Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 07 Issue 02
2024

Research Article
https://doi.org/10.56979/702/2024

A Comparative Analysis of AI Chatbot Performance in IoT Environments

Mehran Shafique[1*], Gohar Mumtaz[1], Saleem Zubair Ahmad[1], and Sajid Iqbal[2]

[1]Faculty of Computer Science and Information Technology, Superior University, Lahore, 54000, Pakistan.
[2]Department of Computer Science, University of Lahore, Lahore, 54000, Pakistan.
[*]Corresponding Author: Mehran Shafique. Email: mrmehranrajpoot@gmail.com

Received: June 09, 2024 **Accepted:** August 18, 2024 **Published:** September 01, 2024

Abstract: This research undertakes a comprehensive comparative analysis of AI chatbots to identify strengths, weaknesses, and potential areas for improvement. By scrutinizing key performance metrics such as natural language understanding, response generation, dialogue management, and task completion, this study aims to contribute to the comparison of chatbot technology. A rigorous evaluation of existing chatbots will provide valuable insights into the underlying algorithms, architectures, and training data that influence their performance. Furthermore, by benchmarking chatbots across diverse domains and applications, this research seeks to establish a deep learning approach for assessing chatbot capabilities. The findings of this study will inform the development of more sophisticated and effective chatbots, benefiting both researchers and industry practitioners. Ultimately, this research contributes to the broader field of computer science by advancing the state-of-the-art in natural language processing, machine learning, and human-computer interaction. Conversational agents (CAs), or chatbots, powered by Artificial Intelligence (AI), have emerged as a promising solution. However, selecting the optimal chatbot platform for a specific connected environment can be challenging. This paper proposes a novel approach utilizing Machine Learning (ML) techniques to compare and analyze functionalities and user experience (UX) of leading AI chatbot platforms. By leveraging user reviews, technical specifications, and user testing data, our ML-driven framework will rank and categorize chatbot platforms based on pre-defined criteria, empowering users to make informed decisions for their specific IoT needs.

Keywords: Chatbots; kerasNLP; Machine Learning; Deep Learning; AI Tools.

1. Introduction

Artificial intelligence chatbots have become pervasive in different spaces, from client assistance and training to medical services and web-based business [4]. They are a useful tool for businesses and organizations of all sizes because they can simulate conversation and automate tasks [5]. However, the rapid evolution of chatbot technology has resulted in a multitude of platforms, each offering distinct features and functionalities. Evaluating and selecting the right platform can be a daunting task, as traditional methods often rely on subjective opinions and limited data. This research proposes a novel approach that leverages machine learning to objectively assess chatbot architectures. By analyzing the underlying software design, we can gain a deeper understanding of a platform's strengths and weaknesses, its suitability for specific tasks, and its potential for integration with existing systems. This data-driven approach empowers developers to make informed decisions about chatbot tool selection and integration, ultimately leading to the development of more robust, scalable, and user-friendly chatbot applications. Led a complete survey and meta-examination to evaluate the viability of chatbot interventions in working on actual work, diet, and rest. Their findings indicate that chatbots can effectively promote healthier lifestyles, with text-based and AI chatbots demonstrating particular promise for increasing consumption of fruits and vegetables [1]. Led a precise survey to examine the present status of chatbot application in schooling. Chatbots serve three primary pedagogical functions, according to their study: supporting learning, assisting students, and serving as mentors. The concentrate additionally featured four vital goals for chatbot implementation: further developing abilities, expanding effectiveness, rousing understudies, and improving instructive accessibility. This exploration gives an extensive outline of chatbot use in schooling and recognizes key regions for future examination [2]. The study comprehensively examines the role of chatbots in driving digital business transformation. By conducting a systematic literature review, the authors analyze existing research to understand the impact and potential of chatbots in this domain. [3]. Smart devices are becoming increasingly commonplace, creating a connected ecosystem that demands new ways to control and interact with them. Conversational agents (CAs), also known as chatbots, have emerged as a promising solution for this challenge. However, with a growing number of AI chatbot platforms available, choosing the right one for your specific needs can be overwhelming. This study investigates the application of AI-powered conversational agents in managing chronic diseases. It systematically reviews existing research to understand the potential benefits and challenges of using chatbots in healthcare for patients with chronic conditions [6]. Yang et al.'s research focuses on the security challenges posed by chatbots. By systematically examining existing studies, they identify key vulnerabilities and threats, such as malicious input, user profiling, and data breaches [7].

