's thesis).

El Hefny, W., Mansy, Y., Abdallah, M., & Abdennadher, S. (2021). Jooka: A bilingual chatbot for university admission. In *Trends and Applications in Information Systems and Technologies*, (Vol., 39, pp. 671-681). Springer International Publishing.

ElGibreen, H., Almazyad, S., Shuail, S. B., Al Qahtani, M., & ALhwiseen, L. (2020). Robot framework for anti-bullying in Saudi schools. In 2020 *Fourth IEEE International Conference on Robotic Computing (IRC)* (pp. 166-171). IEEE.

Følstad, A., & Taylor, C. (2021). Investigating the user experience of customer service chatbot interaction: a framework for qualitative analysis of chatbot dialogues. *Quality and User Experience*, *6*(1), 6.

Goodman, R. S., Patrinely, J. R., Stone, C. A., Zimmerman, E., Donald, R. R., Chang, S. S., ... & Johnson, D. B. (2023). Accuracy and reliability of chatbot responses to physician questions. *JAMA Network Open*, *6*(10), e2336483-e2336483.

Grosuleac, A., Budulan, S., & Rebedea, T. (2020). Seeking an empathy-abled conversational agent. In *RoCHI* (pp. 103-107).

Gu, J. C., Ling, Z. H., & Liu, Q. (2019). Utterance-to-utterance interactive matching network for multi-turn response selection in retrieval-based chatbots. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, *28*, 369-379.

Gupta, P., Rajasekar, A. A., Patel, A., Kulkarni, M., Sunell, A., Kim, K., & Trivedi, A. (2022, December). Answerability: A custom metric for evaluating chatbot performance. In *Proceedings of the 2nd Workshop on Natural Language Generation, Evaluation, and Metrics* (GEM) (pp. 316-325).

Hijjawi, M., Bandar, Z., & Crockett, K. (2015). A novel hybrid rule mechanism for the Arabic conversational agent ArabChat. *Global Journal on Technology*, (8), 185-194.

Hijjawi, M., Bandar, Z., & Crockett, K. (2016). The enhanced Arabchat: An Arabic conv. agent. *International Journal of Advanced Computer Science and Applications*, *7*(2), 332-347.

Hijjawi, M., Bandar, Z., Crockett, K., & Mclean, D. (2014). ArabChat: An arabic conversational agent. In *2014 6th International Conference on Computer Science and Information Technology (CSIT)*, 227-237, IEEE.

Hijjawi, M., Qattous, H., & Alsheiksalem, O. (2015). Mobile Arabchat: An Arabic mobile-based conversational agent. *IJACSA*, *6*(10), 111-119.

Hori, T., Wang, W., Koji, Y., Hori, C., Harsham, B., & Hershey, J. R. (2019). Adversarial training and decoding strategies for end-to-end neural conversation models. *Computer Speech & Language*, *54*, 122-139.

Hu, T., Xu, A., Liu, Z., You, Q., Guo, Y., Sinha, V., ... & Akkiraju, R. (2018). Touch your heart: A tone-aware chatbot for customer care on social media. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

Hung, V., Elvir, M., Gonzalez, A., & DeMara, R. (2009). Towards a method for evaluating naturalness in conversational dialog systems. In *2009 IEEE International Conference on Systems, Man and Cybernetics* (pp. 1236-1241). IEEE.

Hussein, A. A., Hussein, K. M., Saleh, H. H., & Farhan, I. H. (2020). Survey\towards a sustainable information and communication technologies (ICT) in Iraq. *Journal of Physics: Conference S*eries, 1*530*(1), 012089. IOP Publishing.

Io, H. N., & Lee, C. B. (2017, December). Chatbots and conversational agents: A bibliometric analysis. In 2017 IEEE *International Conference on Industrial Engineering and Engineering* Management (IEEM) (pp. 215-219). IEEE.

Jwalapuram, P. (2017). Evaluating dialogs based on Grice's maxims. In *Proceedings of the Student Research Workshop Associated with RANLP* (pp. 17-24).

Kapočiūtė-Dzikienė, J. (2020). A domain-specific generative chatbot trained from little data. *Applied Sciences*, *10*(7), 2221.