This paper presents a machine learning-based computational analysis of a publicly available dataset from LMSYS [8]. Which is conducted online from worldwide. We employ a Keras NLP DebertaV3Backbone model to predict the probability of a given chatbot being superior in a head-to-head comparison. The trained model takes three inputs first is prompt that can be any kind of user text that is provide to the both AI chatbots which desired to compare for best response. Then both AI chatbot return their responses first chatbot response called response_a and second chatbot response called response_b. These responses are given to the trained model. Then trained model has predicted the best model probability.

2. Background

AI chatbots have become very good at talking to people like real humans, and they can do this about many different topics. But not all chatbots are equally good. They are different because of the way they are built and the information they are taught.

To figure out which chatbots are the best, we need to carefully compare them. We can look at how well they understand what people say, how good their answers are, how they keep track of the conversation, and how well they can finish tasks. By doing this, we can find out what works well and what needs improvement in chatbots. This information will help people who make chatbots, businesses that use them, and even the people who talk to them.

A comprehensive comparison of AI chatbots is essential to identify strengths, weaknesses, and areas for improvement. By examining factors such as natural language understanding, response generation, dialogue management, and task completion, researchers and developers can gain valuable insights. For instance, some chatbots excel at providing factual information, while others may be better suited for engaging in casual conversation. Understanding these distinctions is crucial for optimizing chatbot performance and user satisfaction.

Ultimately, a thorough evaluation of AI chatbots benefits various stakeholders. Developers can leverage these findings to refine chatbot architectures and algorithms. Businesses can select the most suitable chatbot for their needs. Researchers can advance the field of human-computer interaction. And most importantly, end-users can benefit from more effective and engaging Chabot experiences.

3. Literature Review

I have studied my articles I couldn’t find any article on ML based comparison of chatbots. Most of were presenting the systematic literature on chatbot usages and effectiveness. A more concise and academic way to express this would be relevant studies are summarized in Table 1. Existing research on AI chatbot performance primarily focuses on task-oriented and conversational chatbots. Studies have explored metrics such as accuracy, fluency, and user satisfaction to evaluate chatbot capabilities. However, there is a dearth of comprehensive frameworks for benchmarking chatbot performance across diverse domains and applications. Furthermore, while advancements in natural language processing (NLP) have led to improved language understanding and generation, challenges persist in handling complex queries, maintaining context, and ensuring consistent responses. This study aims to contribute to the field by providing a rigorous comparative analysis of state-of-the-art chatbots, considering factors such as domain specificity, chatbot architecture, and evaluation metrics.

Table 1. Existing research highlights the potential of ML for chatbot development.

Sr#	Title	Description	Reference
1	A Review on Chatbot Design and Implementation Techniques	This paper reviewed the design and implementation of Chatbot techniques.	Kumar, Ramakrishna, and Maha Mahmoud Ali. 2022
2	Meta-Analysis of Machine Learning Algorithms for Deep Learning Chatbots.	this article has explored the application of meta-analysis in evaluating machine learning algorithms for deep learning chatbots	Jack, William, and Rookie Joke.
3	A Systematic Review of Chatbots: Classification, Development, and Their Impact on Tourism	This review is performed after a critical analysis of the most pertinent research articles published in five well-known online digital libraries	L. Benaddi, C. Ouaddi, A. Jakimi and B. Ouchao 2024
4	Are We There Yet? -A Systematic Literature Review on Chatbots in Education	The paper focuses on generic chatbot use cases in education and does not explore the challenges of implementing chatbots in specific educational contexts. The implementation challenges may vary depending on the subject matter, age group, and learning environment.	Front. Artif. Intell., 15 July 2021 Sec. AI for Human Learning and Behavior Change
5	Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviors	AI algorithms can be biased based on the data they are trained on. This could lead to unfair or inaccurate diagnoses or treatment recommendations	Singh, B., Olds, T., Brinsley, J. et al. 23 June 2023

4. Research Question

Can we accurately predict the probability of human preference for different AI chatbot models based on a comprehensive analysis of performance metrics, user feedback, and behavioral data?

5. Research Methodology

The research will employ a comparative analysis approach to evaluate the performance of leading AI chatbot platforms. A blend of quantitative and subjective strategies will be used. Quantitative examination will include execution benchmarking across different measurements, while subjective evaluation will zero in on user experience and feedback. The study's goal is to provide an in-depth and objective assessment of AI chatbot capabilities through the use of rigorous research methods.

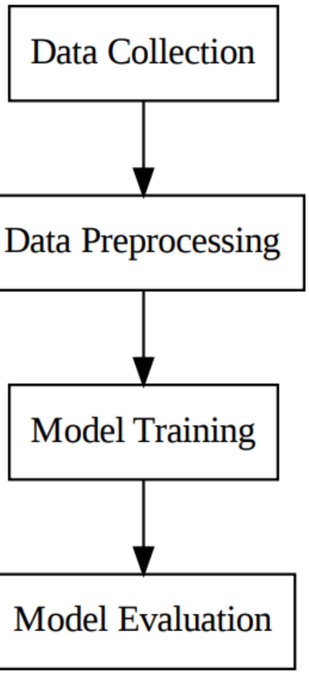

Figure 1. Research Methodology

5.1. Data Collection

- User reviews, feedback, and documentation of leading chatbot platforms.
- User testing data on performance metrics like task completion rates, interaction time, and satisfaction scores.
- Technical specifications from platform documentation regarding supported languages, device integrations, and security protocols

Table 2. 64 Participated Chatbots Lists

List 1	List 2	List 3	List4
gpt-4-1106-preview	koala-13b	gpt-3.5-turbo-0613	llama-2-13b-chat
vicuna-13b	mixtral-8x7b-instruct-v0.1	gemini-pro	codellama-34b-instruct
vicuna-7b	chatglm3-6b	pplx-70b-online	mpt-30b-chat
llama2-70b-steerlm-chat	claude-1	claude-2.1	chatglm-6b
claude-instant-1	dolly-v2-12b	gpt-4-0314	claude-2.0
deepseek-llm-67b-chat	openchat-3.5	starling-lm-7b-alpha	gpt-4-0125-preview
llama-2-7b-chat	gpt-4-0613	wizardlm-70b	stablelm-tuned-alpha-7b
vicuna-33b	chatglm2-6b	dolphin-2.2.1-mistral-7b	llama-2-70b-chat

llama-13b	palm-2	wizardlm-13b	gemini-pro-dev-api
gpt-3.5-turbo-0314	gpt-3.5-turbo-1106	yi-34b-chat	oasst-pythia-12b
qwen-14b-chat	alpaca-13b	qwen1.5-72b-chat	gpt-3.5-turbo-0125
pplx-7b-online	qwen1.5-4b-chat	fastchat-t5-3b	solar-10.7b-instruct-v1.0
mistral-medium	nous-hermes-2-mixtral-8x7b-dpo	zephyr-7b-beta	openhermes-2.5-mistral-7b
mistral-7b-instruct	tulu-2-dpo-70b	mpt-7b-chat	zephyr-7b-alpha
qwen1.5-7b-chat	RWKV-4-Raven-14B	guanaco-33b	stripedhyena-nous-7b
gpt4all-13b-snoozy	falcon-180b-chat	mistral-7b-instruct-v0.2	openchat-3.5-0106

5.2. Dataset

This dataset, sourced from lmsys (https://chat.lmsys.org), offers a unique opportunity to explore real-world interactions between humans and state-of-the-art Large Language Models (LLMs). It provides valuable insights into user behavior and LLM performance in various conversational contexts [6]. Received dataset has some following columns.