Kasinathan, V., Abd Wahab, M. H., Idrus, S. Z. S., Mustapha, A., & Yuen, K. Z. (2020). Aira chatbot for travel: Case study of AirAsia. *Journal of Physics: Conference Series 1529*(2), 022101.

Kim, J., Oh, S., Kwon, O. W., & Kim, H. (2019). Multi-turn chatbot based on query-context attentions and dual wasserstein generative adversarial networks. *Applied Sciences*, *9*(18), 3908.

Kim, S., Kwon, O. W., & Kim, H. (2020). Knowledge-grounded chatbot based on dual wasserstein generative adversarial networks with effective attention mechanisms. *Applied Sciences*, *10*(9), 3335.

Kuligowska, K. (2015). Commercial chatbot: performance evaluation, usability metrics and quality standards of embodied conversational agents. *Professionals Center for Business Research, 2*(02),1-16.

Liu, B., & Mei, C. (2020). Lifelong knowledge learning in rule-based dialogue systems. *arXiv preprint arXiv:2011.09811*.

Mageira, K., Pittou, D., Papasalouros, A., Kotis, K., Zangogianni, P., & Daradoumis, A. (2022). Educational AI chatbots for content and language integrated learning. *Applied Sciences, 12*(7), 3239.

Mai, T. N. T., & Maxim, S. (2021). Enhancing Rasa NLU model for Vietnamese chatbot. *International Journal of Open Information Technologies, 9*(1), 31-36.

Makatchev, M., Fanaswala, I., Abdulsalam, A., Browning, B., Ghazzawi, W., Sakr, M., & Simmons, R. (2010). Dialogue patterns of an arabic robot receptionist. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 167-168). IEEE

Mavridis, N., AlDhaheri, A., AlDhaheri, L., Khanii, M., & AlDarmaki, N. (2011). Transforming IbnSina into an advanced multilingual interactive android robot. In *2011 IEEE GCC Conference and Exhibition (GCC)* (pp. 120-123). IEEE

Mohialden, Y. M., Younis, M. T., & Hussien, N. M. (2021). A novel approach to Arabic Chabot, utilizing Google colab and the internet of things: A case study at a computer center. *Webology, 18*(2), 946-954.

Naous, T., Hokayem, C., & Hajj, H. (2020, December). Empathy-driven Arabic conversational chatbot. In *Proceedings of the Fifth Arabic Natural Language Processing* Workshop (pp. 58-68).

Niculescu, A. I., Kukanov, I., & Wadhwa, B. (2020). DigiMo-towards developing an emotional intelligent chatbot in Singapore. In *Proceedings of the 2020 Symposium on Emerging Research from Asia and on Asian Contexts and Cultures* (pp. 29-32).

Noori, Z., Bandarl, Z., & Crockett, K. (2014). Arabic goal-oriented conversational agent based on pattern matching and knowledge trees. *Proceedings of the World Congress on Engineering, 1*(1),207-212.

Nuruzzaman, M., & Hussain, O. K. (2018). A survey on chatbot implementation in customer service industry through deep neural networks. In *2018 IEEE 15th International Conference on e-Business Engineering* (ICEBE) (pp. 54-61). IEEE.

Nuruzzaman, M., & Hussain, O. K. (2020). IntelliBot: A dialogue-based chatbot for the insurance industry. *Knowledge-Based Systems, 196*, 105810.

Octavany, O., & Wicaksana, A. (2020). Cleveree: an artificially intelligent web service for Jacob voice chatbot. *Telkomnika (Telecommunication Computing Electronics and Control), 18*(3), 1422-1432.

Oguntosin, V., & Olomo, A. (2021). Development of an e-commerce chatbot for a university shopping mall. *Applied Computational Intelligence and Soft Computing, 2021*, 1-14.

Omoregbe, N. A., Ndaman, I. O., Misra, S., Abayomi-Alli, O. O., Damaševičius, R., & Dogra, A. (2020). Text messaging-based medical diagnosis using natural language processing and fuzzy logic. *Journal of Healthcare Engineering* , 1-14.

Palasundram, K., Sharef, N. M., Nasharuddin, N., Kasmiran, K., & Azman, A. (2019). Sequence to sequence model performance for education chatbot. *International Journal of Emerging Technologies in Learning (iJET), 14*(24), 56-68.