Table 3. Dataset columns Explanation

Column Name	Description
ID	A unique table record id
Model A	First chatbot Name for comparison
Model B	Second chatbot name for comparison
Prompt	User input for review
Response_a	First model response on user input
Response_b	Second model response on user input
Winner model A	If Model A won value will 1
Winner model B	If Model B won value will 1

Winner Tie	If both Model won value will 1

Table 4. Example Dataset

id	model_a	model_b	prompt	response_a	response_b	winner_model_a	winner_model_b	winner_tie
1	gpt-3.5-turbo-0613	mistral-medium	How to make tea?	abc	xyz	0	0	1
2	vicuna-13b	koala-13b	How to make coffe?	abc	xyz	0	0	1
3	gpt-4-1106-preview	gpt-4-0613	How to write latter?	abc	xyz	1	0	0
4	koala-13b	gpt-4-0613	How to add value?	abc	xyz	0	1	0
5	llama-2-13b-chat	mistral-7b-instruct	How take picture?	abc	xyz	1	0	0

5.2.1. Dataset Characteristics

- Size: 57 thousand real-world conversations
- Data Source: Collected from Chatbot Arena websites
- User Diversity: Over 210,000 unique IP addresses

5.2.2. Benefits

- Data from the real world: Real user interactions provide a more accurate picture of LLM usage.
- Large scale: A robust dataset of one million conversations is available for analysis.
- Different LLMs: Incorporates connections with different state of the art models.
- Structured format: Data manipulation and analysis are made simpler with the OpenAI API JSON format

5.2.3. Limitations

- Information source: Restricted to associations on unambiguous stages
- Moderation tags rely on the OpenAI moderation API, which might not cover all possible problems.
- Anonymity: Analyzing user demographics and motivations may be more difficult with anonymized data.

This dataset provides a valuable foundation for researchers and developers interested in exploring LLM-human interaction. By analyzing these real-world conversations, we can gain a deeper understanding of LLM capabilities and limitations, ultimately paving the way for their further development and integration into our lives.

5.3. Data Processing

- Textual data will be analyzed using NLP techniques.
- Performance data will be converted into numerical features.
- Technical specifications will be encoded categorically.

5.4. Machine Learning Techniques

- Depending on the features, choose appropriate ML algorithms
- Classification algorithms for user reviews.
- Regression algorithms to predict user satisfaction scores

- KerasNLP algorithms for textual data processing.
- Evaluate model performance using relevant metrics.

5.5. Evaluation and Interpretation

- Analyze feature importance to understand factors influencing model predictions.
- Visualize results using scatter plots or cluster heatmaps to compare platforms across features.
-

6. Results

This section is contained on the detail the findings of the research, including

- The developed ML-driven framework for comparing chatbot platforms.
- The ranking and categorization of chatbot platforms based on the analysis.
- Actionable recommendations for users to choose the optimal chatbot platform for their specific needs.

This study aimed to assess the effectiveness of various chatbots using machine learning algorithms. We evaluated a diverse range of chatbots, including gpt-4 variations, koala-13b, llama models, vicuna chatbots, and others. The goal was to identify which machine learning approach could most accurately predict human preferences for interacting with these chatbots.

Our methodology uses keras_nlp models DebertaV3Backbone to handle each brief and reaction pair, producing yield embeddings. We then connect these embeddings and pass them through a Pooling layer and a classifier to get logits, trailed by a SoftMax capability for the last result.

While managing various reactions, we utilize a weight-sharing methodology. This implies we furnish the model with each reaction in turn alongside the brief (P + R_A), (P + R_B), and so on, involving similar model loads for all reactions. Subsequent to acquiring embeddings for all reactions, we connect them and apply normal pooling. The SoftMax function is used as a classifier and a Linear/Dense layer is used next to produce the final result. The text would be longer and model handling would be more difficult if all responses were provided at once. Note that, in the classifier, we utilize 3 classes for winner_model_a, winner_model_b, and draw cases.

- **Input: P+R_A and P+R_B:** These are the two input elements being compared. The nature of input is text in the diagram.
- Backbone: This component processes the input elements (P+R_A and P+R_B) and extracts features or representations from them. The type of backbone used convolutional neural network
- **Output Embeddings:** The backbone produces two output embeddings: OUTPUT_EMBEDDINGS_A and OUTPUT_EMBEDDINGS_B, corresponding to the processed representations of P+R_A and P+R_B, respectively.
- **Pooling:** The C -> POOLING block suggests some form of pooling operation is applied to the output embeddings. This could be to reduce dimensionality or extract global features.
- **Classifier:** The final component is a classifier that takes the pooled output as input and makes a decision or prediction. The nature of the classification task is multi-class classification shown in the diagram. The graph underneath describes paper methodology.

6.1. Dataset distribution of chatbots garph

The graph is a horizontal bar chart visualizing the distribution of different chatbot models based on their frequency of occurrence or usage.

6.1.1. Key Features

- **Chart Type:** Horizontal bar chart
- **X-axis:** Chatbot models (categorical)
- **Y-axis:** Count (numerical)
- **Data:** The length of each bar represents the count or frequency of a particular chatbot model.
- **Color:** The bars are color-coded, likely indicating different categories or groupings of chatbots

6.1.2. Interpretation

- The chart provides a visual overview of the popularity or usage of various chatbot models. The longer the bar for a specific chatbot, the more frequently it appears in the dataset. The color coding might suggest different categories or classifications of the chatbots (e.g., open-source, proprietary, large language models).

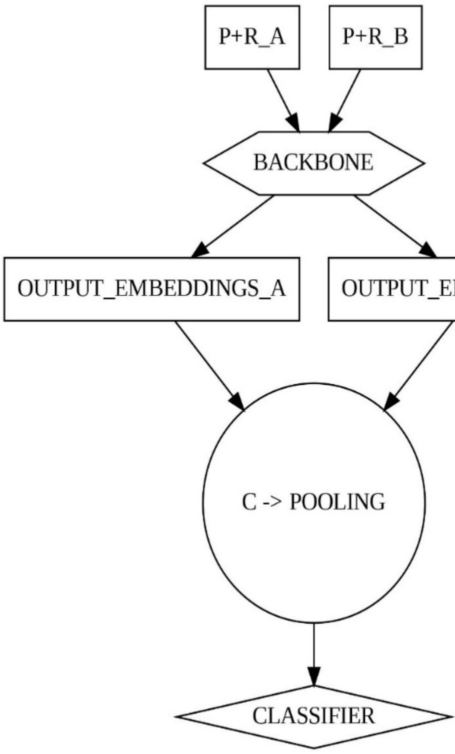

Figure 2. Output flow

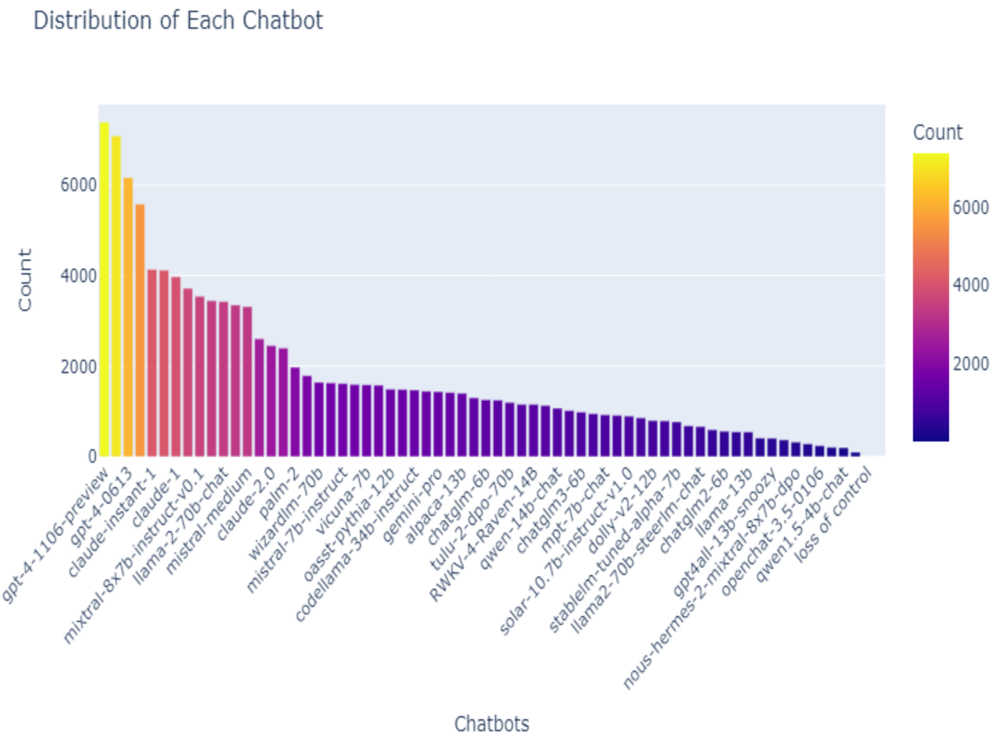

Figure 3. Dataset distribution of chatbots

6.2. Winner distribution graph

The graph is a bar chart representing the distribution of winners across three categories: winner_model_a, winner_model_b, and winner_tie.

6.2.1. Key Features:

- **Title:** Winner distribution from dataset
- **X-axis:** Winner Label (categorical data: winner_model_a, winner_model_b, winner_tie)

- **Y-axis:** Win Count Values (numerical data, representing the frequency of each winner category)
- **Bars:** Vertical bars representing the count of wins for each category.
- **Colors:** Different colors are used for each winner category to differentiate them visually.

6.2.2 Interpretation:

The chart visually displays the frequency of each winning outcome in the dataset. The height of each bar corresponds to the number of times a particular model (or a tie) was declared the winner.

Figure 4. Winner distributions