Peng, D., Zhou, M., Liu, C., & Ai, J. (2020). Human–machine dialogue modelling with the fusion of word-and sentence-level emotions. *Knowledge-Based Systems, 192*, 105319.

Peras, D. (2018). Chatbot evaluation metrics. *Economic and Social Development Book of Proceedings*, 89-97.

Pereira, J., & Díaz, O. (2018, April). A quality analysis of facebook messenger's most popular chatbots. In *Proceedings of the 33rd annual ACM symposium on Applied Computing* (pp. 2144-2150).

Pérez-Soler, S., Guerra, E., & De Lara, J. (2020). Model-driven chatbot development. In *International Conference on Conceptual Modeling* (pp. 207-222). Cham: Springer International Publishing.

Prassanna, J., Nawas, K. K., Jackson, J. C., Prabakaran, R., & Ramanathan, S. (2020). Towards building a neural conversation chatbot through Seq2seq model. *International Journal of Scientific and Technology Research, 9*(3), 1219-1222

Qaffas, A. A. (2019). Improvement of Chatbots semantics using wit. ai and word sequence kernel: Education Chatbot as a case study. *International Journal of Modern Education and Computer Science, 11*(3), 16.

Radziwill, N. M., & Benton, M. C. (2017). Evaluating quality of chatbots and intelligent conversational agents. *arXiv preprint arXiv:1704.04579.*

Ranavare, S. S., & Kamath, R. S. (2020). Artificial intelligence based chatbot for placement activity at college using dialogflow. *Our Heritage, 68*(30), 4806-4814.

Reshmi, S., & Balakrishnan, K. (2018). Empowering Chatbots with business intelligence by big data integration. *International Journal of Advanced Research in Computer Science, 9*(1).

Riek, L. D., Mavridis, N., Antali, S., Darmaki, N., Ahmed, Z., Al-Neyadi, M., & Alketheri, A. (2010, April). Ibn sina steps out: Exploring Arabic attitudes toward humanoid robots. In *Proceedings of the 2nd International Symposium on New Frontiers in Human–Robot Interaction, AISB, 1*,88-94.

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

Sajjapanroj, S., Longpradit, P., & Polanunt, K. (2020). A prototype of Google dialog flow for school teachers' uses in conducting classroom research. *Asian Journal of Distance Education, 15*(2), 133-146.

Schuetzler, R. M., Grimes, G. M., & Scott Giboney, J. (2020). The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems*, *37*(3), 875- 900.

Schumaker, R. P., Ginsburg, M., Chen, H., & Liu, Y. (2007). An evaluation of the chat and knowledge delivery components of a low-level dialog system: The Az-Alice experiment. *Decision Support Systems, 42*(4), 2236- 2246.

Shawar, B. A., & Atwell, E. (2007). Different measurement metrics to evaluate a chatbot system. In *Proceedings of the workshop on bridging the gap: Academic and Industrial Research in Dialog Technologie*s (pp. 89-96).

Singh, D., Suraksha, K. R., & Nirmala, S. J. (2021). Question answering Chatbot using deep learning with NLP. In *2021 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT*) (pp. 1-6). IEEE.

Song, H., Wang, Y., Zhang, K., Zhang, W. N., & Liu, T. (2021). BoB: BERT over BERT for training persona-based dialogue models from limited personalized data. *arXiv preprint arXiv:2106.06169*

Suhaili, S. M., Salim, N., & Jambli, M. N. (2021). Service chatbots: A systematic review. *Expert Systems with Applications, 184*, 115461.

Sweidan, S. Z., Abu Laban, S. S., Alnaimat, N. A., & Darabkh, K. A. (2021). SIAAA-C: a student interactive assistant android application with chatbot during Covid-19 pandemic. *Computer Applications in Engineering Education, 29*(6), 1718-1742.

Sweidan, S. Z., Laban, S. S. A., Alnaimat, N. A., & Darabkh, K. A. (2021). SEG-COVID: a student electronic guide within Covid-19 pandemic. In *2021 9th International Conference on Information and Education Technology (ICIET)* (pp. 139-144). IEEE.

Thorat, S. A., & Jadhav, V. (2020). A review on implementation issues of rule-based chatbot systems. In Proceedings of the *International Conference On Innovative Computing & Communications (ICICC).*

Tran, V. K., & Nguyen, L. M. (2019). Gating mechanism based natural language generation for spoken dialogue systems. *Neurocomputing, 325*, 48-58.