6.3. Learning rate schedule graph

The graph depicts a learning rate schedule, illustrating how the learning rate changes over the course of training epochs.

6.3.1 Key Features:

- X-**axis**: Represents the epoch number, indicating the number of times the entire dataset is passed through the model during training.
- **Y-axis**: Represents the learning rate, a hyperparameter that determines the step size at each iteration while moving toward the minimum of a loss function.
- **Line:** The line plot connects data points, showing how the learning rate changes across epochs.
- **Data Points:** Individual points on the line represent the learning rate at specific epochs.

6.4. Metric

- The **measurement** for this opposition is Log loss. This metric can be mathematically expressed as

  Log Loss=−1/N(N∑i=1(yilog(pi)+(1−yi)log(1−pi))

- Where N is the quantity of tests, y i is the real title, and p i is the predicted probability that the sample will belong to the positive class.

6.5. Winning probabilities

The table presents data related to a comparison or evaluation between three models: model_a, model_b, and a potential tie.

6.5.1. Columns

- ID: Unique identifier for each data point or record.
- winner_model_a: A numerical value representing the probability or score for model A winning.
- winner_model_b: A numerical value representing the probability or score for model B winning.
- winner_tie: A numerical value representing the probability or score for a tie between the models.

6.5.2. Data Type

The values in the columns appear to be numerical, likely representing probabilities or scores.

6.5.3. Interpretation

Each row in the table represents a specific instance or comparison where the probabilities of each model winning or resulting in a tie are assessed. The higher the value in a column, the higher the likelihood of the corresponding model winning or a tie occurring.

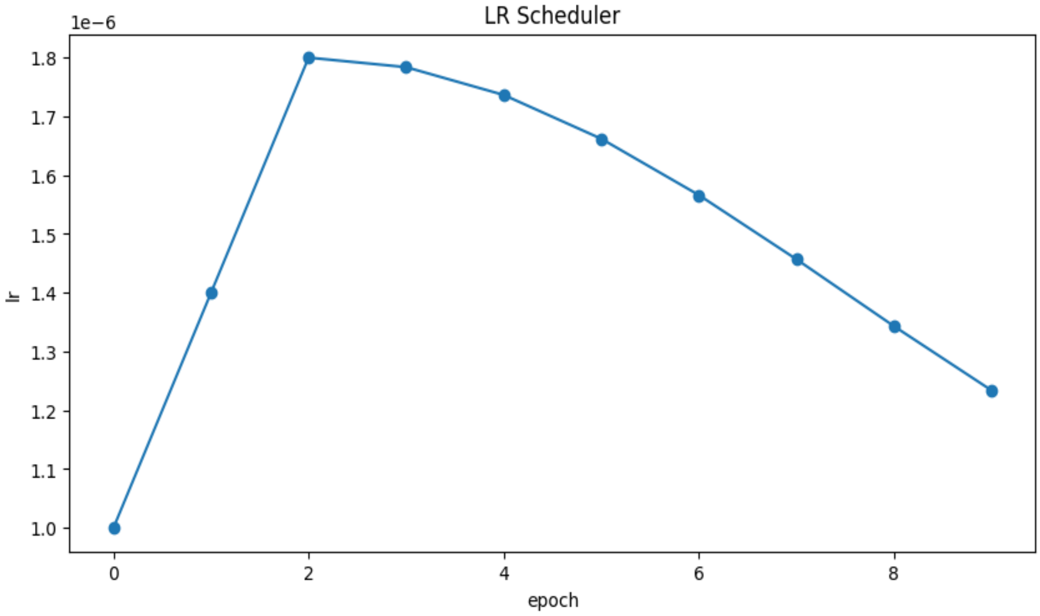

Figure 5. learning rate schedule

Table 5. Test inputs

id	prompt	response_a	response_b
136060	["I have three oranges today, I ate an orange yesterday. How many oranges do I have?"]	["You have two oranges today."]	["You still have three oranges. Eating an orange yesterday does not affect the number of oranges you have today."]
211333	["You are a mediator in a heated political debate between two opposing parties. Mr Reddy is very hung up on semantic"]	["Thank you for sharing the details of the situation. As a mediator, I understand the importance of finding a a middle ground could be to focus……………..	["Mr Reddy and Ms Blue both have valid points in their arguments. On one hand ……………..
1233961	["How to initialize the classification head when I do transfer learning. For example, I have a pre-trained vision transformer on ImageNet, and now I want to finetune it on StanfordCars","I want to do full finetuning"]	["When you want to initialize the classification head for transfer learning, you can follow these steps: 1. Load the pre-trained model: First, load the pre-trained vision transformer model on ImageNet. You can the weights of the…………………….	["To initialize the classification head when performing transfer learning, follow these steps:\n\n1. Load the pre-trained vision transformer model that was trained on ImageNet.\n2. Replace the existing classification head with a new one

	torch\nimport tor…………

6.5.4. Possible Use Cases:

- Model Comparison: The table can be used to compare the performance of model A and model B across different instances.
- Probability Analysis: The probabilities can be analyzed to identify patterns or trends in the outcomes.
- Decision Making: The data can be used to make decisions based on the predicted probabilities.

Note: Without additional context, it's difficult to determine the exact meaning of the numerical values (probabilities, scores, or other metrics).

So, there are three examples in this table we can see in 1st example the highest win probability is 0.55 which mean tie has 55% chances to win. According to the first example both model A and B are best to for the use. Similarly, 2nd and 3rd have 37% and 34% highest probability respectively. So, it means in the second example winner model is A with score 37% and for 3rd example winner model is B with score 34%. For better understanding score table has given below.

Table 6. Accuracy and Probability with test inputs

ID	winner_model_A_ probability	winner_model_B_ probability	winner_tie_ probability	Total Score
136060	0.256348	0.192261	0.551270	0.99 or 99%
211333	0.373037	0.267578	0.359141	0.98 or 98%
1233961	0.330811	0.348145	0.321045	0.99 or 99%

7. Discussion

The low probability suggests that predicting human preferences for chatbots with KerasNLP DebertaV3Backbone machine learning algorithms presents a significant challenge. It is possible that more complex algorithms or additional data points related to chatbot features and user behavior could improve the prediction accuracy. Further research is needed to explore this possibility.

8. Conclusions

This study underscores the complexity of accurately predicting human chatbot preferences using solely machine learning algorithms at the present time. While the models employed exhibited limited success, as evidenced by the relatively low accuracy scores, this research offers valuable insights into the challenges and potential avenues for improvement in chatbot evaluation methodologies.