Vanjani, M., Aiken, M., & Park, M. (2019). Chatbots for multilingual conversations. *Journal of Management Science and Business Intelligence, 4*(1), 19-24.

Venkatesh, A., Khatri, C., Ram, A., Guo, F., Gabriel, R., Nagar, A., & Raju, A. (2018). On evaluating and comparing open domain dialog systems. *arXiv preprint arXiv:1801.03625.*

Wael, T., Hesham, A., Youssef, M., Adel, O., Hesham, H., & Darweesh, M. S. (2021). Intelligent Arabic-based healthcare assistant. In *2021 3rd novel intelligent and Leading Emerging Sciences Conference (NILES)* (pp. 216-221). IEEE.

Wei, C., Yu, Z., & Fong, S. (2018). How to build a chatbot: chatbot framework and its capabilities. In *Proceedings of the 2018 10th International Conference on Machine Learning and Computing* (pp. 369-373).

Wijaya, Y., & Zoromi, F. (2020). Chatbot designing information service for new student registration based on AIML and machine learning. *JAIA-Journal of Artificial Intelligence and Applications, 1*(1), 01-10.

Wu, Y., Li, Z., Wu, W., & Zhou, M. (2018). Response selection with topic clues for retrieval-based chatbots. *Neurocomputing, 316*, 251-261

Wu, Y., Wu, W., Li, Z., & Zhou, M. (2018). Learning matching models with weak supervision for response selection in retrieval-based chatbots. *arXiv preprint arXiv:1805.02333.*

Yang, M., Tu, W., Qu, Q., Zhao, Z., Chen, X., & Zhu, J. (2018). Personalized response generation by dual-learning based domain adaptation. *Neural Networks, 103*, 72-82.

Yuwono, S. K., Wu, B., & D'Haro, L. F. (2019). Automated scoring of chatbot responses in conversational dialogue. In *9th International Workshop on Spoken Dialogue System Technology* (pp. 357-369). Singapore: Springer.

Zahour, O., El Habib Benlahmar, A. E., Ouchra, H., & Hourrane, O. (2020). Towards a Chatbot for educational and vocational guidance in Morocco: Chatbot E-Orientation. *International Journal, 9*(2), 2479-2487.

Zamora, J. (2017). I'm sorry, dave, i'm afraid i can't do that: Chatbot perception and expectations. In *Proceedings of the 5th International Conference on Human Agent Interaction* (pp. 253-260).

Zhang, L., Yang, Y., Zhou, J., Chen, C., & He, L. (2020). Retrieval-polished response generation for chatbot. *IEEE Access, 8*, 123882-123890.

Zhang, S., Dinan, E., Urbanek, J., Szlam, A., Kiela, D., & Weston, J. (2018). Personalizing dialogue agents: I have a dog, do you have pets too? *Proccedings of the 56 th Annual Meeting.*

Zhang, Y., Galley, M., Gao, J., Gan, Z., Li, X., Brockett, C., & Dolan, B. (2018). Generating informative and diverse conversational responses via adversarial information maximization. *Advances in Neural Information Processing Systems, 31*.

Zhou, L., Gao, J., Li, D., & Shum, H. Y. (2020). The design and implementation of xiaoice, an empathetic social chatbot. *Computational Linguistics*, *46*(1), 53-93.

Author Information

Wassem Saad Nsaif
Department of Computer Science, Sciences College, University of Diyala, Iraq
Contact e mail: *Wassem.saad@uodiyala.edu.iq*

Huda Mohammed Salih
Department of Computer Engineering, College of Engineering, University of Diyala, Iraq

Hassan Hadi Saleh
Department of Computer Science, College of Education for Pure Science, University of Diyala, Iraq

Bashar Talib Al-Nuaimi
Department of Computer Science, Sciences College, University of Diyala, Iraq

To cite this article:

Nsaif, W.S., Salih, H.M., Saleh, H. H., & Al-Nuaimi, B. T. (2024). Chatbot development: Framework, platform, and assessment metrics. *The Eurasia Proceedings of Science, Technology, Engineering & Mathematics (EPSTEM), 27,* 50-62.