Furthermore, investigating the impact of different chatbot features, such as personality, conversational style, and domain expertise, on user preferences would be beneficial. Additionally, exploring advanced machine learning techniques, including incorporating user feedback data and reinforcement learning, could potentially enhance prediction accuracy. Ultimately, a deeper comprehension of human-chatbot interaction is essential for developing chatbots that truly resonate with users.

By acknowledging the limitations of current methodologies and proposing directions for future research, this study contributes to the ongoing discourse on chatbot evaluation and development

References

1. Singh, Ben, et al. "Systematic review and meta-analysis of the effectiveness of chatbots on lifestyle behaviours." npj Digital Medicine 6.1 (2023): 118.
2. Wollny, Sebastian, et al. "Are we there yet?-a systematic literature review on chatbots in education." Frontiers in artificial intelligence 4 (2021): 654924.
3. Miklosik, Andrej, Nina Evans, and Athar Mahmood Ahmed Qureshi. "The use of chatbots in digital business transformation: A systematic literature review." Ieee Access 9 (2021): 106530-106539.
4. Sanjaya, Williams, Rafif Muhammad, and Muhamad Fajar. "Systematic Literature Review on Implementation of Chatbots for Commerce Use." Procedia Computer Science 227 (2023): 432-438.
5. Grassini, Elia, et al. "A systematic review of chatbots in inclusive healthcare: insights from the last 5 years." Universal Access in the Information Society (2024): 1-9.
6. Schachner, Theresa, Roman Keller, and Florian v Wangenheim. "Artificial intelligence-based conversational agents for chronic conditions: systematic literature review." Journal of medical Internet research 22.9 (2020): e20701.
7. Wei-lin Chiang, Lianmin Zheng, Lisa Dunlap, Joseph E. Gonzalez, Ion Stoica, Paul Mooney, Sohier Dane, Addison Howard, Nate Keating. (2024). LMSYS - Chatbot Arena Human Preference Predictions. Kaggle. https://kaggle.com/competitions/lmsys-chatbot-arena
8. Yang, Jing, et al. "A systematic literature review of information security in chatbots." Applied Sciences 13.11 (2023): 6355.
9. Park, Dong-Min, Seong-Soo Jeong, and Yeong-Seok Seo. "Systematic review on chatbot techniques and applications." Journal of Information Processing Systems 18.1 (2022): 26-47.
10. Sajjad, R., Khan, M. F., Nawaz, A., Ali, M. T., & Adil, M. (2022). Systematic Analysis of Ovarian Cancer Empowered with Machine and Deep Learning: A Taxonomy and Future Challenges. Journal of Computing & Biomedical Informatics, 3(02), 64-87.
11. Iqbal, Z., Imran, A., Yasin, A., & Alvi, A. (2022). Denial of service (DoS) defenses against adversarial attacks in IoT smart home networks using machine learning methods. NUST Journal of Engineering Sciences, 15(1), 18-25.
12. Rahman, H., Bukht, T. F. N., Imran, A., Tariq, J., Tu, S., & Alzahrani, A. (2022). A deep learning approach for liver and tumor segmentation in CT images using ResUNet. Bioengineering, 9(8), 368. MDPI.
13. Ullah, I., Imran, A., Ashfaq, A., & Adnan, M. (2022). Exploiting machine learning models for identification of heart diseases. Journal of Computing & Biomedical Informatics, 3(2), 1-20.
14. Ullah, T., Khan, J. A., Khan, N. D., Yasin, A., & Arshad, H. (2023). Exploring and mining rationale information for low-rating software applications. Soft Computing, 1-26.
15. Khan, N. D., Khan, J. A., Li, J., Ullah, T., & Zhao, Q. (2024). Mining software insights: uncovering the frequently occurring issues in low-rating software applications. PeerJ Computer Science, 10, e2115.
16. Khan, N. D., Khan, J. A., Li, J., Ullah, T., Alwadain, A., Yasin, A., & Zhao, Q. (2024). How do crowd-users express their opinions against software applications in social media? A fine-grained classification approach. IEEE Access.
17. Fatima, E., Kanwal, H., Khan, J. A., & Khan, N. D. (2024). An exploratory and automated study of sarcasm detection and classification in app stores using fine-tuned deep learning classifiers. Automated Software Engineering, 31(2), 69.